

Les données dans l'éducation – les données pour l'éducation

Bases et pistes de réflexion en vue de l'élaboration
d'une politique d'utilisation de données pour l'es-
pace suisse de formation



Impressum

Les données dans l'éducation – les données pour l'éducation. Bases et pistes de réflexion en vue de l'élaboration d'une politique d'utilisation de données pour l'espace suisse de formation.

educa.ch, sur mandat du comité de coordination de la numérisation de l'éducation (CC N)

Avec des contributions de :

Florent Thouvenin, Alfred Früh, Aurelia Tamò-Larrieux (Center for Information Technology, Society, and Law [ITSL], Université de Zurich)

Hannes Lubich (Institut für mobile und verteilte Systeme, FHNW), Martin Leuthold (SWITCH)

Andreas Kellerhals (ancien préposé aux Open Government Data, OGD)

Matthias Stürmer (Forschungsstelle für digitale Nachhaltigkeit, Université de Berne)

Équipe de projet educa.ch :

Nelly Buchser-Heer

Manuela Gloor

Simon Graber

Michael Jeitziner

Benjamin Volland

Markus Willi

Karl Wimmer

Irene Ziörjen

Photo de couverture : educa.ch/Marion Bernet

CC BY-NC-ND (creativecommons.org)

Berne, mai 2019

Proposition de citation :

educa.ch (2019) : Les données dans l'éducation – les données pour l'éducation. Bases et pistes de réflexion en vue de l'élaboration d'une politique d'utilisation de données pour l'espace suisse de formation. Berne.

Management summary

Intitulé « Les données dans l'éducation – les données pour l'éducation », le présent rapport a été réalisé par educa.ch sur mandat de la Confédération (SEFRI) et des cantons (CDIP). Il présente des bases et des pistes de réflexion en vue de l'élaboration d'une politique en matière d'utilisation de données pour l'espace suisse de formation. Ce rapport ne propose pas directement une politique d'utilisation de données, mais il esquisse au contraire les marges de manœuvre dans les limites desquelles une telle politique peut être formulée : il offre l'aperçu thématique requis à cet effet et contribue ainsi à approfondir la discussion sur l'utilisation sur l'utilisation des données relatives à l'éducation dans le cadre des structures existantes du système étatique de formation. Il aborde les questions des bases légales correspondantes, les aspects touchant à la sécurité, à l'interopérabilité et aux possibilités offertes par les données ouvertes. Enfin, ce rapport énonce des pistes de réflexion en vue de l'élaboration d'un cadre d'action praticable et applicable à l'échelle de toute la Suisse pour l'utilisation des données de l'éducation.

Mandat

Dans le cadre de leur collaboration dans le domaine de l'éducation et de la formation, la Confédération (Département fédéral de l'économie, de la formation et de la recherche [DEFR]) et les cantons (Conférence suisse des directeurs cantonaux de l'instruction publique [CDIP]) traitent également les questions qui ont trait à la numérisation. À cet effet, ils ont conjointement mis sur pied le « comité de coordination de la numérisation de l'éducation » (CC N). Par le biais de ce dernier, ils ont mandaté l'agence spécialisée educa.ch afin qu'elle prépare les bases destinées à l'élaboration d'une future politique en matière d'utilisation des données dans le système éducatif suisse. Le présent rapport « Les données dans l'éducation – les données pour l'éducation » met en œuvre ce mandat.

Ce rapport a été rédigé entre mai 2018 et mai 2019. Les travaux correspondants ont été planifiés en étroite concertation avec le Secrétariat général de la CDIP et le Secrétariat d'État à la formation, à la recherche et à l'innovation (SEFRI). Ils ont été réalisés en impliquant des spécialistes qui ont mis en lumière des aspects pertinents propres à l'utilisation des données dans le système éducatif suisse dans des contributions thématiques qui font partie du présent rapport. Il y avait une grande unanimité entre tous les acteurs pour considérer que le rapport devait mettre l'accent sur les avantages potentiels de l'utilisation des données pour le système éducatif suisse. Ce qui est à la base du présent rapport, c'est la conviction, partagée par tous les participantes et participants, selon laquelle les méthodes et les résultats de l'utilisation des données seront admis de manière accrue dans le cadre de l'évolution future du système de formation, du système scolaire et du système d'enseignement en Suisse – et que les acteurs du système éducatif suisse auront accès aux données pertinentes pour leur contexte et qu'ils devront donc être suffisamment compétents pour pouvoir les utiliser à leurs propres fins. Il s'agit ainsi de rendre possible l'utilisation des données dans le système éducatif suisse en conformité avec les droits individuels propres aux exigences de la protection des données. Au niveau systémique, les objectifs stratégiques y afférents de la politique de l'éducation et de la formation et de son pilotage devront eux aussi être repris par les instances compétentes.

Écosystème des données de l'éducation

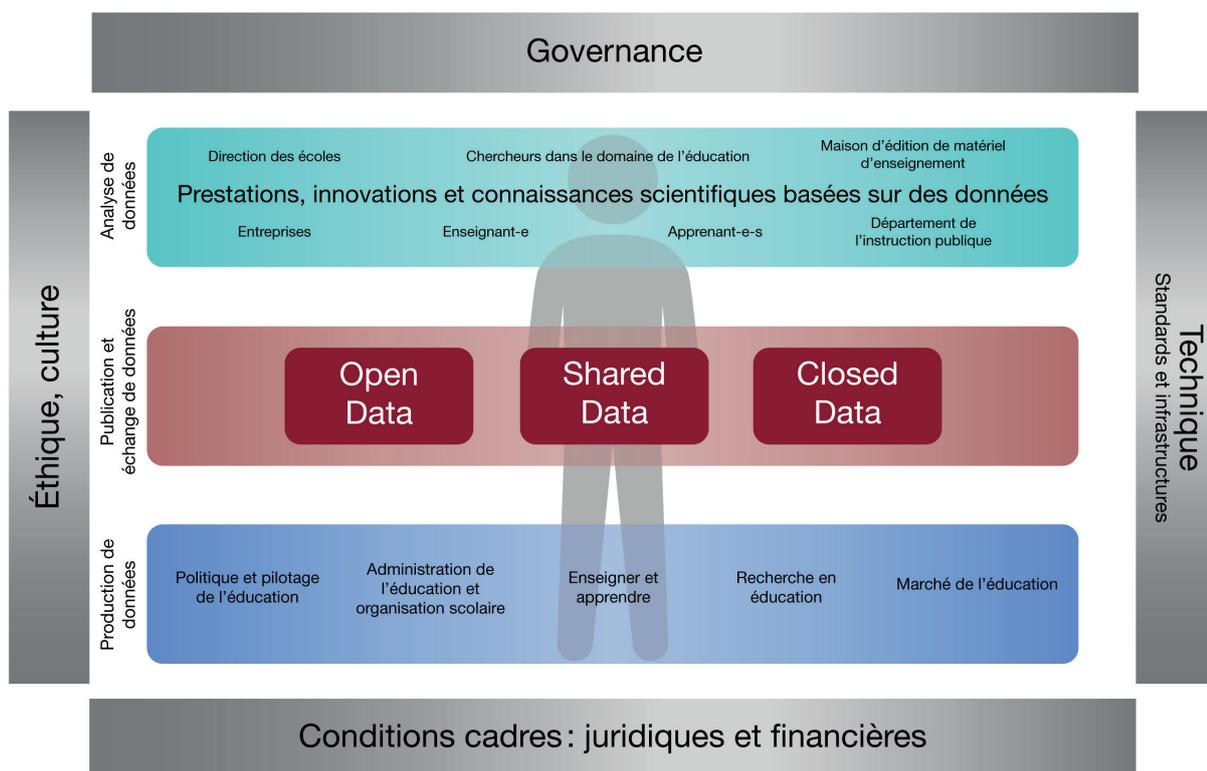
Dans ses parties de présentation, le rapport contient d'abord des clarifications de concepts, notamment la définition de l'expression « données de l'éducation », à savoir les « données qui sont générées dans le cadre du système éducatif qui concernent les énoncés relatifs à ce dernier ou qui permettent de tirer des déductions à ce propos » (voir chapitre 2), et qui résume les possibilités relatives à leur typologie. En outre, le présent document décrit les différents contextes d'utilisation des données de l'éducation ainsi que leurs groupes cibles, qui sont subdivisés en cinq types d'acteurs :

- **les responsables de la politique et du pilotage de l'éducation** dont les instances utilisent les données pour planifier, contrôler et piloter le système éducatif suisse,
- **les responsables de l'administration de l'éducation et de l'organisation scolaire** dont les structures utilisent les données dans le cadre de l'organisation et de l'administration du fonctionnement quotidien des institutions éducatives,

- **les responsables de la recherche en matière d'éducation** qui utilisent les données en tant qu'outil d'observation, d'enquête et d'explication de ce qui se passe effectivement au moment où la formation est dispensée,
- **les acteurs du marché de l'éducation** dont les participants (prestataires et clients) utilisent les données à des fins commerciales, et
- **les responsables de l'enseignement et de l'apprentissage** au cours desquels on utilise les données des enseignantes et des enseignants ainsi que les données des élèves dans le cadre de l'enseignement et de l'apprentissage quotidiens.

Sur la base d'un choix d'exemples de bonnes pratiques *good practices*, le rapport résume et met en évidence les avantages potentiels des données de l'éducation. Dans ce contexte, c'est dans le cadre du monitoring de l'éducation que l'utilité du recours aux possibilités offertes par l'utilisation des données est particulièrement frappante. Des exemples tirés du Luxembourg, d'Estonie et du Danemark qui, pour l'essentiel, en sont déjà à un stade de maturité avancé, montrent comment, aujourd'hui déjà, des données sont enregistrées, rassemblées et préparées dans des systèmes d'éducation nationaux. Grâce à un processus de prise de décision basé sur des preuves empiriques *evidence-based*, l'analyse différenciée de l'éducation que permet ce type de monitoring aide les différents acteurs à prendre des mesures appropriées lors de la mise en œuvre de projets de grande ampleur, et ce, tant au niveau local et régional que national. Il faut mettre en regard de ces possibilités d'utilisation des données de l'éducation structurées sur le plan systémique d'autres possibilités d'utilisation qui visent, en fin de compte, à renforcer l'individualisation de l'enseignement et de l'apprentissage, par ex. en agissant sur des situations d'enseignement ou en regroupant des contenus d'apprentissage. Les études réalisées par les responsables de la recherche en matière de l'éducation attribuent à ces outils un potentiel considérable, mais hormis certains projets individuels, ils n'ont guère encore été utilisés à grande échelle.

Ensuite, les auteurs ont élaboré un « écosystème des données de l'éducation » jouant un rôle de modèle qui constitue le cœur du présent rapport. Selon ce modèle, il est possible de préparer de manière standardisée, à partir des différentes sources de données disponibles dans le système éducatif, un écosystème de données qui satisfait à la fois aux exigences organisationnelles et communautaires ainsi qu'aux exigences en matière de gestion du personnel. Une fois subdivisées en trois types de données, à savoir en données ouvertes (*Open Data*), données partagées (*Shared Data*) et données fermées (*Closed Data*), ces données peuvent alors être libérées de manière contrôlable aux fins d'utilisation. Enfin, des contributions de spécialistes des questions touchant aux bases légales, à la sécurité de l'information, au principe des données ouvertes, aux possibilités de standardisation et d'interopérabilité des données de l'éducation présentent les conditions cadres requises pertinentes au niveau systémique qu'il s'agit de prendre en compte lors de la mise en place de cet écosystème et identifient des lacunes dans ce domaine.



Graphique 1 : Écosystème des données de l'éducation (selon Gollietz 2018)

Défis à relever

À l'heure actuelle, il faut mettre en regard du potentiel important que recèle l'utilisation des données de l'éducation sa faible acceptation auprès des milieux intéressés et leur scepticisme quant au fait que les données numériques seront effectivement utilisées. Les interviews menées dans le cadre de ce rapport montrent que ces opinions sont largement répandues parmi les acteurs du système éducatif. Ces avis sont notamment justifiés par le fait que de nombreux aspects juridiques et procéduraux de la protection des données et de la disponibilité des données dans le domaine de l'éducation ne sont pas encore clarifiés à ce jour. En outre, des lacunes dans l'équipement technique de base sont critiquées et, de manière générale, on présume que les compétences actuelles en matière d'utilisation des données sont plutôt faibles.

Afin que le potentiel de l'utilisation des données de l'éducation puisse s'épanouir et déployer tous ses effets, ce qui fait défaut en priorité, ce sont des réglementations praticables qui devraient créer un équilibre entre les droits de l'individu à la sécurité de l'information et à l'utilisation minimale des données, et les besoins de la société d'utiliser ces données afin d'améliorer le système éducatif et les résultats de l'apprentissage. Ces réglementations devront être réunies dans un cadre d'action qui puisse instaurer, pour chaque individu, la transparence sur les possibilités d'utilisation des données de l'éducation adaptées à la situation concernée. Ce faisant, les points suivants devront être pris en compte :

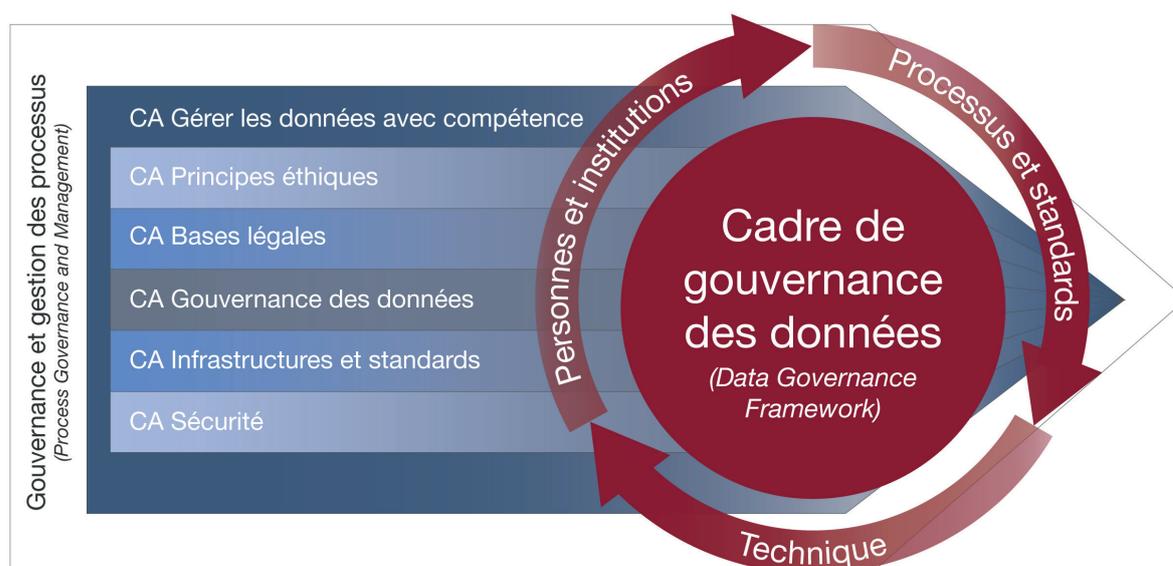
- Ce cadre devra s'appliquer à l'échelle de toute la Suisse : dans une société numérisée, des solutions isolées n'ayant aucun sens ; elles font obstacle à l'efficacité et créent des inégalités (par ex. pour les établissements scolaires de cantons différents ou pour des prestataires de services qui seraient tenus de respecter des prescriptions différentes en matière de droit de la protection des données). La Confédération, les cantons et les communes doivent collaborer afin de garantir la perméabilité – également au-delà des seuls domaines de l'éducation et de la formation.

- Ce cadre sera établi en adoptant une approche globale et intégrative : il ne s'agit pas de trouver des solutions individuelles, mais au contraire de réglementer de manière globale les modalités d'utilisation des données. Ce faisant, il faut inclure dans le processus de recherche de solutions toutes les conditions cadres propres à l'écosystème des données de l'éducation.
- Ce cadre devra être agile et flexible : pour pouvoir réagir rapidement et de manière adéquate aux futurs défis, il faut prévoir des solutions qui permettent de réévaluer continuellement la situation et qui incluent l'intégration, à un rythme élevé, des adaptations requises pouvant aller jusqu'à des remaniements de portée fondamentale.
- Ce cadre devra être développé conjointement : sa mise en œuvre ne se fera pas uniquement par le biais des instances et institutions actuelles concernées du système éducatif, mais elle nécessite au contraire une étroite collaboration entre tous les groupes cibles. Il s'agit de pondérer entre eux leurs intérêts respectifs et de les reprendre dans une solution consensuelle. Et ce, non seulement en raison des aspects financiers, mais surtout également parce que des compétences techniques sont requises à cet effet.

Les défis à relever en relation avec l'utilisation des données de l'éducation revêtent un degré d'urgence élevé, et il faut donc s'y attaquer immédiatement. Aujourd'hui déjà, des prestataires privés (et aussi étatiques) de supports de cours numériques et d'autres services numériques sont en effet en train de collecter des données détaillées sur le comportement d'apprentissage et sur les résultats obtenus par les élèves. Très souvent, en vertu de conventions d'utilisation actuellement en vigueur, les prestataires sont les propriétaires exclusifs de ces données. En règle générale, les institutions qui génèrent des données ne peuvent absolument pas accéder à ces données, et les personnes qui génèrent des données ne peuvent y accéder que de manière limitée (par ex. par le biais du droit d'accès dont elles disposent en vertu du droit de la protection des données). En outre, il n'est pas certain que ces institutions et ces personnes soient au clair sur l'étendue et la nature de la collecte de données effectuée par les prestataires de services. On peut aussi douter qu'à l'heure actuelle, la plupart de ces institutions soient en mesure d'utiliser ces données de manière lucrative, quand bien même elles y auraient accès.

1.1 Champs d'action

Compte tenu du fait que le présent rapport a pour but de préparer les bases requises en vue de l'adoption d'une future politique en matière d'utilisation de données, différents champs d'action (CA) thématiques ont été identifiés à cet effet en collaboration avec les expertes et les experts sollicités. Considérés dans leur globalité, ils contribuent, d'une part, à ce que l'utilisation des données puisse véritablement avoir lieu et, d'autre part, à ce que cette utilisation se déroule dans un cadre dûment réglementé.



Graphique 2 : Champs d'action (CA)

Le champ d'action « Gérer les données avec compétence » constitue la base préalable à toute utilisation des données. Il a pour but de faire en sorte que tout acteur du système éducatif doit pouvoir disposer de certaines compétences de gestion des données correspondant au rôle qu'il exerce dans le système. Les champs d'action « Principes éthiques » et « Sécurité » constituent aussi simultanément une directive d'objectifs : la future utilisation des données dans le système éducatif doit être sécurisée et doit être guidée par des principes éthiques appropriés. Les champs d'action « Bases légales » et « Infrastructures et standards » offrent tous deux des instruments de régulation, mais à des niveaux différents. Le champ d'action « Gouvernance des données » vise déjà la mise en œuvre des résultats obtenus dans les différents champs d'action (qui est responsable, comment les processus doivent-ils se dérouler, etc.). Enfin, le champ d'action « Gouvernance et gestion des processus » assume, en tant qu'« interface », la tâche de coordination entre les différents champs d'action et, partant, constitue le point de départ en vue d'établir un cadre d'action cohérent dans ce domaine.

Pistes de réflexion et d'élaboration

Une utilisation dûment réglementée des données de l'éducation ne tombe pas du ciel. Elle ne s'impose pas « simplement comme cela » grâce à une augmentation de la pratique dans ce domaine. Il s'agit plutôt de considérer les données de l'éducation comme des infrastructures importantes du système éducatif, étant précisé qu'il faut investir des moyens et ressources dans sa mise en place dont il faudra ensuite assurer la maintenance, et dont l'utilisation devra être réglementée.

La dernière partie du présent rapport présente des pistes de réflexion concrètes axées sur les différents champs d'action qui ont tous pour but de contribuer à l'élaboration d'une politique globale d'utilisation des données pour l'espace suisse de formation.

Pistes de réflexion spécifiques aux différents champs d'action

CA 1 : Gérer les données avec compétence

Encourager la compréhension pour la transformation numérique en général et pour le rôle de l'utilisation des données en particulier, notamment auprès des commissions chargées de prendre des décisions. Cela inclut la sensibilisation à l'importance de l'utilisation des données, la mise en place d'une prise de conscience en matière de données (*Data Awareness*) et d'une initiation aux compétences de base pour la maîtrise des données (*Data Literacy*) dans le système éducatif

CA 2 : Bases légales

Développement continu et coordonné, à l'échelle de toute la Suisse, du droit d'accès aux données en vertu de la législation sur la protection des données afin de garantir la portabilité des données dans l'espace éducatif suisse.

Étudier l'opportunité d'introduire, à l'échelle de toute la Suisse, un droit d'accès aux données de l'éducation qui sont détenues à titre privé.

Étudier l'opportunité d'introduire un droit d'accès privilégié renforcé et standardisé, valable pour toute la Suisse, qui serait accordé aux responsables de la recherche en éducation afin de garantir un accès approprié aux données de l'éducation pour la recherche, l'évaluation et la planification.

CA 3 : Sécurité

Définition et implémentation de mesures de protection minimales pour la sécurité de l'information dans la totalité des institutions éducatives suisses.

CA 4 : Principes éthiques

Élaborer une première proposition relative à des principes éthiques de base à observer dans le cadre de l'utilisation des données de l'éducation.

CA 5 : Infrastructures et standards

Implémentation d'un standard uniforme, valable pour toute la Suisse, relatif aux données de l'éducation.

Mise en place d'un entrepôt de données (*Data Warehouse*) national pour les données de l'éducation.

Tirer entièrement parti du potentiel des données ouvertes dans le domaine de l'éducation à titre de base préalable en vue d'adopter de nouvelles pistes créatives de résolution de problèmes, pour exploiter de nouvelles connaissances ainsi que pour accroître l'efficacité dans l'administration.

CA 6 : Gouvernance des données

Définition de structures et processus de gouvernance des données appropriées, en accord avec les mesures à prendre dans les autres champs d'action.

CA 7 : Gouvernance et gestion des processus

Garantir la coordination de la totalité des mesures à prendre.

Propositions de mise en œuvre

Centre de compétences pour l'utilisation des données

Mise en place d'un réseau de compétences pour l'utilisation des données dans le système éducatif. Il exploite et met en valeur les compétences techniques pertinentes en matière d'utilisation des données axées sur les différents champs d'action précités et sert de centre d'aiguillage et d'information à tous les acteurs du système éducatif pour répondre à leurs questions dans ce domaine.

Pour une concrétisation rapide, il est recommandé de mettre l'accent, dans un premier temps, sur des problématiques spécifiques à l'éducation. Lors d'une étape ultérieure, il pourrait être tout à fait judicieux d'institutionnaliser ce réseau et de le transformer en un centre de compétences général pour l'utilisation des données au sein duquel on créerait un service chargé de la formation (voir Frey, Rogg, Savolainen, Schmid, & Wandeler 2018).

Encouragement ciblé à des projets pilotes afin d'établir un cadre adéquat

Soutenir des projets pilotes qui feront la lumière sur le fonctionnement de cet « écosystème des données de l'éducation » dans la réalité, concrétiser des objectifs d'utilisation des données, et tester des pistes de solutions dans le champ de tension qui existe entre les intérêts respectifs des différents groupes cibles. Les solutions qui ont fait leurs preuves devront être mises en œuvre immédiatement.

Ce mode de faire rendra simultanément les acteurs capables de maîtriser le système (le caractère urgent et la nécessité d'agir deviendront visibles, et tant les avantages potentiels que les défis à relever deviendront observables et pourront être expérimentés dans la réalité quotidienne du système éducatif). Cela permettra d'analyser dans le détail la nécessité d'agir et d'élaborer des solutions pratiques conformes aux besoins (y compris la définition des rôles et des responsabilités).

Table de matières

1	Introduction	11
2	Données et types de données	14
3	Écosystème des données de l'éducation	21
3.1	Aperçu d'un modèle	21
3.2	Aspects déterminants de l'utilisation des données	22
3.3	Cadre	23
3.4	Contextes d'utilisation	29
3.5	Conclusion : l'espace suisse de formation en tant qu'espace de données dispersées – absence d'une culture d'utilisation des données	44
4	Discussion de certains concepts clés	46
4.1	Concepts juridiques fondamentaux	47
4.2	Sécurité de l'information	66
4.3	Open Data	83
4.4	Interopérabilité	96
5	Utilisation des données dans le domaine de la formation en comparaison internationale	105
5.1	Méthode	105
5.2	Efforts déployés dans des pays européens	105
5.3	Exemples pratiques (Europe)	106
5.4	Efforts déployés aux États-Unis	108
5.5	Conclusion	108
6	Potentiels et défis de l'utilisation des données dans l'éducation	110
6.1	Introduction	110
6.2	Potentiels	111
6.3	Cas d'application	115
6.4	Défis à relever	122
7	Une politique cohérente d'utilisation des données pour l'espace suisse de formation	129
7.1	Les données dans l'éducation – les données pour l'éducation	129
7.2	Nécessité d'un cadre commun	129
7.3	Caractère urgent	130
7.4	Champs d'action	131
7.5	Conclusion	134
7.6	Pistes de réflexion en vue de l'élaboration d'une politique d'utilisation des données pour l'espace suisse de formation	134
8	Glossaire	138
9	Bibliographie	144
10	Annexes	153

1 Introduction

L'un des effets les plus lourds de conséquences de la numérisation réside dans l'ampleur et la vitesse à laquelle nous pouvons enregistrer, mémoriser et traiter aujourd'hui des données. En effet, les données numériques prennent naissance continuellement et partout – dans toute économie nationale, dans tout secteur, dans tout établissement, et pour toute utilisatrice ou utilisateur de terminaux et applications numériques. La quantité d'information numérique générée augmente en conséquence à un rythme vertigineux. Une étude estime qu'en 2020, quelque 1,8 MB de données par seconde seront générées pour chaque personne dans le monde (IDC 2014), et que l'étendue des données mémorisées, qui est de 33 zettabytes aujourd'hui, passera à 175 zettabytes d'ici à 2025 (IDC 2018). Les possibilités qui résultent de l'interaction entre ces flux de données et une performance de calcul des ordinateurs en constante amélioration sont considérables. Réalisée en 2011, une étude de McKinsey, par exemple, estime qu'à eux seuls, les gains d'efficacité possibles qui résultent de l'utilisation systématique et intégrée des données existantes s'élèvent à 100 milliards d'euros par année pour les administrations gouvernementales européennes (Manyika et al. 2011). Dans la vie quotidienne également, l'évaluation et l'exploitation des données a déjà clairement déployé ses effets. Des diagnostics de maladies assistés par ordinateur à la planification des interventions de la police en passant par l'estimation du risque de récidive des délinquants jusqu'à l'évaluation de la solvabilité individuelle, la publicité personnalisée ou encore la traduction automatisée de textes écrits et de la langue parlée – l'exploitation de grandes quantités de données est à l'origine de tous ces types d'utilisation. Le fait que ces processus peuvent aussi receler un potentiel disruptif considérable est arrivé dans le débat public, au plus tard avec la publication des enquêtes sur les conséquences de l'automatisation sur le marché du travail (Degryse 2016 ; Frey & Osborne 2017 ; Conseil fédéral 2017).

Dans le secteur de l'éducation également, la collecte et l'utilisation de données numériques a nettement gagné en importance ces dernières années (Cope & Kalantzis 2015 ; Cope & Kalantzis 2016 ; Lang, Wise, Siemens, & Gasevic, 2017). Outre le transfert et la saisie continuelle de données « analogiques » sous forme électronique, l'attention se focalise toujours plus sur les données qui sont tirées d'applications numériques telles que les systèmes d'encouragement à l'apprentissage ou les plateformes d'apprentissage qui sont désormais placées au premier plan. En effet, l'utilisation de ces données promet non seulement de nouvelles chances de monitoring et de pilotage du système d'éducation et de ses institutions (Wolter et al. 2018 ; Williamson 2016), mais offre aussi des possibilités d'adaptation de contenus et méthodes d'apprentissage qui n'étaient guère réalisables jusqu'à présent, ainsi que des possibilités de mesure des performances et des carrières dans le domaine de l'éducation liées aux conditions préalables et aux intérêts des apprenants individuels (Baer & Norris 2017).

Alors qu'un nombre croissant d'acteurs du système d'éducation suisse considèrent ces développements comme recelant d'importants potentiels, nombreux sont ceux qui font toutefois preuve d'attentisme ou de scepticisme quant à l'utilisation effective des contenus numériques et des données générées par ce mode d'utilisation. Ces critiques s'expliquent non seulement par les lacunes qui existent actuellement en matière d'équipement technique de base, et par le niveau de compétences disponibles plutôt limité dans ce domaine à ce jour, mais surtout par le fait que beaucoup d'aspects juridiques et procéduraux de la protection des données et de la disponibilité des données dans le domaine de l'éducation ne sont pas encore clarifiés jusqu'à présent. Dans les écoles et l'administration scolaire, un « cadre praticable relatif à la sécurité du droit » (Döbeli Honegger, Hielscher, & Hartmann 2018, p. 128) fait donc défaut. Ce cadre devrait créer un équilibre entre les droits de l'individu à la sécurité de l'information et à l'utilisation minimale des données, et les besoins de la société d'utiliser ces données afin d'améliorer le système d'éducation et les résultats de l'apprentissage. Cela semble d'autant plus important que, déjà aujourd'hui, des prestataires privés

(et étatiques) de supports de cours numériques et d'autres services numériques sont en train de collecter des données détaillées sur le comportement d'apprentissage et sur les résultats obtenus par les élèves. La plupart du temps, en vertu de conventions d'utilisation actuellement en vigueur, les prestataires sont les propriétaires exclusifs de ces données. En règle générale, les établissements qui génèrent des données ne peuvent absolument pas y accéder, et les personnes qui génèrent des données ne peuvent y accéder que de manière limitée (par ex. par le biais du droit d'accès dont elles disposent en vertu du droit de la protection des données). En outre, il n'est pas certain que ces établissements et ces personnes soient au clair sur l'étendue et la nature de la collecte de données effectuée par les prestataires de services. On peut aussi douter qu'à l'heure actuelle, la plupart de ces établissements soient en mesure d'utiliser ces données de manière lucrative, quand bien même elles y auraient accès.

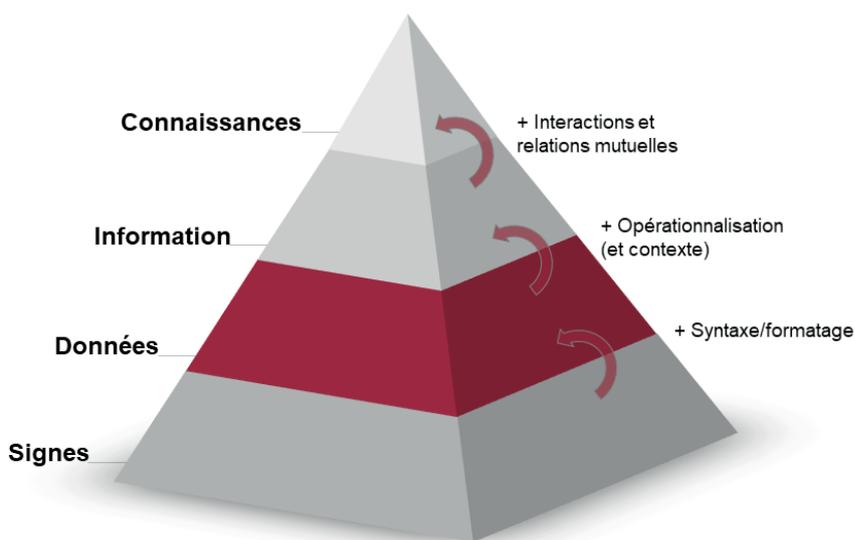
De manière générale, des réflexions qui feraient la lumière sur les défis d'ordre systémique à relever dans ce domaine font défaut, et on déplore le manque d'informations formulées sur la nécessité d'agir et les mesures à prendre qui iraient au-delà de la résolution de problèmes individuels (par ex. en lien avec des dispositions du droit de la protection des données relatives à la transmission de données sur les élèves à des prestataires de services web). Les études qui mettent l'accent sur ce niveau – par ex. l'étude de Jarchow et Estermann (2015) sur les chances et les risques des applications du big data pour l'économie et la société – laissent de côté le secteur de l'éducation. Et même le plan d'action de la stratégie « Suisse numérique » ne prévoit aucune mesure concrète pour le secteur de l'éducation pour la mise en œuvre des objectifs stratégiques dans le domaine « Contenus et données numériques ». Les objectifs et mesures énoncés par le Secrétariat d'État à la formation, à la recherche et à l'innovation (SEFRI) dans le domaine « Protection des données et sécurité des données dans l'espace de formation numérique » ne se focalisent à leur tour que sur une partie seulement des questions touchant à l'accès aux données, à la propriété des données et à la protection des données en visant la création d'une identité numérique fédérée à l'échelle de toute la Suisse pour l'ensemble des acteurs du système éducatif suisse (voir [projet FIDES](#)). En introduisant une telle identité numérique, on crée certes une condition préalable nécessaire à l'adoption d'une politique commune d'utilisation des données de l'éducation, mais c'est insuffisant. Ce qui est nécessaire et urgent, c'est une approche globale qui, outre l'instauration de la sécurité et de la confiance, puisse aussi permettre l'innovation. La « Stratégie pour la gestion de la transition numérique dans le domaine de l'éducation (2018) » de la Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP) a fait un premier pas dans cette direction. Outre la nécessité de garantir la sécurité de l'information, elle souligne aussi l'importance de l'utilisation effective des données dans le système éducatif suisse en tant qu'objectif stratégique.

Entre-temps, la Confédération et les cantons ont reconnu que la complexité de la thématique de l'utilisation des données dans le système éducatif nécessite une coordination commune. Dans le cadre du champ thématique 2 « Accès aux données, propriété des données et protection des données » de leur comité de coordination de la numérisation de l'éducation, ils ont donc confié à leur agence spécialisée commune educa.ch le mandat consistant à élaborer les bases en vue d'une future politique d'utilisation des données. Le présent rapport est le résultat de cette mesure. Il a pour but de prescrire au comité de coordination de la numérisation de l'éducation (CC N) des axes prioritaires stratégiques en lui adressant des recommandations d'action concrètes. Il offre globalement une orientation pour tous les acteurs et contribue ainsi à approfondir la discussion politique et sociétale sur l'utilisation des données de l'éducation. À cette fin, il explique les fondements des données et des analyses de données (chapitre 2) ; il éclaire les contextes et institutions au sein desquels, aujourd'hui déjà, on collecte, on mémorise et on évalue des données (chapitre 3). Il met aussi en évidence les consensus et les divergences entre les exigences posées par les différents acteurs à une politique d'utilisation des données dans le domaine de l'éducation (chapitre 3.4). Pour approfondir encore cette thématique, différents spécialistes explicitent de manière détaillée des concepts clés propres à l'utilisation des données ainsi que leur importance pour une future utilisation des données : il s'agit des concepts juridiques (chapitre 4.1), de la sécurité de l'information (chapitre 4.2), des Open Data (chapitre 4.3) et de l'interopérabilité (chapitre 4.4). Au chapitre 5, on présente les résultats d'une recherche et d'une enquête comparative entre différents pays. Cette enquête fournit un aperçu des approches existantes au plan international s'agissant de la réglementation de l'utilisation des données dans l'éducation. Un résumé des résultats de ces analyses approfondies comportant un aperçu complet des opportunités et défis de l'utilisation des données figure au chapitre 6. Cet aperçu forme la base à partir de laquelle les propositions et recommandations d'action sont ensuite formulées (chapitre 7).

Dans le présent rapport, les questions qui ont trait à l'utilisation d'œuvres protégées par le droit d'auteur, par ex. des supports de cours et du matériel d'enseignement numériques, ont été mises entre parenthèses et ne seront pas traitées. Il faudra y répondre dans le contexte du droit d'auteur et de la propriété intellectuelle, ce qui ne peut être réalisé dans le cadre de cette étude.

2 Données et types de données

Si une réponse généralement acceptée à la question de savoir ce que sont des « **données** » fait défaut à ce jour (Voss 2013), on les définit la plupart du temps comme une suite (syntactiquement ordonnée) de **signes** (ou comme des fonctions continues de signes) qui servent à collecter, à mémoriser et à représenter des faits et des observations provenant du monde réel (Mertens et al. 2017 ; Beynon-Davies 2002). L'attribution d'une suite bien définie de signes à un phénomène donné est désignée par terme de codification ou d'opérationnalisation. Elle permet de traduire des informations en données et, à l'inverse, de traduire en retour ces données sous forme d'**informations**. Si ces informations sont intégrées dans des contextes plus larges, et si elles sont complétées par des interactions réciproques, il est possible d'en déduire des **connaissances**. Les relations entre les signes, les données, les informations et les connaissances sont souvent représentées sous la forme de ce qu'on appelle la « hiérarchie des connaissances » (selon Krcmar 2015 ; Rehäuser & Krcmar 1996).¹

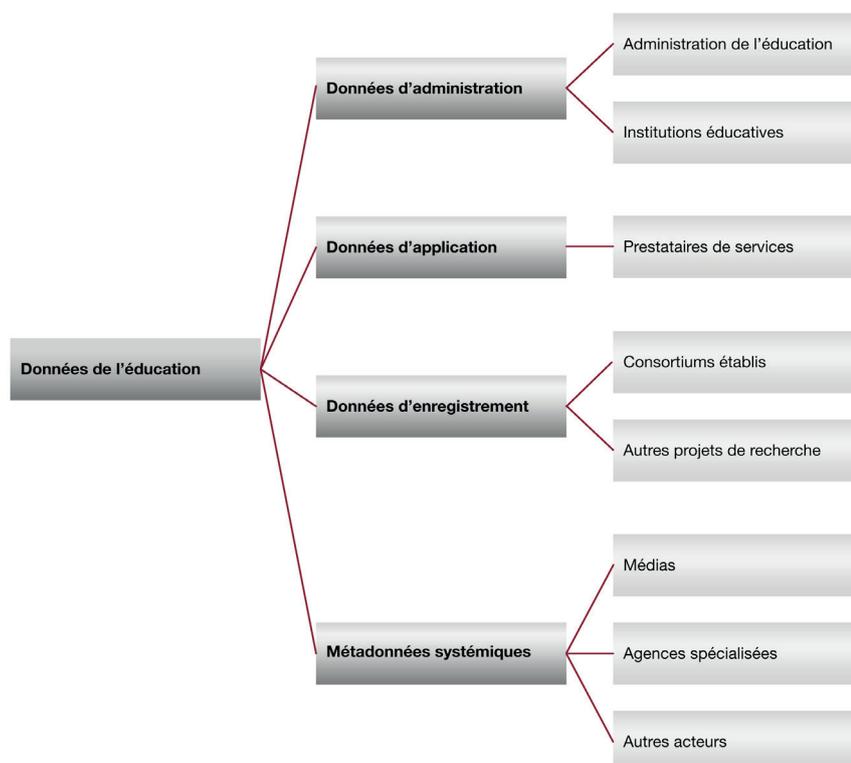


Graphique 3 : Hiérarchie des connaissances

Lorsque la collecte, la mémorisation ou la représentation de données s'effectuent sous forme électronique, et qu'elles permettent à des machines de traiter les données ainsi engendrées, on parle dans ce cas de « **données numériques** ». Le présent rapport se conforme à cette définition.

¹ Très souvent, les transitions entre les différents niveaux de cette hiérarchie ne sont pas techniquement descriptibles sans ambiguïté. Partant, lorsqu'on traduit des informations en données et, à l'inverse, lorsqu'on déduit de ces données des informations et des connaissances, il existe toujours une certaine marge d'interprétation. En fonction du contexte social, culturel et individuel concerné, cette marge peut donner lieu à des interprétations différentes du même ensemble de données ou du même résultat d'analyse. Dans la pratique, l'établissement et la communication claire de règles et de modes d'emploi correspondants (par ex. sous la forme de modes d'emploi de codification et de rapports méthodologiques) se sont avérés utiles afin de réduire au maximum les divergences d'interprétation individuelles dans l'utilisation des données. En outre, ces écarts peuvent être enregistrés au moyen d'indicateurs correspondants – il s'agit de ce qu'on appelle des mesures de fiabilité.

Si l'on s'en tient à une définition la plus large possible du terme de « données », toutes les données qui sont générées dans le cadre du système éducatif qui portent sur des énoncés au sujet de ce dernier ou qui permettent des déductions à ce propos en tant que « données se rapportant à l'éducation » ou plus brièvement, en tant que « **données de l'éducation** » peuvent être désignées comme telles. Ces données permettent – du moins en théorie – de procéder à une observation, une description et une évaluation complètes du système éducatif et de ses acteurs individuels relevant aussi bien du niveau « micro » que du niveau « macro ». Elles fournissent des informations sur les personnes impliquées comme les élèves ou les collaboratrices et collaborateurs travaillant dans l'administration, sur les établissements auxquels ils appartiennent, sur le système éducatif qui relie ces établissements, et sur le contexte social, culturel et économique dans lequel ce système est intégré. À cet égard, elles ne se rapportent pas exclusivement aux acteurs étatiques comme les écoles ou les départements cantonaux d'instruction publique, mais contiennent en outre des informations qui sont générées et mises à disposition par des acteurs non étatiques pertinents (par ex. organisations non gouvernementales, représentantes et représentants des médias, ou participantes et participants issus de l'économie privée). Dès lors, les données se rapportant à l'éducation et à la formation relèvent de nombreux niveaux et se rencontrent dans une multitude de contextes auprès des acteurs les plus divers. Le graphique ci-dessous détaille la structure des données de l'éducation sous forme de schéma selon leur origine ou selon leurs sources.



Graphique 4 : Données selon leur contexte de naissance

Selon leur contexte de naissance, les données de l'éducation peuvent être décrites comme suit :

- **Les données d'administration** comprennent toutes les données qui sont générées et enregistrées dans le cadre de l'exploitation (quotidienne) et de la gestion de l'exploitation de l'éducation. Elles constituent le canal le plus important pour la génération des données sur l'éducation en Suisse. Sous une forme fortement agrégée, ces données représentent la colonne vertébrale du monitoring actuel de l'éducation (Wolter et al. 2018). En font partie, outre les informations sur la structure et la mise en place du système éducatif, les conditions cadres politiques, économiques et démographiques, notamment les données sur les inputs effectués dans le système, par exemple sous la forme du nombre d'enseignantes et d'enseignants, d'apprenantes et d'apprenants, ou d'établissements d'éducation, ainsi que les informations sur les dépenses nécessaires à leur fonctionnement. En règle générale, ces informations concernent la gestion du système scolaire et relèvent du niveau de l'unité administrative responsable concernée. Elles contiennent donc des informations sur chaque élément du système dans son ensemble.

- **Les données d'enregistrement** sont les données qui sont générées dans le cadre de recensements et d'enregistrements institutionnalisés (par ex. enquêtes COFO, PISA, etc.) et « spontanés » (par ex. recherche universitaire ou de hautes écoles) réalisés dans le domaine de l'éducation. Elles servent souvent à obtenir des informations qui ne peuvent pas, ou ne peuvent pas encore être obtenues à partir des seules données d'administration ; elles comportent donc aussi un degré de pertinence élevé pour le monitoring actuel de l'éducation. En font partie notamment les résultats et les outputs du système éducatif sous la forme d'informations sur les compétences des apprenants, sur l'évaluation du matériel d'enseignement, sur l'évaluation des techniques et concepts de l'enseignement, ou sur la saisie du taux de satisfaction envers le système éducatif. La plupart du temps, ces informations sont obtenues à l'aide de sondages ou d'enquêtes, souvent en combinaison avec des tests standardisés. Les informations tirées de ces enquêtes portant sur l'ensemble du système se fondent généralement sur une sélection d'échantillons représentatifs.
- **Les données d'application** sont les données qui sont produites dans le cadre de l'acquisition et de l'utilisation de produits et prestations de services qui sont proposés par des acteurs – il s'agit la plupart du temps d'acteurs de l'économie privée – aux fins de pilotage et d'exploitation du système éducatif ainsi que des institutions éducatives. Font partie de ces acteurs, outre les maisons d'édition de matériel d'enseignement, notamment les exploitants de plateformes d'apprentissage ou d'autres prestations de services numériques. En règle générale, les données qui prennent naissance dans le cadre de ces applications sont générées chez les prestataires respectifs concernés. Leur degré de pertinence est limité à la partie de l'ensemble du système correspondant aux offres qui sont effectivement consultées et utilisées. Avec la diffusion croissante des outils de travail numériques (programmes de management, plateformes d'apprentissage, logiciels d'apprentissage, etc.), la quantité de données produites chez les prestataires de ce genre de services augmente continuellement. En Suisse, il est donc prévu de procéder à une intégration complète de ces sources de données dans le monitoring de l'éducation ; ce projet constitue une partie de la stratégie de numérisation de la CDIP (CDIP 2018b).
- **Les métadonnées systémiques** contiennent toutes les données qui prennent naissance lors de la description externe du système éducatif, de ses acteurs, de ses résultats et de ses processus.² En d'autres termes, il s'agit de données qui ne sont pas générées en priorité dans ou pour l'exploitation du système éducatif, mais qui sont au contraire produites à partir d'un point de vue externe de rang hiérarchique supérieur par le biais de la description, et surtout par le biais de l'évaluation de cette exploitation. Relèvent en priorité de cette catégorie, outre des informations générales sur la mise en place et la structure du système éducatif, des positionnements, des attentes et des scénarios tels qu'ils sont formulés, par exemple, par des agences spécialisées, des associations, des partis politiques, des groupes de pression ou des laboratoires d'idées. S'y ajoutent les données qui reflètent les opinions et les jugements de valeur des groupes cibles pertinents sur le système éducatif et ses processus. Les comptes rendus et commentaires des médias en font également partie.

Au sein de chaque contexte différent, une multitude d'informations sont générées sur les acteurs individuels du système éducatif, sur leurs caractéristiques et leur comportement ainsi que sur leurs relations mutuelles. Jusqu'à présent, il n'existe aucun standard global en la matière pour la Suisse et, de ce fait, aucune classification généralement acceptée de ces données. Bien que la création d'un tel standard serait souhaitable et d'une nécessité urgente, un tel travail sortirait du cadre du rapport actuel.³ C'est pourquoi, aux fins d'uniformisation de la discussion ultérieure, en nous inspirant du schéma de données normalisées des *Common*

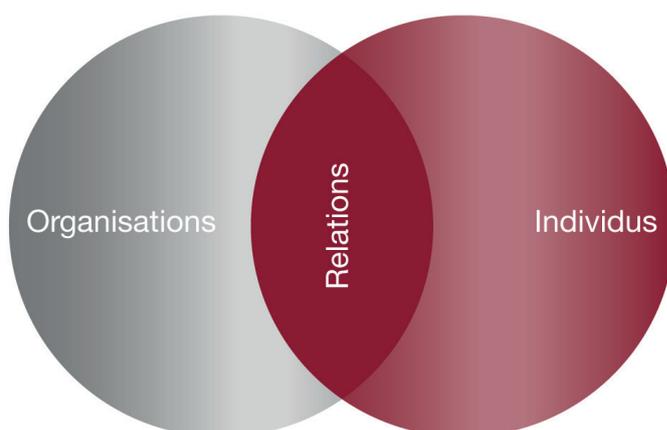
² De manière générale, le concept de « métadonnées » désigne toute forme de données structurées qui mettent à disposition les informations sur d'autres données. Elles servent à décrire et documenter des ensembles de données dans le but de pouvoir les retrouver et de les rendre comparables (Gregory, Heus, & Ryssevik 2009). Dans ce sens, il existe des métadonnées sur tout ensemble de données qui est obtenu dans le cadre d'activités administratives, de l'utilisation d'applications, de recensements ou d'enregistrements. Les « métadonnées » désignent donc des informations au sujet de données provenant du système éducatif. Afin de les délimiter par rapport aux simples métadonnées, nous entendons ici par « métadonnées systémiques » des informations sur le système éducatif en tant que tel.

³ À titre d'exemple de standard de portée relativement large pour les données de l'éducation, nous renvoyons au standard américain Common Education Data Standards (CEDS). Dans sa version actuelle (7.1), il met à disposition des définitions pour 4259 concepts ou éléments (à titre supplémentaire, il existe 1813 définitions pour des éléments qui sont utilisés afin de garantir l'interopérabilité du CEDS avec des standards similaires aux États-Unis, par ex. le standard de l'Ed-Fi Alliance). Ces éléments sont assignés à 163 catégories, qui sont à leur tour subdivisées en 59 unités d'organisation afin d'être finalement attribués à l'un des douze domaines de formation. En outre, ce standard fixe des modèles de données avec des règles de codification à observer pour ces éléments, et il offre des instruments pour l'intégration de sources de données non standardisées.

Education Data Standards (« CEDS Normalized Data Schema »)⁴, nous proposons d'adopter un modèle de données fortement simplifié de rang hiérarchique élevé qui est axé en priorité sur les acteurs et sujets concernés. Ce schéma distingue entre deux groupes principaux de sujets de données :

- **Individus** : relèvent de cette catégorie toutes les données qui décrivent les propriétés et les caractéristiques, le comportement et les prestations de personnes physiques.
- **Organisations** : toutes les données qui décrivent les propriétés et les caractéristiques, les états, les situations ou les prestations d'unités non naturelles telles que des classes d'école, des établissements scolaires, l'administration de l'éducation ou une ressource d'apprentissage. Outre les données qui sont produites originellement pour ce type d'informations (par ex. le nombre de bâtiments faisant partie d'un établissement scolaire et les sites où ils se trouvent), on enregistre à cet effet à titre supplémentaire des informations par le biais de l'agrégation de données sur des individus (par ex. nombre d'apprenants qui étudient dans un établissement scolaire, ou pourcentage de personnes qui ont interrompu leurs études dans un cycle de formation).

La distinction entre ces deux groupes de sujets de données n'est pas nette, ni exempte de chevauchements puisqu'en fin de compte, les organisations sont composées des individus qui en font partie, et puisque le comportement individuel est déterminé, du moins pour une part, par sa position au sein d'une organisation. Au sein du standard CEDS, ce modèle sert surtout de guide pour la structuration des bases de données relationnelles. Cela signifie qu'il sert de moyen auxiliaire lors du développement pratique d'un schéma relationnel normalisé⁵ pour les données collectées dans un contexte de formation. Dans ce sens, distinguer entre ces deux groupes principaux de sujets de données permet de regrouper des éléments de données, quand bien même tous les membres d'un type de données ne peuvent être décrits de manière suffisamment précise par les mêmes informations. Prenons, par ex., les parents et les apprenants d'une école primaire. Il s'agit ici d'individus. Ils partagent donc certaines propriétés et caractéristiques telles que, par ex., une date d'anniversaire ou une taille corporelle enregistrées. Par ailleurs, des informations qui porteraient sur le traitement régulier des devoirs à domicile seraient peu judicieuses lorsqu'il s'agit de décrire les parents, alors que des informations sur le secteur économique de l'employeur n'apporteraient aucune donnée pertinente lorsqu'il s'agit de décrire des enseignants de l'école primaire. L'espace de données généré par l'intersection de ces deux types de sujets peut être représenté de manière simplifiée sous la forme d'un diagramme de Venn comme suit :



Graphique 5 : Données selon les sujets de données (selon CEDS 2017).

Les sujets de ce modèle de données sont reliés à d'autres sujets, resp. sont reliés les uns aux autres par le biais de relations. Ces relations permettent l'attribution mutuelle entre les sujets (par ex. l'énoncé selon lequel deux enfants sont les élèves d'une classe déterminée dans un établissement scolaire déterminé qui est dirigé par un enseignant identifiable avec précision), ce qui permet ainsi de reproduire les rapports sociaux et hiérarchiques (par ex. l'individu A est le supérieur hiérarchique de l'individu B, le département de l'ins-

4 Voir sous : <https://ceds.ed.gov/dataModelNDS.aspx>

5 Au sujet de la « normalisation » : [https://fr.wikipedia.org/wiki/Forme_normale_\(bases_de_donn%C3%A9es_relationnelles\)](https://fr.wikipedia.org/wiki/Forme_normale_(bases_de_donn%C3%A9es_relationnelles))

truction publique est compétent pour impartir une instruction à la direction de l'établissement scolaire), tout comme la position des individus au sein de l'organigramme des organisations (rôles, appartenances à certaines classes ou institutions, etc.). Les données qui rendent possibles ces mises en liens entre les sujets de données et qui, par conséquent, décrivent des relations, sont désignées par le terme de **données relationnelles**. En outre, dans la plupart des contextes de recensements et d'enregistrements de données, on génère une série d'attributs qui mettent à disposition les informations sur des propriétés, des comportements et des prestations des différents sujets de données :

- **Les données de propriétés** fournissent des renseignements sur la situation démographique, socio-économique, physique, psychologique et politique d'une personne physique ou d'une organisation. Pour les individus, elles permettent une description complète de la situation biologique et sociétale. En voici des exemples : l'âge ou la religion d'un élève, le revenu d'un enseignant, ou l'opinion politique d'une employée de l'administration. Pour les organisations, elles décrivent la situation géographique et les circonstances matérielles ainsi que le cadre juridique. En font partie, par exemple, la forme juridique de l'organisme responsable, l'équipement en matériel d'enseignement, les informations sur le budget, ou la structure d'un support de cours.
- **Les données de comportement** informent sur les activités et les conditions sous lesquelles ces activités sont observées. Elles permettent de décrire des modèles récurrents de comportement et d'en déduire, moyennant certaines conditions, l'existence de certains processus (mentaux) et de certaines capacités. En font partie, par ex., au niveau des individus, des informations sur les absences ou sur l'organisation et la participation à des activités extrascolaire, mais aussi des traces d'utilisation de matériel d'apprentissage telles que des temps d'accès à certaines applications d'apprentissage ou des visites à la bibliothèque scolaire. Au niveau des organisations, outre des fixations d'objectifs, elles incluent surtout des pratiques et des modes de comportement qui résultent de manière émergente de l'interaction entre certains membres, par ex. des formes de processus de prise de décision ou l'ampleur de l'interaction et de la collaboration collégiale.
- **Les données de prestations** décrivent les résultats de processus d'appréciation et d'évaluation des prestations. Elles servent en priorité à évaluer le niveau de compétence d'une personne physique ou d'une organisation et – en comparant l'évolution au fil du temps – à évaluer les changements. En font partie également, au niveau des individus, outre les notes scolaires classiques, les auto-évaluations, la participation à des mesures d'encouragement spécifiques ou le nombre de leçons d'apprentissage terminées par unité de temps pour une ressource d'apprentissage numérique. En font partie, dans une mesure particulière, au niveau de l'organisation, les moyens dépensés pour atteindre les objectifs et la relation entre les objectifs atteints et les ressources utilisées à cet effet. Par exemple, le rapport sur l'éducation du Centre suisse de coordination pour la recherche en éducation (CSRE) mesure les prestations (de certaines parties) du système éducatif suisse (Wolter et al. 2018) au moyen d'indicateurs d'effectivité, d'équité et d'efficience.

Comme dans le cas des sujets de données, cette catégorisation d'attributs est grossière, peu précise et potentiellement incomplète. Ainsi, les évaluations des prestations des collaborateurs ou des élèves se basent, en partie, sur leur comportement et sur les conditions sous lesquelles un comportement peut être observé, et portent également sur des propriétés du sujet de données. En outre, l'ampleur de concrétisation de nombreux attributs n'est pas fixe, comme supposé implicitement par une catégorisation statique, mais elle change au contraire au fil du temps. Une telle structuration ne convient donc que de manière limitée à la standardisation des ensembles de données, resp. des banques de données. Dans le cadre du présent rapport, cette catégorisation sert en priorité à visualiser concrètement l'étendue des données collectées dans les différents contextes où elles ont pris naissance.

La distinction grossière entre les types de données sur la base du sujet qui est décrit par lesdites données se rencontre aussi de manière similaire dans la pratique administrative et juridique. Toutefois, dans ce cas, on distingue surtout entre les **données personnelles** et les **données matérielles** : sont désignées en tant que données personnelles⁶ toutes les informations qui se rapportent à une personne physique ou morale déterminée ou seulement identifiables (par 3, al. 2, LPD). Leur collecte, leur traitement et leur mise à dispo-

6 On parle parfois également – ce qui n'est toutefois pas cohérent avec la terminologie de la LPD – de « données à caractère personnel ».

sition sont soumis à la législation sur la protection des données et, partant, à des restrictions particulières. Ce faisant, il s'agit d'accorder une importance particulière aux « **données sensibles** », désignées également par l'expression « **données particulièrement dignes de protection** » puisqu'en raison de leur importance, il existe un risque particulier d'infraction au droit de la protection de la personnalité lié aux possibilités de leur traitement et de leur croisement avec d'autres données (par. 3, al. 4, let. a, de la loi cantonale zurichoise sur l'information et la protection des données, IDG ZH). Il s'agit, par exemple, des informations sur les opinions philosophiques ou sur l'état de santé d'un apprenant. Relèvent en outre des collectes de données personnelles les déductions qui permettent d'identifier des aspects essentiels de la personnalité des personnes physiques, à savoir des informations pouvant servir à établir ce qu'on appelle des **profils de personnalité** (par. 3 [d], LPD). Dans certaines circonstances, ces dernières peuvent être générées ou déduites, par exemple, lorsque des données personnelles peuvent être combinées avec des métadonnées axées sur le comportement (par ex. le temps passé sur une page web ou les mouvements du curseur effectués sur cette dernière).⁷

L'Office fédéral de la statistique opère donc une différenciation en quatre niveaux de protection différents selon le niveau de **sensibilité des données** (OFS 2006). Relèvent du niveau 0 (données matérielles) toutes les données axées sur des personnes qui ne posent aucun problème, comme par exemple des données météorologiques. Le niveau 1 (données personnelles simples) comprend les données dont ne résulte, à elles seules, aucun potentiel de risque pertinent pour la protection de la personnalité. En font partie, par ex., des informations démographiques telles que l'âge ou le sexe d'une personne physique. Le niveau 2 (données personnelles qualifiées) décrit les données dont résulte déjà, à elles seules, un certain risque de mise en danger de la personne concernée. Relèvent explicitement de cette catégorie les données sur le niveau de formation et sur l'acquisition de ce dernier. Enfin, le niveau 3 comprend toutes les données personnelles qui ont été définies comme sensibles ou particulièrement dignes de protection par le législateur. En font partie, outre des attributs observables (par ex. des données sur les prestations comme les notes scolaires et les évaluations, ainsi que des données sur le comportement comme les informations sur les absences et sur les décisions en matière de consommation) provenant des attributs déductibles de ces observations (par ex. caractéristiques de personnalité et risque de changement d'affectation ou de transfert d'un collaborateur). Le but poursuivi par les restrictions liées à ces données relatives au traitement de données consiste à protéger les personnes sur lesquelles on collecte des informations privées et sensibles du risque d'identification et des inconvénients éventuels qui en résultent, et de garantir ainsi leurs droits fondamentaux et leurs droits de protection de la personnalité.⁸

Par conséquent, des données personnelles pour lesquelles on a durablement éliminé la référence à une personne et qui ne permettent plus une identification, ou seulement au prix d'une charge de travail disproportionnée, ne sont plus des données à caractère personnel. Dans la pratique, on désigne souvent cela par l'expression « **anonymisation de fait** » ou « **pseudonymisation** » de données. Ce faisant, on remplace des attributs clairement attribuables à une personne, par ex. le nom, l'adresse ou le numéro AVS par une variable d'identification (alpha)numérique.⁹ En outre, les données pouvant être mises en relation avec des personnes, à savoir des attributs qui, certes, ne permettent pas d'identifier clairement ou immédiatement une personne déterminée, mais qui permettent de constater l'identité d'une personne à l'aide d'autres informations, sont manipulées de telle sorte qu'un risque d'identification devient limité. Cela concerne surtout des informations de nature géographique (comme le nom de l'école ou le numéro de commune) ou des valeurs exceptionnelles d'autres attributs (par ex. un revenu très élevé ou très faible, une langue rare parlée à la maison, etc.) (Höhne 2010). Toutefois, l'augmentation constante du volume de données personnelles issues d'autres sources (provenant par ex. des réseaux sociaux) et l'existence de technologies toujours plus performantes réduisent la charge de travail requise pour la réidentification probabiliste d'individus. Ainsi, la distinction entre les données à caractère critique pouvant être attribuées à des personnes et les don-

7 Les métadonnées se rapportant au comportement sont parfois résumées par le concept de « paradonnées » dans la littérature résumant les études recensées dans ce domaine (Gregory, Heus, & Ryssevik 2009).

8 Compte tenu du fait qu'un volume non négligeable de données concerne des personnes mineures dans le secteur de l'éducation, une prudence particulière est de mise à cet égard.

9 En règle générale, les attributs se rapportant directement à des personnes sont archivés dans un fichier séparé qui autorise la « réidentification » des personnes impliquées à l'aide d'un autre fichier-clé. Ces fichiers d'aide doivent être conservés de manière physiquement séparée des données personnelles anonymisées générées (Häder 2009). Si elles sont détruites de façon irréversible et que cela rend ainsi impossible une identification exempte d'ambiguïté des personnes impliquées, on parle dans ce cas d'« anonymisation absolue ».

nées à caractère non critique attribuables à des personnes s'estompe de plus en plus, ce qui nécessite des mesures supplémentaires de protection des données. Les méthodes usuelles de manipulation de ces données, comme par ex. l'agrégation de données à un niveau d'observation plus élevé, ne sont pas non plus toujours appropriées pour garantir l'anonymisation requise à tous les niveaux du système éducatif. Par exemple, des moyennes de classe ne fournissent certes aucune information sur un apprenant spécifique, mais contiennent très certainement des informations sur l'enseignant (Rudin 2014, p. 34). La distinction entre données personnelles et données anonymisées n'est donc pas toujours précise, et elle nécessite un examen dans chaque cas individuel (voir chapitre 4.1.3.2, p. 51).

Une autre classification des types de données pertinentes pour la pratique concerne la concession de licences, resp. la limitation de l'accès aux données à des fins d'utilisation générale. Pour décrire le degré de limitation d'accès aux données, l'Institut pour les données ouvertes (Open Data Institute) a développé le concept de « spectre de données » (Open Data Institute 2018). Selon ce système de classification, toutes les données peuvent être qualifiées de manière différente tout au long d'un éventail continu qui va des données ouvertes aux données propriétaires. Si des données sont mises à disposition sans restriction aux fins d'utilisation libre, de transmission ultérieure et de réutilisation, on parle dans ce cas de données ouvertes (également : **Open Data**). Les données propriétaires (également : **Closed Data**) sont à l'opposé de cet éventail. Elles sont exclusivement accessibles à leur propriétaire. Entre ces deux pôles figurent trois catégories de données semi-ouvertes (appelées **Shared Data** ou données partagées) qui se distinguent selon leur degré de limitation d'utilisation : il s'agit des données **Named-access** qui ne sont accessibles qu'à des personnes ou organisations dûment nommées. C'est la catégorie d'accès la plus restrictive. En revanche, les données **Group-based-access** sont accessibles à certains groupes de personnes conformément à des critères spécifiques (par exemple à des chercheurs travaillant dans des hautes écoles dûment enregistrées). Enfin, les données **Public-access** constituent la catégorie la plus ouverte parmi les **Shared Data**. Elles sont en principe accessibles à tous. Mais leur utilisation est liée au respect de certaines conditions préalables. Pour les données non sensibles de l'administration publique, le Conseil fédéral a pris clairement position en faveur du principe de l'ouverture des données « dont la percée [devrait] être favorisée dans l'ensemble de l'administration, dans la recherche et, en partie aussi, dans l'économie privée » (Stratégie « Suisse numérique » 2016, 4.2.2)¹⁰. Il reste à clarifier dans quelle mesure il faut aussi viser la mise en œuvre de ce principe dans le système éducatif suisse. Il semble important que des discussions de ce genre soient menées au plan national et que, ce faisant, l'on tienne compte de la diversité des différents acteurs qui interagissent dans le système éducatif ainsi que de leurs intérêts respectifs (voir chapitre 3.4).

¹⁰ Dans la version actuelle de la stratégie « Suisse numérique » de septembre 2018, sous le point 4.7.3, l'étude du principe de l'Open Data est poursuivie.

3 Écosystème des données de l'éducation

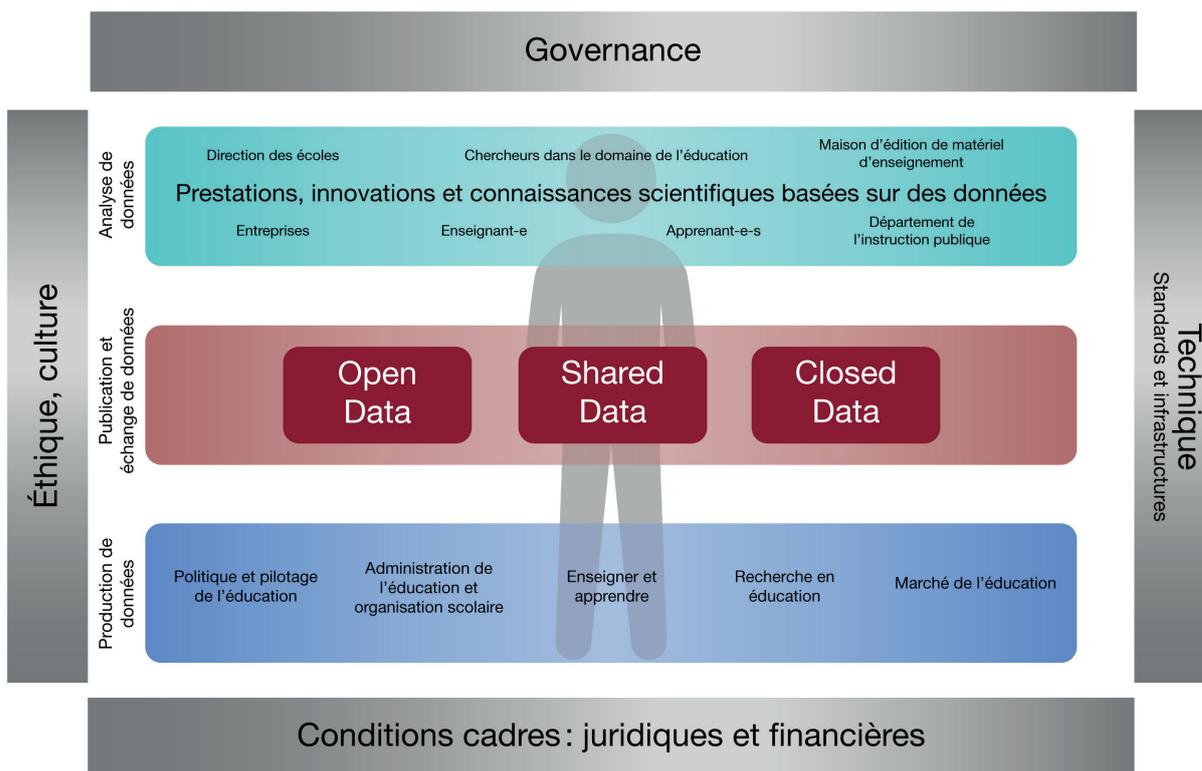
« La question des données n'est pas encore clarifiée... c'est trop complexe... nous ne sommes pas encore aussi avancés », « Quelles questions devrions-nous/devons-nous nous poser ? » et « Dans quelle direction pouvons-nous réfléchir à l'avenir ? ». Des déclarations de ce genre montrent clairement que la thématique des données et de leur utilisation soulève des questions et des incertitudes, voire qu'elles sont en partie liées à un certain malaise. À cet égard, il semble en particulier que le concept d'utilisation des données a une connotation négative : dans ce contexte, quelqu'un utilise quelque chose qui appartient à quelqu'un d'autre – et même souvent sans lui en demander la permission. Est-il autorisé à le faire ? Et quelles conséquences cela pourrait avoir pour celui dont les données ont été volées ?

Cette attitude généralement incertaine s'explique à maints égards par les intrications propres à cette thématique – qui est en outre très fortement mise en relation avec la technique, voire qui est réduite à cette dernière. Elle rend plus difficile une compréhension rapide et donne ainsi le sentiment qu'en tant que non-initié, il est impossible d'en saisir toute la complexité. Conséquence : il arrive souvent que des points de vue très étroits et simplificateurs soient défendus par certains (par ex. : l'utilisation des données, cela veut dire qu'il faut respecter la protection des données).

Le chapitre suivant a pour but de décrire de manière compréhensible au plus large public possible le concept d'utilisation des données ainsi que les conditions cadres qui la réglementent. Un modèle simplifié de l'« écosystème des données de l'éducation » nous servira de guide à cet effet.

3.1 Aperçu d'un modèle

Le modèle de l'« écosystème des données de l'éducation » (graphique 1) est axé, dans les grandes lignes, sur le cycle de vie usuel des données : les aspects centraux qui sont considérés comme des dimensions importantes du concept d'utilisation des données résident dans la production des données, la publication des données, l'échange des données ainsi que dans l'analyse des données. Ces aspects sont entourés par les conditions cadres juridiques et financières (en bas du graphique), les infrastructures techniques et les standards (à droite), les principes culturels et éthiques (à gauche) ainsi que la gouvernance en charge de la régulation (en haut). Les fondements stratégiques, qui constituent le cadre le plus extérieur, sont les critères d'orientation de référence. Enfin, le sujet de données en tant qu'individu autodéterminant se trouve au centre du graphique et fait simultanément partie d'entités plus larges (par ex. classe d'école, établissement scolaire, association sportive, commune, canton, système éducatif).



Graphique 1 : Écosystème des données de l'éducation (selon Gollietz 2018)

3.2 Aspects déterminants de l'utilisation des données

Tel qu'il est défini dans le présent rapport, le concept d'utilisation des données est composé de trois étapes clés. Sur la base de la hiérarchie des connaissances qui a été introduite précédemment (voir Graphique 2), l'utilisation des données commence avec la traduction de signes en données, puis se poursuit avec leur transformation au moyen de procédures techniques de traitement de l'information (production de données), et se termine avec l'évaluation de ces données et l'exploitation de ses résultats (analyse des données). En outre, compte tenu du nombre important d'acteurs hétérogènes dans le système éducatif l'échange de données entre les différents acteurs joue un rôle important (publication et échange de données).

3.2.1 Production de données

La production de données se rapporte à la collecte, l'enregistrement, la consolidation et la mémorisation de données. Cet aspect inclut, d'une part, tous les processus où des informations sous forme de signes sont collectées et regroupées et sont traduites en données au moyen de règles d'opérationnalisation. Cela pré-suppose notamment des décisions à prendre sur qui doit collecter, quand et comment quels types d'informations. Les données ainsi saisies seront ensuite consolidées et mises au net et, le cas échéant, seront transformées, c.-à-d. modifiées et/ou croisées avec d'autres données. Enfin, dans le cadre de la production de données, des décisions seront prises sur la forme de détention et de dépôt des données ainsi collectées et consolidées. Cela concerne la forme de la conservation des données (par ex. avec quel logiciel et dans quelle structure les données devront-elles être mémorisées ?), archivage (par ex. règles applicables à la migration dans de nouveaux systèmes ou sur de nouveaux supports) et, le cas échéant, effacement des données. Lors de la production de données, il s'agit d'élaborer, de traiter et d'obtenir des données sans qu'il ne soit possible de déduire à nouveau des informations à partir des données elles-mêmes.

3.2.2 Publication et échange de données

La publication de données décrit tous les processus et décisions qui concernent l'utilisation ultérieure et la transmission d'ensembles de données existants. Cet aspect comprend, d'une part, des décisions d'ordre général portant sur la question de savoir si oui ou non, et si oui, avec quelles limitations (*open, shared, closed*) des données seront transmises (par ex. la question de savoir s'il est impératif d'anonymiser des données personnelles). Mais elle se rapporte également à des problématiques pratiques relatives au formage, à la transformation et à la structuration des données mises à disposition (par ex. la question de savoir comment l'anonymisation doit être techniquement mise en œuvre). À cet égard, la capacité visée de raccordement des données, à savoir la question de savoir si oui ou non, et si oui, comment les données pourront être reliées à d'autres sources de données joue un rôle important.

3.2.3 Analyse des données

L'analyse des données décrit l'utilisation des ensembles de données existants afin de répondre à des problématiques concrètes. En règle générale, ces données sont conçues dans le but de soutenir et d'améliorer les processus de prise de décision. Cet aspect inclut donc tous les processus et décisions qui concernent la sélection de données appropriées, de méthodes statistiques et d'outils techniques de traitement de l'information adéquats. Mais il contient aussi l'interprétation et la classification des connaissances qui ont été obtenues par les analyses.

Pour qu'une utilisation des données s'effectue de manière judicieuse, toutes ces trois dimensions sont importantes : les analyses de données présupposent l'existence d'une collaboration et d'une interaction optimales entre la production de données et la publication et l'échange de données. Car ce n'est qu'ainsi que l'on peut s'attendre à des données d'un haut niveau de qualité et, sur cette base, à des résultats qualitatifs dûment validés de l'utilisation des données.

3.3 Cadre

3.3.1 Conditions cadres juridiques et financières

L'utilisation des données dans le système éducatif est réglementée par différentes lois en fonction du contexte, du niveau et des acteurs impliqués. Ces règles peuvent encore être précisées dans les ordonnances et les directives, resp. dans les lignes directrices y afférentes (voir Graphique 6).

En principe, la *Constitution fédérale suisse* confère à toute personne un droit fondamental « d'être protégée contre l'emploi abusif des données qui la concernent »¹¹. L'individu reçoit ainsi un droit d'autodétermination informationnelle, à savoir le droit de déterminer lui-même si oui ou non il entend divulguer à qui et à quel moment des situations objectives, des pensées, des émotions ou des sentiments personnels de sa vie. Partant, la protection des données ne protège pas les données en tant que telles, mais au contraire les droits fondamentaux des personnes. Les *constitutions cantonales* prévoient une garantie correspondante.

La *loi fédérale sur la protection des données*¹² s'adresse à l'Administration fédérale ainsi qu'à toutes les personnes privées qui traitent les données personnelles. Les *lois cantonales sur la protection des données* réglementent le traitement des données effectué par les administrations cantonales, dont font partie également le corps enseignant et les directions des écoles en tant que membres des autorités. Les collaboratrices et collaborateurs d'établissements publics sont en outre soumis au *secret de fonction*. De plus, il faut tenir compte de la *loi fédérale sur le principe de la transparence dans l'administration*¹³ ainsi que des possibilités résultant du principe de *l'entraide administrative*¹⁴ qui autorisent l'échange de données.

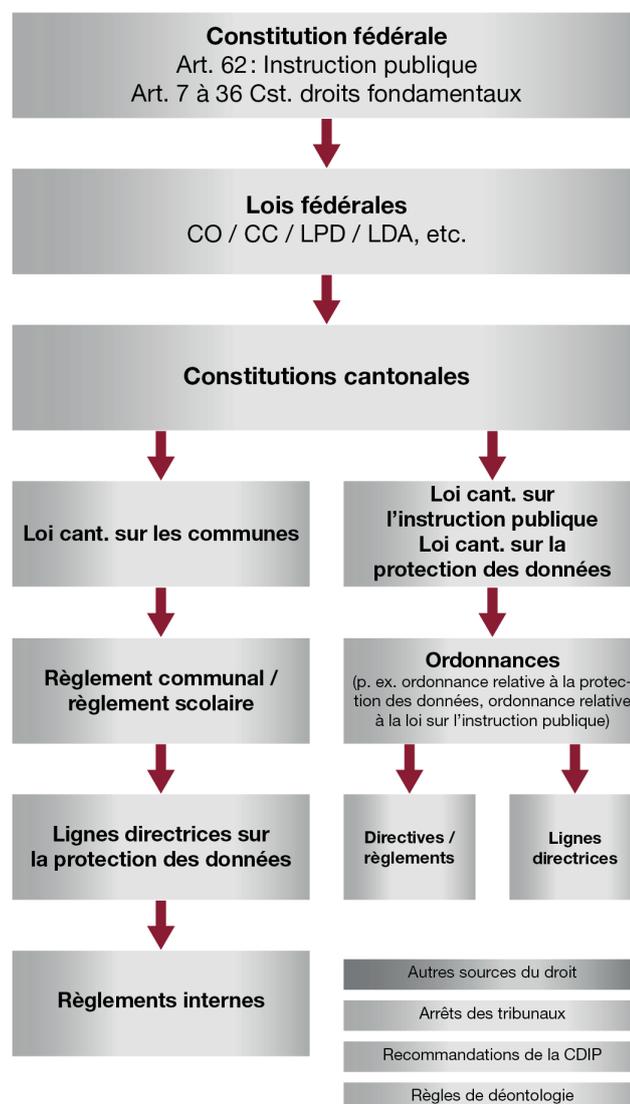
11 Art. 13, al. 2, Cst.

12 Loi fédérale sur la protection des données du 19 juin 1992, état au 1^{er} mars 2019 (LPD), RS 235.1.

13 Loi fédérale sur le principe de la transparence dans l'administration du 17 décembre 2004, état au 19 août 2014 (LTrans), RS 152.3.

14 Art. 44, al. 2, Cst. : « La Confédération et les cantons s'accordent réciproquement l'entraide administrative et l'entraide judiciaire. » L'entraide administrative est fournie sur demande et au cas par cas ; elle est accordée si des intérêts publics ou privés prépondérants ne s'y opposent pas. L'entraide administrative est limitée en particulier par le secret de fonction et par la protection des données.

Les défis à relever dans ce domaine résident dans les différences cantonales¹⁵, des marges d'interprétation, des lacunes juridiques et/ou des zones grises découlent de l'utilisation croissante des appareils numériques (utilisation mobile) et des applications (du Big Data) (voir chapitre 4.1 et chapitre 6.4.3, p. 127).



Graphique 6 : Sources générales du droit et hiérarchie des niveaux du droit scolaire suisse (Marris 2016)¹⁶

Dans le cadre du droit actuel en vigueur, les conditions cadres financières restent mises entre parenthèses. Elles constituent pourtant un facteur pertinent qu'il faut intégrer et prendre en compte dans l'élaboration d'une politique d'utilisation des données cohérente et lors de la fixation de ses objectifs ainsi que de ses ressources et moyens. Le présent rapport se limite toutefois à traiter de la préparation technique des bases préalables à un tel processus et, pour cette raison, renonce à se livrer à des réflexions sur les coûts. Il est cependant évident que l'encouragement à une utilisation dûment réglementée des données dans le système éducatif suisse génère des dépenses, et que, pour cette raison, ce projet ne pourra pas être mis en œuvre gratuitement. Mais il est tout aussi évident qu'une utilisation dûment réglementée des données promet de générer des gains d'efficacité et d'effectivité. En outre, si l'on considère le système éducatif dans son ensemble, on constate qu'il gagnera en stabilité dès lors que de nombreux conflits possibles pourront

¹⁵ Dans le cadre du présent rapport, le Centre d'information et de documentation IDES de la CDIP a réalisé une première étude comparative des lois cantonales sur la formation et des ordonnances y afférentes (y compris les lois sur l'école secondaire du deuxième degré et les lois sur la formation professionnelle). Cette étude a été mise à disposition de l'ITSL pour ses articles spécialisés.

¹⁶ Une présentation concrète de la hiérarchie des niveaux figure à l'Annexe I.

être évités, et qu'on pourra ainsi réduire au maximum de coûteux liens de dépendance envers des prestataires de services de l'économie privée exerçant leurs activités dans ce domaine.

3.3.2 Infrastructures techniques et standards

Si l'on prévoit d'utiliser à grande échelle des données numériques dans le système éducatif, il est indispensable de disposer d'une infrastructure technique qui rend possible une telle utilisation. Cela implique qu'il faut doter les institutions d'un équipement complet de terminaux numériques, et qu'il faut mettre à leur disposition des capacités (de serveurs) suffisantes nécessaires à l'évaluation rapide des données ainsi collectées, au développement ou à l'acquisition de programmes correspondants.

Des standards techniques constituent un outil de travail d'importance décisive doublement nécessaire : d'une part, afin de garantir la sécurité informatique au sens de l'IT security et, d'autre part, afin de soutenir l'échange et l'utilisation des données pour assurer l'interopérabilité des données. Même si nous ne disposons d'aucune vue d'ensemble fiable sur les standards actuellement pris en compte en Suisse, nous pouvons partir du principe que la pratique de la sécurité informatique et la certification des produits et systèmes informatiques sont axées sur les standards généralement répandus en la matière (par ex. normes ISO/IEC-27000 et ISO/IEC-15408). En revanche, les standards liés à des modèles de données et à l'échange et aux modèles de données sont moins répandus (voir chapitre 4.2 et chapitre 4.4).

3.3.3 Éthique et culture

En raison de l'utilisation croissante de grands volumes de données – notamment dans le contexte de l'intelligence artificielle (IA), de l'Internet des Objets (IoT) et du Big Data – il en résulte non seulement des problématiques sociétales et juridiques, mais encore des problématiques éthiques. À l'avenir, qui devrait récolter quelle quantité de données provenant de qui et dans quel but, et qui devrait être autorisé à les traiter ? Comment devrait se présenter à l'avenir la « culture d'entreprise » en matière de traitement de données¹⁷ dans le système éducatif suisse ? Par quels principes éthiques cette culture d'entreprise devrait-elle être guidée ?¹⁸

3.3.4 Gouvernance

Les structures de gouvernance des données, en tant que dernier élément cadre d'importance cruciale pour l'écosystème des données sont le thème cible qu'il s'agit de discuter dans le présent rapport. Elles définissent et opérationnalisent l'utilisation effective des données de l'éducation :

- en clarifiant les rôles et les responsabilités,
- en fixant des lignes directrices et des standards,
- en définissant des processus et des méthodes,
- en rendant les personnes et les organisations capables de les maîtriser,
- en créant la confiance et la sécurité,
- en prévoyant un système de gestion des risques.

3.3.5 Bases stratégiques

Des documents stratégiques de base sur l'utilisation des données existent dans différents domaines politiques et administratifs et à différents niveaux. La compilation suivante de ces documents contient des bases stratégiques qui sont pertinentes au niveau national. Elles sont brièvement présentées sous l'angle de la contribution qu'elles apportent à la clarification du concept d'utilisation des données dans le système éducatif suisse.

¹⁷ Une gestion compétente des données fait aussi partie de cette culture d'entreprise requise en matière de traitement des données, ce qui nécessite des compétences correspondantes ainsi qu'une acceptation générale de l'utilisation des données par les acteurs concernés.

¹⁸ Dans le domaine de la recherche médicale, on est déjà sensibilisé à ce type de problématiques (voir par ex. ASSM 2019 ; Hauser et al. 2017).

- **Conseil fédéral (2018) : stratégie « Suisse numérique »**

La stratégie faîtière du Conseil fédéral relative au recours à la numérisation ne traite pas directement les défis à relever qui sont en rapport avec l'utilisation des données dans le système éducatif suisse. Toutefois, dans le champ d'action « Formation, recherche et innovation », il est stipulé, pour le domaine de la scolarité obligatoire et du niveau secondaire II, qu'il faut rendre la population suisse capable d'utiliser activement les technologies de l'information et de la communication (TIC) « pour participer de manière compétente à des processus numériques politiques, sociaux, économiques et culturels ». Les lignes directrices sur la gestion des données et sur l'utilisation des données sont définies de manière séparée de ce qui précède dans le champ d'action « Données, contenus numériques et intelligence artificielle ». Il contient notamment des objectifs qui sont en lien avec les droits en matière de données, les droits d'accès aux données et les droits d'utilisation des données ainsi que sur le recours à l'intelligence artificielle.

- **CDIP et SEFRI (2016) : coordination des stratégies de la Confédération et des cantons relatives à l'intégration des TIC dans le domaine de la formation**

La coordination des stratégies de la Confédération et des cantons relatives à l'intégration des TIC dans le domaine de la formation repose sur la collaboration et le partenariat entre la Confédération et les cantons dans le domaine de l'éducation et de la formation. La base de ce partenariat réside dans la loi sur la coopération dans l'espace suisse de formation (LCESF) de 2016 et dans la convention de coopération qui en fait partie. Dans le domaine de la numérisation du système éducatif suisse et des changements qui en découlent, elle est mise en œuvre par le comité de coordination de la numérisation de l'éducation (CC N). Il a pour mandat de « créer continuellement les conditions préalables à une politique cohérente à l'échelle de toute la Suisse afin de structurer la mutation numérique dans le système éducatif suisse ». Font aussi partie de ses tâches le traitement des questions ayant trait à une politique commune d'utilisation des données dans le système éducatif suisse. Le CC N joue ainsi un rôle clé dans l'élaboration de cette politique : il constitue la plateforme d'échange d'informations et de discussion des travaux qu'entreprennent, chacun pour eux-mêmes, la Confédération et les cantons. C'est par le biais du CC N que ces travaux doivent être harmonisés entre eux du point de vue de leur contenu, de leurs structures et de leurs processus afin qu'ils puissent être regroupés ensemble en un seul recueil de réglementation applicable aux deux partenaires de formation dont les domaines de tâches et de compétences seront certes harmonisés, mais dont le contenu doit être conjointement défendu par ces deux partenaires.

- **SEFRI (2017) : plan d'action pour le domaine de la formation, de la recherche et de l'innovation durant les années 2019-2020**

Le *Plan d'action pour le domaine de la formation, de la recherche et de l'innovation durant les années 2019-2020* a été élaboré par le SEFRI en tant que partie du rapport intitulé *Défis de la numérisation pour la formation et la recherche en Suisse* (SEFRI 2017b). Les mesures prévues dans le domaine « Formation » de ce plan d'action mettent l'accent sur les niveaux de formation de la scolarité obligatoire et du niveau secondaire II. Elles sont axées sur une plus forte intégration des TIC dans le système éducatif suisse ; ce qui est mis au premier plan, c'est l'amélioration des compétences TIC des élèves et des personnes qui ont terminé leurs études, la mise en place d'infrastructures adéquates dans le système éducatif suisse, la professionnalisation et le développement des compétences pour le corps enseignant, et le renforcement de la coordination et de la communication entre la Confédération et les cantons dans le cadre de la collaboration en matière d'éducation. La tendance a été de laisser de côté l'aspect qui touche à l'utilisation des données. Ces mesures ont toutefois pour but d'aider à clarifier des questions relatives à l'utilisation des données parce qu'elles esquissent de futurs domaines d'application et mettent en évidence les objectifs qu'elles ont contribué à atteindre.

- **Conseil fédéral (2018) : mesures dans le domaine des données, de l'utilisation des données et de l'intelligence artificielle (stratégie « Suisse numérique »)**

Les mesures prévues dans le champ d'action « Données, contenus numériques et intelligence artificielle » sont globalement appropriées pour contribuer à la clarification du concept d'utilisation des données dans le système éducatif suisse. On peut s'attendre, d'une part, à disposer de bases de portée générale (par ex. conditions cadres juridiques) et, d'autre part, à étudier des possibilités concrètes d'utilisation des données dans un secteur éloigné de celui de la formation qu'on pourra ensuite adapter au système éducatif dans le sens de bonnes pratiques, et qui pourront alors être reprises en conséquence. En font partie notamment les travaux de révision relatifs à la loi fédérale sur la protection des données

(LPD), le Programme national de recherche « Big Data », la séparation entre la mémorisation et l'utilisation de données statistiques, le projet pilote relatif à l'utilisation multiple de données de base, les projets de grande ampleur d'utilisation des données dans la statistique publique, l'« initiative sur les données » ainsi que les résultats du groupe de travail « Intelligence artificielle ».

▪ **CDIP (2018b) : Stratégie sur la numérisation**

Avec la Stratégie sur la numérisation de la CDIP, les cantons veulent « jouer un rôle actif dans la transition numérique, en tirer le meilleur profit pour les processus éducatifs et l'organisation de l'éducation et contribuer à maintenir la compétitivité de notre pays ». Cette stratégie contient des objectifs qui concernent sept domaines différents. Il s'agit d'objectifs que les cantons veulent atteindre soit par des mesures propres, soit par des mesures communes, soit par des mesures soutenues par la CDIP. Le concept d'utilisation des données est un fil rouge qui traverse l'ensemble de la Stratégie sur la numérisation. Le premier objectif est explicitement consacré à l'utilisation des données :

1. Utilisation des données du système éducatif : protection, sécurité et confiance

Objectif stratégique : clarifier la gestion des données du système éducatif, leur utilisation et les conditions requises pour garantir la sécurité de l'information (confidentialité, disponibilité et intégrité).

Objectifs opérationnels :

1.1 La Suisse s'est dotée au niveau national de lignes directrices relatives à l'utilisation des données du système éducatif et à la sécurité de l'information.

1.2 Les possibilités d'utilisation des données du système éducatif sont pleinement exploitées.

1.3 Il existe une identité numérique reconnue dans toute la Suisse pour chaque élève et pour chaque personne en formation ou travaillant au sein des établissements de formation. Cette identité numérique apporte une contribution majeure à la sécurisation de l'utilisation des données du système éducatif.

1.4 Tous les élèves et toutes les personnes en formation ou travaillant au sein des établissements de formation ont accès aux données générées en lien avec leur propre identité numérique et peuvent voir l'utilisation qui en est faite.

1.5 Il existe une institution mandatée par les autorités cantonales qui s'occupe de l'utilisation des données du système éducatif et de la sécurité de l'information.

D'autres objectifs propres à la Stratégie sur la numérisation de la CDIP permettent de répondre aux défis à relever en lien avec l'utilisation des données dans le système éducatif. Il y a lieu de mentionner notamment :

- L'extension du monitoring national de l'éducation aux questions qui ont trait à la numérisation
- L'échange de données et d'informations entre les établissements scolaires et avec les différents groupes cibles du système éducatif
- Le recours aux TIC dans l'enseignement, et le fait de tirer parti du potentiel des moyens adaptatifs d'enseignement et d'apprentissage
- Le renforcement des compétences numériques dans les directions des écoles, au sein du corps enseignant et chez les élèves
- L'accès des chercheurs aux données d'application et à leur exploitation en vue de l'innovation (pédagogique)

Dans l'ensemble, l'existence de ces objectifs présuppose que le concept d'utilisation des données a donc été clarifié à titre de premier objectif, par ex. parce que ces objectifs se fondent sur le fait que des données sont disponibles, resp. sur le fait que ces données sont générées conformément à leur but d'utilisation et dans le respect des conditions cadres juridiques y afférentes.

- **Cantons : stratégies de numérisation dans le domaine de la formation**

La collection *Numérisation dans le système éducatif : concepts cantonaux* du Serveur suisse de documents pour l'éducation et la formation (edudoc) comprend 26 entrées (état : 20.03.2019).¹⁹ Le plus ancien document de concept a été publié en 2011, et le plus récent remonte à 2018. Ces documents nous guident dans les embranchements propres au système fédéraliste de l'éducation en Suisse, qui marque aussi de son empreinte la numérisation et la transformation numérique. Dans la mesure où des objectifs communs en lien avec l'utilisation des données sont saisissables, ces derniers sont reproduits dans la Stratégie sur la numérisation de la CDIP. C'est pourquoi nous renonçons ici à une analyse plus détaillée de cette thématique.

- **Conseil fédéral (2018a) : stratégie Open Government Data (OGD) de la Suisse**

Le premier objectif de la *Stratégie pour des données administratives ouvertes en Suisse 2019-2023* (ou stratégie OGD) consiste à rendre disponibles, pour le grand public, à titre de données ouvertes, de nouvelles collections de données ainsi que des collections de données existantes de l'Administration fédérale sur un portail web centralisé à l'adresse opendata.swiss aux fins d'utilisations diverses. Les mesures prévues à cet effet sont développées de manière globale et devraient contribuer à la percée de l'idée de l'Open Data, par ex. en organisant, au-delà des seuls collaborateurs de l'Administration fédérale, une « table ronde » entre les offreurs de données et les utilisateurs de données en renforçant de manière ciblée la compétence des intéressés en matière de gestion de données, et en incitant le grand public à utiliser les données.

- **CdC (2018) : lignes directrices des cantons relatives à l'administration numérique**

Les *lignes directrices des cantons relatives à l'administration numérique* de la Conférence des gouvernements cantonaux (CdC) constituent la compréhension commune des gouvernements cantonaux des thèmes de la numérisation et de l'e-government. Globalement, au-delà des différents champs d'action et pistes de réflexion individuelles définis dans ce document, elles visent l'application de la règle du « digital first » sous forme de principe. Elles ont pour but de provoquer une mutation fondamentale au sein des administrations pour se diriger vers l'utilisation de données numériques qui ne concerne pas uniquement le recours aux nouvelles technologies, mais qui prévoit au contraire la transformation des modalités d'organisation, des modes de travail, des modèles d'affaires et de l'économie des données.

Dans l'ensemble, il s'avère que les documents de base stratégiques précités fixent certes, chacun individuellement, des objectifs et principes importants pour l'utilisation des données dans le système éducatif, mais qu'ils ne peuvent être que difficilement regroupés en un cadre d'action directeur – au sens de conditions cadres concrètes – pour la pratique de l'utilisation des données dans le système éducatif parce qu'ils sont ancrés dans des contextes trop différents et parce qu'ils sont esquissés à partir de points de vue trop différents. S'agissant de la fixation des objectifs, il est certain que la Stratégie sur la numérisation de la CDIP constitue le point d'orientation pertinent en la matière. Elle est en effet légitimée par les tâches et compétences des cantons dans le système éducatif, et c'est la seule stratégie qui contienne des formulations substantielles d'objectifs pour l'utilisation des données dans le système éducatif. Ces formulations sont toutefois trop peu concrètes pour pouvoir déjà clarifier également les défis à relever propres à l'utilisation des données et les questions correspondantes qui sont soulevées par les autres objectifs de la stratégie de la CDIP. À cet effet, on peut et l'on doit s'inspirer de la stratégie « Suisse numérique » du Conseil fédéral et des clarifications sur les données, l'utilisation des données et l'intelligence artificielle visées dans ce document, de la stratégie OGD ainsi que des lignes directrices de la CdC. Il y a lieu de reprendre dans les documents précités les principes à observer lors de l'utilisation des données dans le système éducatif – et, par conséquent, il faut aussi en déduire les interactions mutuelles entre les objectifs de la stratégie de numérisation de la CDIP. En tant que cadre le plus extérieur de l'écosystème des données de l'éducation, ces bases stratégiques nécessitent encore une concrétisation. Le CC N assume un rôle central à cet effet. Il est le comité approprié pour s'attaquer de manière coordonnée aux questions qui se profilent. C'est à cette instance que les services compétents de la Confédération et des cantons doivent s'adresser pour recevoir des réponses adéquates à leurs questions.

¹⁹ La collection edudoc « Numérisation dans le système éducatif : concepts cantonaux » est accessible sous : <https://edudoc.ch/collection/digitalisierungskonzepte?ln=fr>

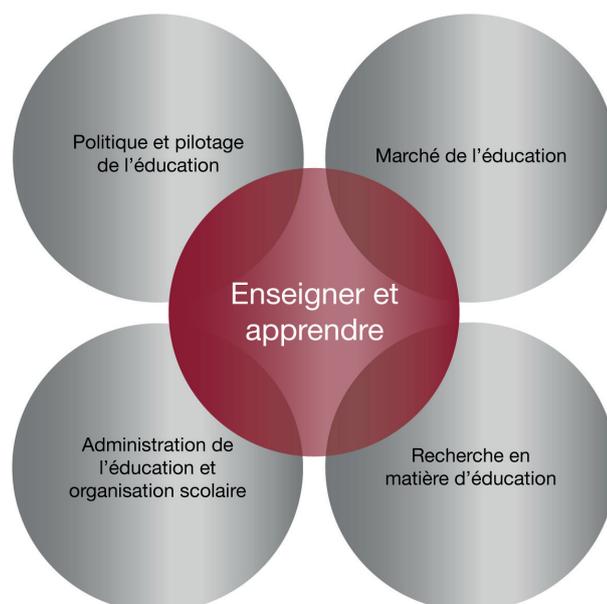
3.4 Contextes d'utilisation

En règle générale, les organisations et acteurs qui enregistrent et collectent les données de l'éducation ne considèrent pas qu'ils le font comme s'il s'agissait d'une fin en soi. Ces enregistrements servent plutôt à répondre à des problématiques concrètes ou à améliorer des applications concrètes. Très souvent, les contextes dans lesquels des ensembles de données sont collectés sont les mêmes que ceux dans lesquels ces données sont utilisées. Par exemple, des données sur le nombre d'élèves sont récoltées au sein de l'administration de l'éducation afin de lui permettre d'identifier ses besoins en matière de ressources et de se préparer à de probables changements. Dès lors, la production et l'utilisation des données de l'éducation sont étroitement enchevêtrées. Pour parvenir à comprendre l'étendue et la nature de la collecte de données dans le système éducatif, il est donc crucial de comprendre à quelles fins ces données sont enregistrées et collectées, et quelle importance il faut accorder aux éventuelles analyses résultant de ces données.

Dans les lignes qui suivent, nous étudions le positionnement dans le système éducatif ainsi que le rôle des acteurs clés, de leurs tâches et obligations, resp. de leurs intérêts respectifs qui déterminent les contextes d'utilisation décisifs des données du système éducatif. Dans ce contexte, il s'agit de mettre en évidence l'importance du rôle des données et de l'utilisation des données ainsi que les défis de la numérisation pour les différents domaines concernés. Cinq contextes d'utilisation différents font l'objet d'une description, à savoir :

- **La politique et le pilotage de l'éducation** décrivent l'utilisation de données provenant du secteur de la formation aux fins de planification, de contrôle et de pilotage du système éducatif par les institutions et commissions politiquement responsables. Dans ce domaine, les données servent surtout à soutenir des décisions stratégiques.
- **L'administration de l'éducation et l'organisation scolaire** se rapportent à l'utilisation de données dans le cadre de l'organisation et de l'administration du fonctionnement quotidien des établissements d'éducation et de formation de la Confédération et des cantons.
- **La recherche en éducation** décrit l'utilisation de données de l'éducation dans un contexte essentiellement scientifique en tant qu'outil d'enquête et d'explication de la réalité du déroulement des événements éducatifs.
- **Le marché de l'éducation** met en évidence l'utilisation et les objectifs d'application des données de l'éducation par des prestataires de technologies et services numériques.
- **L'enseignement et l'apprentissage** discutent l'utilisation de données de l'éducation dans le cadre de l'enseignement quotidien et du déroulement courant des événements d'apprentissage. Ce contexte constitue l'élément de liaison d'importance décisive entre les autres contextes.

Le graphique 7 illustre, de manière très simplifiée, selon leur orientation de principe, les contextes axés sur l'amélioration du déroulement des événements d'apprentissage et d'enseignement. Pour des raisons de lisibilité, nous renonçons notamment à représenter les autres liens d'interdépendance ainsi que les relations complexes de dépendances mutuelles. Tous les contextes où l'on fait usage des données de l'éducation ou des informations déduites de ces dernières ne sont pas non plus utilisés de la même façon que le font les médias ou la discussion publique dans le graphique figurant ci-dessous. Ce schéma sert ainsi à représenter de manière simplifiée les contextes d'utilisation les plus pertinents au sein du système éducatif.



Graphique 7 : Contextes d'utilisation

3.4.1 Politique et pilotage de l'éducation

Conformément à l'art. 62 de la Constitution fédérale, les cantons assument la responsabilité principale relative au système suisse d'instruction publique. La surveillance sur les établissements scolaires leur incombe, et ils prennent les décisions fondamentales y afférentes. Par conséquent, chaque canton dispose de ses propres prescriptions juridiques pour le domaine de l'éducation et de la formation. Toutefois, les 26 lois cantonales sur les écoles ou sur la formation reposent, pour l'essentiel, sur les mêmes bases légales, et sont axées, pour l'essentiel, sur des objectifs concordants.²⁰

Outre les cantons, les communes sont le deuxième acteur important de la politique et du pilotage de l'éducation. Différentes compétences leur sont dévolues par les cantons. Dans le domaine de la scolarité obligatoire, il s'agit en particulier de l'engagement des enseignants ainsi que de la mise en place et de la conduite des installations de formation (établissements scolaires et infrastructures au sens étroit du terme). Très souvent, les décisions communales nécessitent d'être approuvées par l'instance cantonale compétente et sont tenues de respecter ses directives et conditions cadres.

La Confédération assume un rôle subsidiaire dans la politique et le pilotage de l'éducation suisses. Ses compétences concernent le domaine de la formation postobligatoire : pour les écoles de maturité gymnasiales, la formation professionnelle et les hautes écoles, la Confédération et les cantons sont des partenaires qui assument la responsabilité de l'instruction publique.

Les cantons sont chargés de coordonner les travaux au niveau national par le biais de la Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP). Cette dernière est l'organe de la collaboration intercantonale dans le domaine de l'éducation et repose sur des conventions revêtant force obligatoire passées entre les cantons (concordats). Tout comme la Confédération, la CDIP agit également à titre subsidiaire et assume les tâches qui ne peuvent pas être assumées par les régions²¹ ou les cantons.

L'art. 61a de la Constitution fédérale constitue la base légale de la collaboration entre la Confédération et les cantons. La loi sur la coopération dans l'espace suisse de formation (LCESF) est fondée sur cet article. La LCESF définit les objectifs et modalités d'organisation de cette collaboration, ainsi que la mise en place et la conduite d'institutions communes afin d'encourager « la qualité et la perméabilité de l'espace suisse

²⁰ Le [site web](#) de la CDIP (CDIP, sans indication de l'année de publication) permet d'accéder facilement aux législations cantonales sur l'éducation et la formation.

²¹ Il existe d'autres organes dans différentes régions de la Suisse, à savoir : la Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP) ainsi que les commissions de collaboration intercantonale suivantes : la Conférence des directeurs de l'instruction publique de la Suisse centrale (BKZ), la Conférence des directeurs de l'instruction publique des cantons de la Suisse orientale et de la Principauté du Liechtenstein (CDIP-Est) et la Conférence des directeurs de l'instruction publique de la Suisse du Nord-Ouest (NW CDIP).

de formation » et de permettre la mise en place d'« une politique de l'éducation objective et cohérente » (art. 1^{er}, al. 2, LCESF). Le comité de coordination de la numérisation de l'éducation (CC N), qui doit servir de « plateforme pour l'échange d'informations structuré de manière stratégique entre la Confédération, les cantons et leurs partenaires dans le système éducatif pour faire face aux défis et tirer parti des avantages de la mutation numérique dans l'éducation » (CDIP et SEFRI 2016, p. 1, al. 2), repose notamment sur la LCESF. Dans le cadre du CC N, la Confédération (représentée par le SEFRI) et les cantons (représentés par la CDIP) se chargent de la coordination de leurs projets et mesures en impliquant d'autres offices fédéraux, des représentants cantonaux ainsi que des acteurs privés afin de relever les défis de la numérisation dans l'éducation.

Le monitoring de la formation est la principale base de la politique et du pilotage de l'éducation. C'est dans ce cadre que sont compilées, regroupées et évaluées des données et informations provenant de la statistique, de la recherche et de l'administration pour créer un système éducatif global. Publié tous les quatre ans, le « Rapport sur l'éducation en Suisse » analyse le niveau national du système éducatif suisse. Il est axé sur les critères de l'effectivité, de l'efficacité et de l'équité. Il est élaboré par le Centre suisse de coordination pour la recherche en éducation (CSRE) sur mandat du Secrétariat d'État à la formation, à la recherche et à l'innovation (SEFRI) et de la Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP). Si l'image du monitoring de l'éducation est assez fermée sur le plan national, elle devient toutefois disparate aux niveaux des conférences régionales et des cantons individuels. Alors que l'Institut de recherche et de documentation pédagogique (IRDp), sur mandat de la Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP), met à disposition chaque année des données sur le système d'éducation de la Suisse romande au sens large du terme (Pagnossin, Armi, & Matei 2014), l'Institut d'évaluation de l'éducation de l'Université de Zurich établit un rapport complémentaire sur l'éducation dans la Suisse du Nord-Ouest (Ender, Moser, Imlig, & Müller 2017) qui vient compléter le « Rapport sur l'éducation en Suisse ». Au niveau cantonal, il ne reste alors plus que certains cantons qui publient périodiquement un compte rendu sur l'éducation. Ce faisant, ils s'orientent moins sur des modèles de monitoring systémiques, mais mettent au contraire plus fortement l'accent sur la nécessité de rendre des comptes aux autorités exécutives chargées de la politique et du pilotage de l'éducation à l'attention du Parlement et du grand public (Departement für Erziehung und Kultur Thurgau 2018).

Les données forment un élément clé du pilotage de l'éducation et constituent une source d'information importante pour les décisions des responsables de la politique de l'éducation. Elles fournissent des indications, selon des degrés d'approfondissement plus ou moins élevés, sur l'input, la structure et les résultats du système éducatif, et permettent ainsi d'avoir une vue d'ensemble sur la réalité du système. Outre la mise à disposition sur l'état actuel du système (par ex. : combien d'enseignants travaillent-ils dans les écoles primaires ?), elles servent surtout de base à la planification des ressources (par ex. à quel rythme les effectifs des élèves vont-ils s'accroître au cours de ces prochaines années ?) et à l'évaluation (par ex. le passage au Plan d'études alémanique LP21 (« Lehrplan 21 ») a-t-il facilité ou non une intégration des apprenants après des changements de cantons ?). Les données qui sont utilisées à ces fins proviennent d'une série de contextes où elles ont pris naissance, étant précisé qu'en règle générale, les sources donnent des indications sur la structure et les inputs du système éducatif, et que les données d'enregistrement mettent à disposition des informations sur les résultats. Ces données sont traitées et analysées soit à l'interne (c.-à-d. au sein des établissements publics d'éducation), soit à l'externe (par ex. lorsqu'elles sont étudiées par des instituts de recherche indépendants). Elles constituent la base du système actuel de monitoring de l'éducation.

La numérisation place la politique et le pilotage de l'éducation devant un triple défi à relever : premièrement, la politique de l'éducation doit tenir compte des changements d'exigences posées aux jeunes gens par l'économie et la société ainsi que par l'« instruction publique » (art. 62 Cst.) et doit être axée sur ces dernières tant sur le plan du contenu que de l'organisation. Récemment, la Confédération et les cantons ont tous deux établi de nouvelles bases stratégiques à cet effet : du côté de la Confédération, il s'agit du rapport du SEFRI sur les *Défis de la numérisation pour la formation et la recherche en Suisse* (SEFRI 2017), du côté des cantons, il s'agit de la *Stratégie de la CDIP du 21 juin 2018 pour la gestion de la transition numérique dans le domaine de l'éducation* (CDIP 2018b). Ces deux documents fondamentaux stratégiques (voir chapitre 3.3.5, p. 25) apportent des clarifications nécessaires, également dans la perspective de l'utilisation des données dans le système éducatif suisse. La dimension fortement fédéraliste de l'éducation et

les nombreux niveaux de compétence correspondants, y compris dans la politique et le pilotage de l'éducation, se traduisent par la lenteur des processus de prise de décision, de sorte qu'il arrive souvent que les décisions ne déploient leurs effets que tardivement. La vitesse vertigineuse et les progrès parfois disruptifs de la numérisation nécessitent toutefois des réactions rapides de grande portée – que ce soit sur le plan du contenu, de l'organisation ou des aspects financiers – que les mécanismes actuels de pilotage de la politique de l'éducation ne parviennent que difficilement à suivre.²² Tel est ici le deuxième défi que recèle la numérisation pour la politique et le pilotage de l'éducation. Jusqu'à présent, les thèmes et les problématiques que soulève la rapide montée en puissance de la numérisation dans l'éducation ne sont que ponctuellement enregistrés par le monitoring de l'éducation. Des recensements et rapports cantonaux qui sont consacrés à la numérisation prennent naissance dans le contexte de projets actuels propres à la politique l'éducation (souvent afin de décrire la situation de départ ou afin de dresser un « état des lieux » initial) et présentent des différences considérables sur le plan du contenu et de la méthodologie. Ainsi, ces recensements sont certes concluants si on les considère chacun isolément, mais ils ne sauraient être regroupés pour former une image globale cohérente dont on pourrait tirer des énoncés de rang hiérarchique supérieur sur l'état actuel de la numérisation dans le système éducatif. Un troisième défi résulte de ce qui précède pour la politique et le pilotage de l'éducation. Il concerne le monitoring de l'éducation lui-même, à savoir l'écart qui s'ouvre de manière croissante entre les sources de données effectivement utilisées et les sources de données potentiellement utilisables. Les données d'enregistrement qui sont utilisées à l'heure actuelle pour le monitoring fournissent des connaissances transversales et ne permettent des comparaisons dans le temps qu'à un niveau fortement agrégé. Quand bien même ces enregistrements peuvent mettre en évidence des changements entre deux points de référence temporels, et au-delà des unités géographiques concernées, des sources de ce genre ne fournissent toutefois pas d'indications claires sur le déroulement des processus en continuels changements : elles sont utilisées en tant que données « les plus récentes actuellement disponibles ». Mais aujourd'hui, par le biais de la numérisation, des ensembles de données beaucoup plus complets provenant de données d'application prennent constamment naissance, et ce, tant au niveau individuel qu'au niveau des acteurs du système éducatif saisis de manière abstraite. Des approches adaptées à ce genre de méthodes relevant de la « Big Data Analysis » permettent notamment de procéder à des comparaisons permanentes structurées de manière récursive qui ne restent pas axées sur le passé, mais autorisent des énoncés prédictifs au sujet de futurs développements (Williamson 2016). À l'heure actuelle, la politique et le pilotage de l'éducation demeurent toutefois très éloignés d'un tel genre d'utilisation de ces nouvelles données. Elle exige des compétences qui sont disponibles aujourd'hui du côté des grands prestataires privés qui exercent leurs activités à l'échelle internationale et qui constituent le cœur de leurs modèles d'affaires, raison pour laquelle ils les conservent jalousement. La politique et le pilotage de l'éducation, auparavant conçus comme relevant de la souveraineté cantonale ou nationale, se déplacent ainsi de manière accrue vers les centres de calcul numérisés où l'expertise technique permettant de calculer, de visualiser et d'analyser les données de l'éducation est disponible : ces grands prestataires déterminent les futurs instruments et préjugent ainsi des futures décisions de la politique et du pilotage de l'éducation à l'ère du numérique.

Le rôle de l'Office fédéral de la statistique dans l'espace suisse de données de l'éducation

L'Office fédéral de la statistique (OFS) est une instance d'importance décisive de la collecte et du traitement des données de l'éducation. Compte tenu de sa mission en tant que centre de compétence national de la statistique publique de la Suisse, il joue un rôle central dans l'écosystème de données du secteur de l'éducation et de la formation. Il produit et publie des informations statistiques sur l'état actuel et l'évolution de domaines importants de la vie de la société, et notamment de la formation (OFS 2018). En outre, il coordonne les travaux statistiques entre la Confédération et les régions (cantons et communes). Les objectifs et les axes prioritaires thématiques de son travail sont chaque fois définis pour une période de législature dans le programme statistique pluriannuel de la Confédération. Ce programme est élaboré « en étroite collaboration avec les autres producteurs de statistiques de la Confédération et des cantons ainsi qu'avec les groupes d'utilisateurs importants issus des milieux politiques, de l'administration, des milieux économiques, des partenaires sociaux et des organisations internationales » (OFS 2016, p. 7).

22 Ce constat ne s'applique que de manière limitée aux établissements scolaires.

Dans le secteur de la formation, l'OFS recense et produit une série de statistiques déterminantes qui fournissent des renseignements sur la structure et l'évolution du système éducatif. En font partie, d'une part, des informations sur les inputs et les paramètres fondamentaux du système tels que le nombre d'établissements de formation, le nombre d'enseignants et d'apprenants, les dépenses publiques en matière d'éducation et de formation, ou le montant des bourses cantonales et des prêts de formation. D'autre part, l'OFS obtient et met à disposition des informations sur certains outputs du système éducatif, par ex. le nombre de diplômes par niveau de scolarité ou l'état de la formation de la population résidante. De surcroît, certains recensements couvrent d'autres aspects pertinents du secteur de la formation. Ainsi, par exemple, le Microrecensement formation de base et formation continue (MZB) et l'Enquête suisse sur la population active (ESPA) fournissent des données sur les activités de formation continue des employés, alors que l'Enquête sur la formation en entreprise (SBW) éclaire la question du côté des employeurs. En outre, depuis l'intégration du nouveau numéro d'assuré AVS à 13 chiffres dans les différents registres de personnes de l'Administration fédérale, le projet d'appariement de données « Analyses longitudinales dans le domaine de la formation » (LABB) a été lancé. Il a pour but de mieux comprendre les déroulements et transitions dans le domaine de la formation au niveau de l'apprenant individuel. De plus, l'OFS élabore, à intervalles réguliers, des scénarios d'évolution du système éducatif, étant précisé qu'un axe prioritaire thématique est consacré aux prévisions sur l'évolution des effectifs des apprenants et des enseignants et sur le nombre des futurs diplômés (Babel, Gaillard, & Strübli 2017).

Ces groupes de données constituent un pilier important du monitoring de la formation, qui est ainsi à l'origine des discussions publiques, politiques et scientifiques dans ce domaine. Les données et les statistiques produites par l'OFS pour le secteur de la formation reposent ainsi sur des approches méthodologiques différentes. Elles vont des enquêtes par sondage écrites et téléphoniques (enquêtes réalisées auprès d'une partie représentative sélectionnée à partir d'un ensemble de population) aux enquêtes complètes (enquêtes réalisées auprès de tous les sujets d'un ensemble de population, par ex. auprès de tous les étudiantes et les étudiants des hautes écoles ayant achevé leurs études), en passant par la consultation électronique de banques de données cantonales ou fédérales existantes jusqu'à la gestion et l'évaluation de registres étatiques (par ex. le Registre des entreprises et des établissements [REE] aux fins d'identification d'écoles et autres institutions de formation).

La transmission et l'utilisation de ces groupes de données sont régies par la loi sur la statistique fédérale (LSF). Elle prescrit que la protection des données et de la personnalité constitue un principe fondamental de la statistique publique. Partant, des règles strictes sont applicables à l'accès aux données de l'OFS (OFS 2006). Ce faisant, il faut observer le principe selon lequel, en règle générale, les données personnelles non anonymisées ne sont pas transmises, et les données non anonymisées particulièrement dignes de protection ne sont jamais transmises (même pas aux services statistiques des cantons et des communes). Les données individuelles anonymisées ne sont également mises à disposition que moyennant la signature d'un contrat sur la protection des données, et uniquement à des fins statistiques et scientifiques. Une interdiction s'applique à l'utilisation de données individuelles destinée à des « actions administratives, à des actions de contrôle, à des actions fiscales ou à des actions de surveillance » (OFS 2006, p. 11).

En outre, au niveau fédéral, l'OFS est l'unique institution habilitée à procéder à des appariements (= croisements de données) entre des ensembles de données publiquement générés (art. 14 LSF). Au niveau cantonal et communal, les services statistiques ne sont autorisés à appairer des données de l'OFS afin d'accomplir leurs tâches statistiques que si elles disposent du consentement écrit de l'OFS à cet effet et qu'elles respectent les obligations correspondantes. Sur demande et moyennant le respect de certaines obligations, l'OFS peut procéder pour des tiers à des appariements de données à des fins non axées sur des personnes telles que la recherche scientifique et la planification.²³ Toutefois, tout appariement de données nécessite une justification écrite qui sera contrôlée par les sections spécialisées concernées, par la section des méthodes statistiques et par le service juridique de l'OFS (OFS 2017).

23 Les conditions préalables à remplir résident dans le fait que les données doivent être exclusivement utilisées à des fins de statistique publique ou de recherche scientifique. Il faut aussi que les données soient anonymisées, qu'elles ne soient pas dé-anonymisées ou qu'elles ne soient pas appariées avec d'autres données, et il faut qu'elles soient effacées une fois l'analyse effectuée. En outre, l'appariement doit être judicieux du point de vue méthodologique, et il doit être réalisable sur le plan technique et organisationnel (OFS 2016b).

3.4.2 Administration de l'éducation et organisation scolaire

Par les expressions « Organisation scolaire » et « Administration de l'éducation », nous décrivons dans les lignes qui suivent un contexte d'utilisation qui regroupe l'établissement scolaire et les services cantonaux et communaux de l'administration de l'éducation, et – dans le cas des écoles professionnelles – également les services fédéraux de l'administration de l'éducation. Certes, cela peut entrer en contradiction avec leur organisation institutionnelle (et la manière dont se considèrent les personnes qui travaillent dans et avec cette organisation), mais il s'agit d'une approche appropriée lorsqu'on les définit en tant que contexte d'utilisation des données. Dans leur rôle d'autorités, tant les établissements scolaires que l'administration de l'éducation sont tenus de mettre en œuvre les prescriptions résultant du processus de pilotage politique du système éducatif dans le cadre d'une offre d'enseignement scolaire qui encadre et soutient l'activité d'enseignement et d'apprentissage du corps enseignant et des élèves. Le contexte d'utilisation « Organisation scolaire et administration de l'éducation » comprend ainsi les structures et processus qu'il faut mettre en place et diriger afin que l'enseignement et l'apprentissage puissent avoir lieu sous la forme de l'enseignement scolaire.

Les législations des cantons sur l'éducation et la formation constituent la base légale pour l'organisation scolaire et l'administration de l'éducation.²⁴ Elles différencient les niveaux scolaires propres au système éducatif et confèrent aux établissements scolaires la mission consistant à offrir l'enseignement scolaire correspondant. Dans ce contexte, il faut accorder une grande importance aux communes qui sont les organes responsables pour le niveau primaire et, très souvent, pour le niveau secondaire I. Au niveau secondaire II, c'est le canton qui est l'organe responsable des écoles de culture générale (gymnases et écoles de formation générale), alors que la Confédération assume également certaines tâches et obligations pour les écoles professionnelles.

Enregistrer, administrer et échanger des données fait partie des tâches clés propres à l'organisation scolaire et à l'administration de l'éducation. À cet égard, il s'agit de données d'administration qui contiennent des informations sur des personnes, des infrastructures et des organisations et qui sont utilisées de diverses façons, à savoir qu'elles sont croisées, analysées et interprétées. Ces données prennent naissance, d'une part, dans le contexte d'utilisation « Enseignement et apprentissage » (voir chapitre 3.4.5, p. 38) et, d'autre part, elles sont générées par les structures et dans les processus de l'organisation scolaire elle-même. Elles sont utilisées par des partenaires de la mise en réseau institutionnelle et/ou sont transmises à ces derniers.

L'utilisation des données dans le contexte d'utilisation « Organisation scolaire et administration de l'éducation » est clairement réglementée pour l'essentiel. Sont déterminantes à cet effet les législations cantonales sur la formation susmentionnées. Elles stipulent, d'une part, les exigences légales fondamentales de la protection des données au sein de l'organisation scolaire et de l'administration de l'éducation, mais autorisent aussi la communication et l'échange de données – y compris des données particulièrement dignes de protection – entre des personnes qui appartiennent toutes à l'organisation scolaire ou à l'administration de l'éducation.²⁵ En outre, les exigences légales propres à la protection des données sont complétées par le secret de fonction. Ce dernier renforce, d'une part, la protection des données personnelles vis-à-vis des tiers qui n'appartiennent ni à l'organisation scolaire, ni à l'administration de l'éducation et, d'autre part, il facilite leur échange en raison des possibilités offertes par l'entraide administrative.²⁶

Les défis actuels de l'utilisation des données dans le contexte d'utilisation « Organisation scolaire et administration de l'éducation » résident dans le domaine pratique. La numérisation des processus génère en effet des quantités de données nettement plus volumineuses que ce n'était le cas dans le cadre de la conduite administrative « analogique ». Ces données sont générées dans une multitude de formats techniques (souvent propriétaires) qui ne peuvent être préparés qu'au prix d'une grande charge de travail, voire ne peuvent être préparés pour un échange de données. Une partie de ces défis à relever trouve aussi son origine dans les infrastructures techniques. Certes, la possibilité d'organiser leur exploitation de manière centralisée et transversale en englobant tous les services, ou de les externaliser à l'attention d'un presta-

24 Le [site web](#) de la CDIP (CDIP, sans date de publication) permet d'accéder facilement aux législations cantonales sur l'éducation et la formation.

25 Voir par exemple l'art. 73 « Protection des données » de la loi sur l'école obligatoire (LEO) du canton de Berne. Consultable sous : [http://www.lexfind.ch/dta/23039/FR\(29.11.2018\)](http://www.lexfind.ch/dta/23039/FR(29.11.2018))

26 Voir par exemple l'aide-mémoire « Secret de fonction » du préposé à la protection des données du canton de Lucerne (2009).

taire externe, accroît la sécurité de l'exploitation, mais elle répartit aussi et formalise de manière nouvelle les responsabilités respectives. Il n'est possible de parvenir que lentement à des améliorations en la matière puisqu'elles présupposent des échanges intensifs et une coordination étroite entre tous les services concernés.²⁷

À la différence de l'administration de l'éducation, l'organisation scolaire doit non seulement se charger des tâches qui sont en rapport avec les données administratives, mais elle est en outre tenue de s'assurer que les droits des enseignants et des élèves sont garantis lors de l'utilisation de services en ligne. Cette responsabilité découle moins des dispositions légales que des contrats que les écoles ont conclus avec leurs prestataires s'agissant de l'utilisation de services en ligne par des enseignants et par des élèves.²⁸ Sont également concernées les « traces d'apprentissage », resp. les données d'application qui sont générées par les élèves. Il incombe par conséquent à l'organisation scolaire de les tenir à disposition conformément aux exigences légales de la protection des données, de les protéger contre des utilisations abusives et contre des corruptions de toute nature. Les défis qui en résultent pour l'organisation scolaire sont considérables.

L'organisation scolaire est non seulement tenue de garantir, en prenant les mesures organisationnelles et les mesures d'infrastructure nécessaires, que les systèmes informatiques et les processus basés sur ces derniers satisfont effectivement aux exigences requises, mais elle doit aussi veiller à ce que ces exigences soient respectées par les prestataires de services en ligne. Il faut partir du principe que toutes les écoles ne sont pas en mesure de tenir suffisamment compte de ces deux aspects constitutifs d'une utilisation des services en ligne conforme aux prescriptions légales.²⁹

3.4.3 Recherche en éducation

Selon [le Centre suisse de coordination pour la recherche en éducation](#) (CSRE), la première mission de la recherche en éducation consiste à étudier et expliquer la réalité du déroulement des événements de la formation. Elle analyse les structures, les processus et les résultats du système éducatif – et aussi du système éducatif extra-institutionnel – en faisant appel à des méthodes scientifiques. Elle aide ainsi les milieux politiques et les responsables de la pratique à poursuivre l'élaboration des questions qui ont trait à l'éducation et à l'école (CSRE-SKBF 2018). Dès lors, le concept de « recherche en éducation » décrit un champ de recherche défini de manière large auquel contribue un nombre important (et croissant) de chercheurs issus de différentes disciplines (Köller 2014 ; Tippelt & Schmidt-Hertha 2018). Le CSRE documente la recherche en éducation pratiquée en Suisse. De manière structurelle, cette recherche bénéficie d'un large ancrage et est déjà bien établie. Ainsi, [la banque de données de recherche du CSRE](#) contient des entrées sur des projets de recherche réalisés par près de 4000 chercheuses et chercheurs travaillant dans plus de 70 instituts domiciliés en Suisse – qui vont des universités et des hautes écoles pédagogiques aux instituts de recherche indépendants en passant par des unités d'organisation de l'administration cantonale et fédérale (état : 30.11.2018).

Au cours de ces dernières décennies, l'importance des données et des évaluations de données pour la recherche en éducation s'est accrue rapidement (Gräsel 2015). En Suisse également, on constate qu'elle revêt une importance croissante. Ainsi, pendant la période 2005-2011, le pourcentage des projets de recherche quantitatifs (à savoir des projets qui procèdent à des analyses statistiques sur la base de mesures expérimentales ou non expérimentales) a progressé de 17 % par rapport à la période de comparaison 1980-1984. À l'inverse, le pourcentage des projets qualitatifs a reculé de 10 %, et celui des autres projets a diminué de 5 % (Grossenbacher & Oggenfuss 2015). Ce faisant, la recherche scientifique travaillant selon une approche empirique s'est avérée être un instrument central pour l'examen de la qualité des écoles

27 La nécessité d'agir a été du moins reconnue dans le domaine de la formation professionnelle. La CDIP a décidé en juin 2018 de lancer le programme « optima », « un programme de coordination dans lequel l'échange de données sera globalement réglementé et mis en œuvre » (CDIP 2018a, p. 1, al. 5).

28 Pour les conditions contractuelles à respecter lors de l'utilisation scolaire de services en ligne, veuillez consulter la « recommandation du CTIE aux écoles publiques concernant les dispositions du droit de la protection des données relatives à la transmission de données des élèves aux prestataires de services web » (CTIE 2012).

29 La protection des données n'est qu'un aspect de la question de la compliance de l'utilisation des TIC dans les écoles. Elle inclut d'autres défis à relever comme la question des acquisitions TIC, celle des contrats, des pratiques d'utilisation, de la mise en place des compétences ainsi que de la gouvernance. Selon des déclarations de membres de l'administration de l'éducation, les projets actuels d'acquisitions TIC en lien avec l'introduction du Plan d'études alémanique LP21 montrent que le niveau des connaissances sur ces thèmes est plutôt faible dans les écoles et au sein des directions des écoles. Voir par exemple l'article de Steimen (2018), du Service de recherche sur la durabilité numérique de l'Université de Berne, à propos de la Conférence sur les marchés publics et les acquisitions TIC 2018.

et de l'enseignement, le diagnostic et le contrôle des tâches propres aux réformes, et elle s'est désormais bien établie pour l'évaluation de résolutions de problèmes potentiels (Tippelt & Schmidt-Hertha 2018). Elle constitue ainsi le cœur de l'actuel système de monitoring de l'éducation (Wolter et al. 2018). Compte tenu de ce qui précède, il est dès lors peu surprenant que le degré d'intensité d'utilisation des données soit très élevé, et que des données provenant d'un grand nombre de contextes de naissance différents soient utilisées.³⁰ En outre, une multitude de données d'enquête et d'enregistrement institutionnalisées et « spontanées » dans le domaine de l'éducation sont générées par les institutions de recherche en éducation elles-mêmes. Il s'agit ainsi d'une source d'importance cruciale pour la génération de données d'enquête. L'enregistrement, la mémorisation et la transmission de ces données ne sont soumises à aucun standard global revêtant force obligatoire. Certes, tant au niveau national qu'international, il existe une série de guides pour la structuration des enquêtes et recensements (par ex. les *Survey guidelines* de l'Institut Leibniz pour les sciences sociales [Silber, Daikeler, Heycke, & Stiegler 2019]) et pour la définition et la mémorisation des attributs (par ex. le standard eCH pour l'échange d'identifications de personnes [Steimer & Stingelin 2014]) ; néanmoins, il arrive souvent que des standards ne soient définis que de manière spécifique à certains cas, et que l'accès aux données soit limité – également pour des raisons de protection du droit de la personnalité. Dès lors, une réutilisation des données par des tiers, ou un appariement avec d'autres jeux de données ne sont souvent pas garantis même si l'enquête a été financée par des fonds publics. Une recherche effectuée dans le catalogue de données du Centre de compétences suisse en sciences sociales (FORS), par exemple, n'a permis de trouver que 40 jeux de données enregistrées pour le mot-clé « Formation » (état : 20.12.2018).

Dans la perception publique de la recherche en éducation, ce sont les données d'enquêtes provenant de comparaisons de prestations nationales et internationales qui dominent, par ex. le *Programme for International Student Assessment* (PISA) et la vérification de l'atteinte des compétences fondamentales (enquêtes COFO). Ces enquêtes mettent en œuvre des tests standardisés pour une série de domaines de compétences (par ex. mathématiques, langue parlée à l'école et première langue étrangère pour les enquêtes COFO). Elles servent ainsi en priorité à la mesure externe des compétences dont disposent les élèves à l'issue de la période de scolarité obligatoire. De surcroît, par le biais de questionnaires, on collecte des informations complètes sur les caractéristiques d'un grand nombre de groupes cibles (par ex. apprenants, parents et écoles dans le cadre de PISA). Elles offrent donc une image transversale complète sur le statu quo du système éducatif, et permettent ainsi de mener une discussion empiriquement fondée sur la qualité des processus éducatifs (Gräsel 2015). Elles servent en outre d'étalon de comparaison pour l'efficacité des systèmes éducatifs au-delà des frontières cantonales ou nationales et ont dès lors une influence considérable sur le pilotage politique des systèmes éducatifs. Toutefois, à elles seules, ces enquêtes transversales ne sont pas suffisantes pour procéder à une analyse causale précise des influences exercées sur les prestations. Pour ce faire, comme dans la recherche médicale, il est nécessaire de disposer d'enquêtes longitudinales établies sur la base d'une répartition randomisée de sujets de données subdivisés en groupes de traitement et groupes de contrôle (Beglinger 2018 ; Connolly, Keenan, & Urbanska 2018). En outre, même des enquêtes institutionnalisées et régulièrement répétées, comme les enquêtes PISA, sont limitées dans le temps. C'est-à-dire qu'elles n'offrent des informations sur des prestations des apprenants et du système éducatif qu'à des dates déterminées et ce, sans que l'on puisse saisir ce qui se passe entre ces dates. De surcroît, les résultats ne sont publiés qu'avec un retard considérable. Au final, cela signifie que des décisions de pilotage reposent souvent sur des observations statiques, et peut-être même sur des données obsolètes.

Un autre défi à relever pour la recherche en éducation réside dans l'analyse de la structure du système éducatif, des interactions entre ses différents niveaux, et des effets de ces interactions sur le comportement d'apprentissage et ses résultats. Les élèves se trouvent dans des classes qui, à leur tour, composent des écoles. Ces écoles sont à leur tour soumises à une administration commune qui se répartit en plusieurs niveaux administratifs. Chacun de ces niveaux influence l'environnement d'apprentissage et, de ce fait, le comportement d'apprentissage et ses résultats. Dans ce contexte, des décisions prises à différents niveaux peuvent soit s'affaiblir, soit se renforcer mutuellement. Afin de comprendre les effets de ces interactions, la recherche en éducation est donc souvent tributaire également d'une combinaison de données pro-

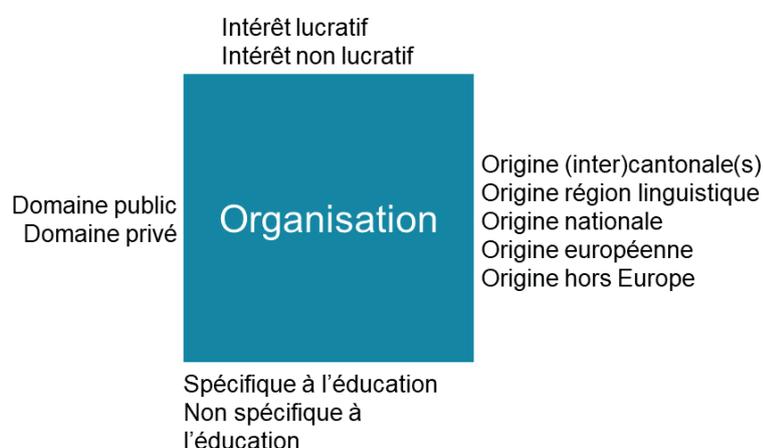
³⁰ Les données de l'administration, collectées et gérées par l'OFS, constituent une importante source de données (voir le texte sur « Le rôle de l'Office fédéral de la statistique dans l'espace suisse de formation des données. »)

venant de différentes sources telles que des données d'enregistrement et des données d'administration. La diffusion croissante des moyens et applications d'enseignement numériques, ainsi que les développements qui ont eu lieu ces dernières années dans le domaine de la technologie des ordinateurs et de l'analyse des données ont en outre étendu les possibilités d'application des méthodes statistiques à des phénomènes qui n'étaient jusqu'ici pas saisissables, ou seulement difficilement saisissables du point de vue quantitatif. En font partie, par ex. l'analyse de l'importance des relations émotionnelles et des sympathies sur les processus d'apprentissage et leurs résultats (par ex. Madaio, Peng, Ogan, & Cassel 2018) ou l'examen détaillé des processus d'apprentissage (par ex. Jovanović, Gašević, Dawson, Pardo, & Mirriahi 2017). Toutefois, le recours à des ensembles de données de ce genre et à de telles techniques d'analyse pour répondre à des questions de la recherche en éducation est encore un champ de recherche comparativement récent, même si on leur attribue un potentiel de changement considérable – en particulier pour le monitoring du système éducatif (par ex. Cope & Kalantzis 2015 ; Mitros et al. 2017).

3.4.4 Marché de l'éducation

Outre les acteurs publics issus des milieux politiques et de l'administration, les prestataires de produits et services provenant de l'économie privée jouent un rôle toujours plus important dans le système éducatif. Ce « marché de l'éducation » comprend notamment des organisations qui offrent des technologies de l'information et de la communication (par ex. infrastructure TIC, connectivité, terminaux, applications logicielles utilisées en mode local, apps mobiles, technique de sécurité, etc.) ainsi que des moyens numériques d'enseignement et d'apprentissage (des contenus préparés sur le plan didactique destinés à l'enseignement) pour être utilisés dans d'autres contextes d'utilisation.

Les fournisseurs de solutions TIC et de moyens numériques d'enseignement et d'apprentissage peuvent être catégorisés selon les quatre dimensions suivantes : selon leur intérêt lucratif ou non lucratif, selon leur origine, selon leur structure de domaine public ou privé, et s'il s'agit d'offres spécifiques à l'éducation ou non (voir Graphique 8). Lorsqu'ils cherchent à se procurer des renseignements, les acteurs et institutions du système éducatif étatique se voient ainsi en principe confrontés à une multitude de prestataires hétérogènes qui leur proposent une offre complexe largement diversifiée et dont le volume est potentiellement croissant.



Graphique 8 : Types d'organisations des fournisseurs de services en ligne

Tout comme dans d'autres secteurs économiques privés, des entreprises qui ne sont toutefois pas axées spécifiquement sur l'éducation et qui exercent leurs activités au niveau mondial dominant également le marché de l'éducation : Apple, Google et Microsoft jouent le rôle de leaders à cet égard. Dans le contexte d'utilisation « Organisation scolaire et administration de l'éducation », par exemple, Microsoft dispose d'un quasi-monopole puisque la quasi-totalité des services administratifs publics misent sur des solutions de Microsoft. En revanche, des marchés partiels spécifiques, par ex. les logiciels destinés à l'administration des écoles, sont dominés, pour une part déterminante, par des prestataires nationaux. À l'heure actuelle, il n'existe aucune vue d'ensemble, à l'échelle de toute la Suisse, sur le volume effectif du marché de l'éducation ou sur la diffusion de technologies ou de solutions logicielles individuelles au sein des établissements de

formation.³¹ Une harmonisation et une coordination des projets d'acquisition correspondants au-delà des frontières cantonales – si tant est qu'elles existent – ne sont mises en œuvre qu'à très petite échelle.

Dans l'ensemble, on peut constater que, dans le cadre du marché de la numérisation, c'est le côté de l'offre (par opposition à celui de la demande) qui connaît une progression importante. Outre de nouvelles petites entreprises et microentreprises, un nombre croissant de prestataires dont les solutions n'ont pas été développées pour une utilisation dans un contexte scolaire se pressent sur le marché. La plupart du temps, un recours à ces produits dans le secteur éducatif nécessite une adaptation des conditions d'utilisation aux exigences juridiques en vigueur et à des aspects spécifiques à l'éducation.³²

Outre les données d'administration qui sont nécessaire à la conclusion des contrats, à la facturation et au commissionnement, on collecte parfois à grande échelle auprès des acteurs du marché des données sur le comportement et les prestations d'utilisateurs individuels. Par exemple, les données provenant des systèmes d'apprentissage numériques fournissent des renseignements détaillés sur les progrès et les difficultés d'apprentissage des apprenants qui utilisent ces systèmes. À l'heure actuelle, ces données sont, pour leur plus grande part, des données propriétaires, c.-à-d. qu'elles sont exclusivement mises à disposition du prestataire qui, en outre, a le pouvoir de définir les objectifs d'utilisation liés à ces dernières – en vertu de la structuration des conditions générales de vente (CGV), des conditions d'utilisation et de la déclaration relative à la protection des données. De plus, en règle générale, certains établissements de formation ne disposent pas d'un pouvoir de marché suffisant pour pouvoir obtenir l'adaptation de ces conditions cadres afin de faire en sorte qu'ils puissent avoir accès à ces données. Dans ce contexte d'utilisation, les instruments de régulation déterminants sont donc plutôt le Règlement général de l'UE sur la protection des données (RGPD) et la loi fédérale sur la protection des données (LPD) de la Suisse, et en particulier les lois cantonales sur la protection des données. En outre, les plans d'études propres aux différentes régions linguistiques et les mesures de contrôle de qualité ont des effets régulateurs sur l'utilisation des moyens d'enseignement (définition du statut des moyens d'enseignement).

3.4.5 Enseigner et apprendre

Ce qui relie de manière centralisée les autres contextes d'utilisation des données, ce sont toutes les activités qui se déroulent dans le cadre du contexte d'utilisation « Enseigner et apprendre », à savoir les activités réalisées dans le cadre de l'élaboration concrète des connaissances par les apprenants ainsi que du soutien mis à leur disposition par le corps enseignant. En dernière analyse, ces activités qui ont lieu dans les autres contextes d'utilisation, y compris l'utilisation des données, visent à optimiser ces processus.

Toutefois, également dans un contexte d'enseignement, des données sont collectées, utilisées et évaluées. Ce faisant, en règle générale, l'objectif est de mettre à disposition tant des enseignants que des apprenants des informations sur le comportement d'apprentissage et sur ses résultats. Cela doit permettre, d'une part, un ajustement entre l'acquisition des compétences atteintes et l'acquisition des compétences attendues. Et, d'autre part, il s'agit de fournir des indications sur les écarts entre les valeurs réalisées et les valeurs attendues afin de procéder aux adaptations nécessaires des processus d'enseignement et d'apprentissage. À cette fin, une série d'informations sur les apprenants et leurs interactions avec les enseignants et les autres camarades de classe ont été collectées par le passé et sont en train d'être collectées actuellement. Cela concerne, par exemple, des indications sur les absences et le comportement figurant dans les carnets de classe, resp. des enregistrements de données personnelles sur les enseignants. Font aussi partie de cette liste des données saisies les évaluations sommatives des prestations³³ que les enseignants sont légalement tenus d'enregistrer. Par comparaison avec d'autres données, les enregistrements sous forme d'« épreuves » ou de « tests » sont fortement standardisés puisqu'on se base sur ces résultats sous forme de « notes scolaires » et de « moyennes de notes scolaires » pour décider de la promotion des élèves d'une année scolaire

31 Toutefois, dans certains cantons, on procède à des enquêtes annuelles sur l'équipement et l'utilisation de solutions TIC dans le domaine scolaire (par ex. fri-tic 2017).

32 Pour structurer le recours à des solutions conforme aux dispositions légales et à des conditions d'utilisation spécifiques, la conclusion de contrats-cadres est un outil approprié. Dans ces contrats-cadres, il est en effet possible de convenir à un niveau hiérarchique supérieur des conditions légales (régime juridique, protection des données, responsabilité civile, etc.) ainsi que des exigences spécifiques au secteur éducatif (interdiction de faire de la publicité, utilisation des données par des élèves mineurs, etc.), ce qui n'est pas possible, la plupart du temps, au niveau des établissements de formation individuels.

33 En d'autres termes, il s'agit d'évaluations de l'atteinte d'objectifs d'apprentissage définis à l'avance réalisées à une date déterminée.

donnée vers la prochaine année scolaire, d'un niveau de formation donné vers le prochain niveau, ainsi que pour attribuer des niveaux de prestations aux élèves. Du côté des enseignants, des données sur leurs prestations sont également enregistrées. Elles visent à évaluer les modalités didactiques de leurs cours. Les enregistrements sont mis en œuvre sous de multiples formes dans le cadre des mesures de contrôle et de développement de la qualité qui ont lieu la plupart du temps à l'interne de l'école, raison pour laquelle elles sont organisées de manière différente ; des enquêtes auprès des élèves sur les cours sont largement répandues, de même que des visites d'enseignants collégiales, des visites de la direction des écoles ou de la surveillance des écoles. Bon nombre de ces informations sont encore générées aujourd'hui, pour l'essentiel, de manière analogique, même si l'utilisation croissante de logiciels d'administration scolaire entraîne une numérisation toujours accrue de ces données.

Dans la vie quotidienne de l'enseignement, on produit aussi dans une mesure croissante des données sur les processus et méthodes d'apprentissage. Ces informations prennent naissance en tant que « traces d'apprentissage » à partir du processus d'enseignement / d'apprentissage concerné. On leur accorde une grande importance dans le cadre de l'enseignement axé sur les compétences basé sur le Plan d'études romand (PER), le Plan d'études alémanique (« Lehrplan 21 ») et les plans d'études destinés à l'enseignement dispensé dans les cycles de formation du niveau secondaire II. À l'heure actuelle, ces données sont collectées, par ex., en tant que « feed-backs » d'autres camarades de classe et en tant que « feed-backs » d'enseignants, resp. en tant que « journaux d'apprentissage » structurés de manière autoréflexive et enregistrés sous une forme qui n'est guère standardisée.

Tant les données provenant des évaluations des prestations que celles qui sont issues des processus d'enseignement / d'apprentissage sont des données personnelles. Du point de vue de la protection des données, elles doivent être considérées comme des données sensibles puisqu'elles permettent non seulement d'évaluer la compétence objective des élèves, mais qu'elles rendent visibles également leur comportement d'apprentissage en tant qu'individus et au sein de groupes sociaux (Odermatt et al. 2009).

La numérisation croissante de l'enseignement et de l'apprentissage a pour effet que de nombreux procédés et processus qui n'étaient jusqu'à présent que difficilement enregistrables dans le cadre du processus d'apprentissage peuvent être enregistrés et quantifiés de manière plus facile aujourd'hui. Ces informations prennent naissance surtout en raison de l'utilisation d'applications qui, outre les réponses correctes qu'elles fournissent aux questions posées, peuvent saisir également de manière détaillée le comportement des utilisateurs. Jusqu'à présent, on recourt notamment à ces informations lorsque l'application utilisée met à disposition un résumé de ces données facile à comprendre (par ex. sous la forme d'un tableau de bord. En revanche, l'absence de plateformes appropriées par le biais desquelles les données seraient mises à disposition, mais aussi des conditions contractuelles formulées de manière unilatérale conclues avec le fournisseur de services (par ex. via les CGV ou les conventions d'utilisation) empêchent que ces données ne soient utilisées par les enseignants et les apprenants. Pour l'instant, l'absence de compétences techniques des enseignants en matière de préparation, d'analyse et d'interprétation des données à l'attention des élèves constitue encore un problème – du moins à l'heure actuelle. Néanmoins, on estime que le potentiel de ces données d'application est très important, tout particulièrement pour le contexte d'utilisation « Enseignement et apprentissage » (Johnson, Adams Becker, Estrada, & Cummins 2015). Car c'est sur ce type de données que se basent les architectures informatiques, déjà disponibles aujourd'hui, des systèmes d'apprentissage / de tests adaptatifs ainsi que des approches dites du « Learning Analytics ». Tous deux sont déjà repris par la recherche internationale en éducation et font l'objet de discussions intensives. Tant les partisans de ce type d'approche que ceux qui sont critiques à cet égard les considèrent comme l'un des défis de principe que devra relever bientôt la numérisation de l'éducation (New 2016 ; Zeide 2017).

3.4.6 Des points de vue différents concernant les données de l'éducation

Les énoncés suivants ont été enregistrés pendant divers entretiens informels menés avec divers acteurs de l'éducation dans les différents contextes d'utilisation des données.³⁴ Ils attirent l'attention sur des problèmes et des déficits dans l'utilisation actuelle des données dans le secteur éducatif et expriment des souhaits – parfois (fortement) divergents – en ce qui concerne une future politique d'utilisation des données.

³⁴ Afin d'obtenir des points de vue supplémentaires, une enquête semi-structurée à ce sujet a en outre été réalisée par courriel. Les résultats correspondants ont été également intégrés ici.

Politique et pilotage de l'éducation

- « Aus dem Bildungsmonitoring-Kontext heraus kenne ich die zugänglichen Datenquellen sehr wohl. Zu den heute aber aufgrund Digitalisierung anfallenden Daten habe ich / haben wir keinen Zugang, sodass sie auch nicht als Grundlage für die politische Steuerung des Systems zur Verfügung stehen. »
- « Alle drei Dimensionen [Datenerfassung, Datenzugang, Datennutzung] sind (in ihrem Zusammenspiel) stark untergenutzt. Das ' Wertschöpfungspotential ', das darin steckt, wird im Bildungswesen nicht annähernd adäquat genutzt. »
- « Die Datenfrage ist noch nicht geklärt... zu komplex... wir sind noch nicht so weit... »
- « Wenn Nutzen für etwas anderes abfällt, geschieht dies eher per Zufall. Es ist nicht transparent, wer zu was Zugang hat – die Auftraggeber sind überfordert. »
- « Bisher ging man davon aus, dass der Anbieter zuständig ist – muss ich mir in Zukunft bei der Lehrmitteleinschätzung auch noch überlegen, was es mit den Daten auf sich hat ? Dann ist diese jedoch nicht mehr im gleichen Rahmen machbar – der Aufwand wäre zu gross ! »
- « Wir wollten die mit den Lernfördersystemen generierten Daten dieses Jahr erstmals auswerten (mittlerweile haben wir viele Daten ; wir sammeln seit 15 Jahren) ; leider konnten diese – wissenschaftlich – jedoch nicht ausgewertet werden, da sie nicht vergleichbar waren (Standardisierung der Erhebungssituation ist nicht geregelt). »
- « Es gibt noch keine Kultur für systematische Auswertung. Bisher galt das Credo : möglichst keine Daten weitergeben. »
- « Gemeinsame Richtlinien wären sehr sinnvoll (Common Agreements als wirkungsvolles Instrument, praktikabler als gesetzliche Vorgaben). »
- « Welche Fragen sollen/müssen wir uns stellen ? Wohin können wir denken → Rahmen ist gewünscht ! Auf höherer Ebene ! »
- « Datenpools sollten verknüpft werden (von Kantonen, Wissenschaft etc.). »
- « Sensibilisierung und Kommunikation sind bei Digitalisierung und Datafizierung zentral. »
- « Leider sind folgende Daten für uns als interkantonales Kompetenzzentrum nicht zugänglich : Adressdaten von kantonalen Mitarbeitern, die in x verschiedenen Gremien für x verschiedene Institutionen tätig sind und überall einzeln erfasst sind. Aus Datenschutzgründen ist eine Verbesserung allerdings kaum möglich. »
- « Für unser Magazin haben wir alle (?) Schulen der Deutschschweiz erfasst. Andere offizielle Institutionen (wie die EDK) haben diese Adressen wohl auch. Ein zentraler Pool für alle Institutionen der Kantone wäre von Vorteil. Aus Datenschutzgründen ist eine Verbesserung allerdings kaum möglich. »
- « Mir ist es ein Anliegen, dass wir in diesem Bereich [Datennutzung] Aussagen bekommen. Dieser Bereich ist heute sehr unübersichtlich. Die Lehrpersonen werden mit diesen Fragen meistens alleine gelassen. Im Kanton X sind wir deshalb daran, eine ' Strategie Datenschutz Schulen ' zu erarbeiten. »
- « Schuladministration : Ein weiterer Ausbau in Richtung Reporting und Steuerung (z. B. Freigabe Kurs- und Klassenbildung) wird angestrebt. »

Administration de l'éducation et organisation scolaire

- « Nous constatons qu'il est parfois difficile d'avoir accès à des données car la logique qui prévaut est encore assez « cloisonnée » dans chacun des services et certaines données ne sont pas facilement accessibles. Le souhait au niveau du DFJC est de pouvoir fournir plus facilement des données utiles au pilotage stratégique mais aussi de l'état du système. »

- « Il faudra créer une instance transversale à tous les services. Les données dont nous disposons ne sont pas accessibles au grand public, ni aux autres institutions pour des raisons de protection des données. »
- « Il n'y a pas de collecte systématique. Des analyses d'usage de site web étaient effectuées avec Google Analytics, mais la solution n'ai plus autorisée en raison des contraintes de sécurité à l'État. »
- « Qu'il s'agisse de la protection des données, d'éthique ou du pilotage de l'école numérique, il nous faut porter une importance aux données issues de l'usage pédagogique. Placer les données scolaires au cœur de la stratégie numérique. » « D'ailleurs le système éducatif français positionne en première action la place centrale des données. »³⁵
- « Les sources de données les plus importantes seraient liées à l'usage des systèmes en ligne. Mais actuellement aucune information d'analyse n'est extraite par exemple du Moodle du DIP. Nous pratiquons des expérimentations avec la solution Hapara Analytics sur la suite Google Apps. Je suis le seul à accéder à ces données. »
- « Oui, j'aurai besoin de beaucoup de données mais elles sont rarement disponibles. Trop de volume pour avoir des statistiques d'usage des réseaux, pas d'outils déployés sur les ordinateurs ou tablettes pour savoir s'ils sont utilisés ou qu'elles ont les applications ou apps utilisées, pas de statistique sur les web sites ... » « Les données ne sont pas accessibles. Le management est peu intéressé par des indicateurs. »
- « Les outils, le management peu enclin à la transparence, les utilisateurs 'criant au flicage'. »
- « Interessant wären alle Daten im Bereich « Learning Analytics » ; Tools und Ressourcen (Zeit und Know-how) fehlen gegenwärtig, aber das ist wohl eine Frage der Zeit. »
- « Alle Schulen verfügen über Daten, die auch für andere Schulen interessant wären. »
- « Es fehlen Tools und Ressourcen für eine systematische Auswertung. »
- « Wichtig ist die Balance zwischen Datenanalysen / Open Data auf der einen und Persönlichkeitsschutz / Sicherheit der Daten auf der anderen Seite. »
- « Monitoring von Bildungsprozessen kann zwar die Effizienz steigern, aber auch einem utilitaristischen Denken Vorschub leisten und damit negative Einflüsse auf die notwendige Freiheit für die Persönlichkeitsentwicklung haben. »
- « Es braucht eine Strategie Datenschutz Schulen. »

Enseigner et apprendre

- « Die Datenschutzrichtlinien sollten harmonisiert werden ! Es ist eine Ressourcenverschwendung und dies ist nicht haltbar. Wir haben alle die gleichen Dinge zu regeln. »
- « Die im Lernnavi generierten Daten hätten wir gerne – zur Unterstützung einer gezielten Förderung des einzelnen Schülers. »
- « An unserer Schule haben alle LPs ein Lehrer-Laptop. Wir denken im Moment darüber nach, Lehrer-Office als « Netzwerk » in der Schule einzurichten. So hat man z. B. als Klassenlehrperson die Möglichkeit, die Noten der SuS in allen Fächern einzusehen. Im Moment arbeiten einige von uns mit der Desktop-Version : Diese ermöglicht nur die isolierte Noten-/Absenzenverwaltung einer einzelnen LP. »
- « Ausserdem fände ich es auch praktisch, wenn man allenfalls wichtige Schulinformationen (Spezialanlässe, besondere Situationen mit SuS o.ä.) elektronisch zur Verfügung hätte (à la Pinnwand) – keine Ahnung, ob es sowas gibt, aber dann hätte jede/-r ohne Lehrerkonferenz/Mailflut immer alle wichtigen Infos

³⁵ Cf. la politique numérique au service de l'école de la confiance : <http://www.education.gouv.fr/cid133192/le-numerique-service-ecole-confiance.html>

auf einen Blick. Ich stelle mir eine elektronische Post-it-Wand vor mit Infos wie : Welche SuS sind heute nicht da (krank, Halbtage)? LPs können dieses Post-it ergänzen und alle sehen es direkt auf ihren Laptops. Andere Infos, die mir gerade in den Sinn kommen : « Zähneputzen diese Woche ! », Traktandenliste LeKo... »

- « An einer Schule gibt es administrative Daten und Lerndaten. Die Administration ist fast komplett digitalisiert, auch wenn die Daten teilweise nicht direkt erfasst werden (z. B. werden Absenzen oft zuerst auf Papier erfasst und entschuldigt und später erst für administrative Zwecke aggregiert). Lerndaten werden kaum elektronisch erfasst, auch nicht systematisch. Sie werden zufällig sichtbar, wenn digitale Anwendungen verwendet werden, bleiben sonst aber meist bei den Schülerinnen und Schülern. Auch Noten sind über Institutionen hinweg nicht greifbar, eine genaue Entwicklung des Leistungsstands oder der Kompetenzen von Lernenden ist nicht dokumentiert. »
- « Das [ideale Organisation des Prozesses der Datenerfassung bzw. Datengenerierung] hat weitreichende ethische und systemische Konsequenzen, weshalb es auf diese Frage keine schnelle Antwort gibt. Generell sind heute Fachleute mit der Datenerfassung betraut, die auch die spezifischen Anforderungen von Schnittstellen kennen. Das ist sicherlich ein sinnvolles System. »
- « Für die Planung und die Gestaltung von Unterricht greifen Lehrpersonen aber kaum auf Daten zurück, auch, weil diese fehlen. »
- « Ich habe meist aus Datenschutzgründen keinen Zugriff darauf. Beispielsweise wäre es sinnvoll, auf frühere Leistungsnachweise von Lernenden zugreifen zu können. Wurden die an einer anderen Institution durchgeführt, sind sie nicht zugänglich, auch um die Rechte von Lernenden zu schützen. »
- « Um Lernprozesse optimieren zu können, könnten Datenanalysen sehr hilfreich sein. Sie widersprechen aber eigentlich der pädagogischen Kultur in der Schweiz. »

Marché de l'éducation

- « Actuellement, très peu de données sont collectées et exploitées. Elles servent uniquement d'indicateurs pour connaître l'usage de la plateforme électronique. »
- « Datensicherheit wird ein immer wichtigeres Thema. »
- « Unsere Daten sind wertlos ohne Kontext. »
- « Es braucht schweizweite Regelungen im Bereich des Datenschutzes. Ansonsten haben einzelne SP Wettbewerbsnachteil. » « Diejenigen, die sich an Regeln halten, sollten wirtschaftlich nicht benachteiligt werden. »
- « Unsere Daten sind für die Bildungsplanung nicht interessant. Sie erheben lieber selber Daten. »
- « Die Ansprüche und Wünsche an die Digitalisierung sind hoch seitens der Lehrer. »
- « Anforderungen im Feld sind naiv. Am liebsten hätten sie alle Daten. »
- « Es braucht mehr Forschung. »
- « Ich wünsche mir klare, aber nicht zu enge Rahmenbedingungen (keine technischen Vorgaben). »
- « Schulentwicklung können wir als Verlag nicht machen ... Es braucht Geräte. »
- « Bisher wurden Daten sehr wenig genutzt : für Support und Statistiken. »
- « Wir müssen den Rahmen kennen. Es braucht vernünftige (einheitliche) Richtlinien. »
- « Es braucht gute Rahmenbedingungen. Sonst vergeben wir auch als Land wichtige Chancen. »
- « Wir wollen nicht Systeme aufbauen, die helfen zu kontrollieren (ethische, didaktische Grundsätze). »

- « Ich hoffe auf ein « Swiss Finish » beim DSG. Es muss praktikabel bleiben. Eigentlich hatte man mit der DSGVO die Grossen (Google und Facebook) im Visier, nun trifft es aber die kleinen Verlage ! »
- « Wir sind uns bewusst, dass spannende Daten generiert werden, haben uns jedoch noch keine Gedanken gemacht, wie diese genutzt werden könnten – unser Fokus ist zurzeit, ein gutes Lehrmittel für die Lernenden zu produzieren. »
- « Bisher gab es noch keine Anfragen von der Bildungsforschung. »
- « Lernen steht bei Lehrmitteln im Vordergrund (nicht die Datenfrage). »
- « Nach Möglichkeit wollen wir keine personenbezogenen Daten haben. Uns reichen anonymisierte Daten vollkommen. Wer genau das ist, der mit der Applikation arbeitet, ist für unser Produkt unerheblich. »
- « Wir haben noch kein Geschäftsmodell für digitale Lehrmittel. »
- « Personenbezogene Daten interessieren uns [als Verlag] aus Datenschutzgründen nicht. » « Anonymisierung ist für Anbieter ein riesiger Vorteil. »
- « Man muss sich bewusst sein, was gemessen werden kann und was nicht. » « Und das Ziel der Datenanalysen muss klar sein. »
- « Wir stehen noch sehr am Anfang bezüglich Konzepten der Datennutzung. Wir haben seit Jahren Apps und Internetplattformen ergänzend zu Printprodukten ; es werden aber nur Nutzungsdaten erhoben, keine Bewegungsdaten. »
- « Es braucht in Zukunft stärkere Zusammenarbeit mit der Bildungsforschung (zurzeit kein einziges Projekt am Laufen). Bisher ergriff nur die Verlagsseite die Initiative. » « Es bestehen Organisationssilos, nicht nur Datensilos. »
- « Die Digitalisierung soll dem Primat der Didaktik folgen. »
- « Meines Erachtens ist es viel wichtiger, gute bzw. bessere Lehrer zu haben (also in die Lehrerausbildung investieren), als sich um die Datengeschichte zu kümmern. »
- « Ich gehe davon aus, dass sich auch die Bildung am künftigen Datenschutzgesetz orientieren wird. Was ist der Sinn und Zweck eines zusätzlichen Berichts ? »
- « Wenn die EDK/Steuerungsebene Interesse an unseren Daten hat, soll sie sie selber hosten. »
- « Wir geben der Forschung sicher keine Daten ! »
- « Die Forschung müsste sich um die Lehrmittelwirkung kümmern ! Zurzeit besteht hier noch ein grosses Desiderat. »

Recherche en éducation

- « Pour les activités de recherche et d'études, les notes et les résultats des élèves ne sont pas accessibles facilement (procédures lourdes pour les autorisations d'accès). »
- « Manque de ressources humaines (et de temps) pour une exploitation plus approfondie. »
- « Ich finde es langweilig, was in der Bildungsforschung gemacht wird. »
- « Der Forschung sollten weniger Hürden in den Weg gestellt werden ! »
- « Datenschützer sind reine Verhinderer – mich interessieren personenbezogene Datensätze nicht. »
- « Lösung : Datenschützer sollten sensibilisiert werden, dass Daten für die Forschung anonymisiert werden (können). » « Grundsätzlich sind Datenschützer eine wichtige Institution. »

- « Item Banking ist die technologische Zukunft. »
- « Wir stehen mit dem Ganzen noch am Anfang, es geht noch viel mehr ! – Pädagogische und politische Fragen werden noch kommen. »
- « Psychometrische Forschung wird in der Schweiz vernachlässigt. »
- « Es sollte nicht überreglementiert werden ! Ansonsten wird Forschung und damit Innovation verhindert. »
- « Es braucht Richtlinien für Transparenz. Wie bei Creative Commons (CC) ein einfaches System, um Nutzungsbedingungen für Daten anzugeben. »
- « Es sollte eine informierte Entscheidung (Informed Choice) sein, Daten der Forschung zur Verfügung zu stellen. »
- « Eineindeutige Identifikatoren (AHV-Nr./Fides) sind aus wissenschaftlicher Sicht ein Traum. Sie ermöglichen Längsschnittanalysen. »
- « Das zukünftige Schlachtfeld sind Klassencockpit, Mindsteps, Lernnavi etc. »
- « Schulrankings (vgl. USA) sind der blanke Horror. »

3.5 Conclusion : l'espace suisse de formation en tant qu'espace de données dispersées – absence d'une culture d'utilisation des données

Dans le système éducatif suisse, la formation est dispensée de plus en plus de manière numérique, ce qui entraîne une production continue de données à tous les niveaux. Tant les élèves que les enseignants et le personnel administratif produisent quotidiennement des données et sont observés pendant qu'ils génèrent ces données – ce qui crée des données supplémentaires. Toutefois, les acteurs individuels de la formation et le système éducatif ne procèdent guère encore à une exploitation et à une évaluation systématiques de ces données jusqu'à présent. Sauf lors de recensements systématiques qui sont réalisés dans le cadre du monitoring de la formation ou, le cas échéant, par le biais de prestataires de services qui développent ultérieurement leurs produits sur la base de ces évaluations de données.³⁶

Cette absence d'exploitation systématique des données de l'éducation peut s'expliquer par les aspects énumérés ci-après – qui sont en lien avec cette dernière et se déterminent mutuellement (ils ont déjà été en partie pris pour thèmes de discussion par différents acteurs de la formation dans le chapitre précédent) :

- Absence de prise de conscience (*Data Awareness*)
- Absence d'infrastructures et de compétences (*Data Literacy*)
- Points de vue étroits et simplificateurs (utiliser des données = se contenter de respecter la loi sur la protection des données)
- Peur de ce qui est inconnu et incontrôlable
- Concentration sur les aspects de protection et de sécurité³⁷
- Résolution de problèmes individuels
- Pensée en « silos » (tant au sein des différents contextes d'utilisation qu'au-delà de ces derniers)
- Cadre inadéquat ou absence de cadre

³⁶ Selon leurs propres déclarations, cela ne se fait jusqu'à présent qu'à petite échelle et dans une faible mesure au sein des maisons d'édition suisses de matériel d'enseignement.

³⁷ Voir également à ce sujet *FIDES – Résultats de la consultation relative au projet de concept* (2016) ; *Écosystème E-ID dans la formation – rapport partiel 3 : analyse de la SITUATION ACTUELLE* (2017) ; entretiens de la CDIP avec les secrétaires cantonaux des départements de l'instruction publique ainsi qu'avec d'autres parties prenantes dans le cadre de l'adaptation de la stratégie TIC de 2007 (rapport interne à l'attention de la CdC du 8 mars 2018).

Tous ces aspects dénotent l'absence d'une culture (d'utilisation) des données tant dans les établissements scolaires que dans le système éducatif global. Or une telle culture est simultanément une condition préalable de base et l'objectif poursuivi en vue d'une utilisation systématique des données dans le système éducatif. Toutefois, une telle approche nécessite de changer d'optique : les analyses de données doivent être considérées comme quelque chose de positif, à savoir comme un instrument (efficient) qui apporte son appui aux élèves, au corps enseignant, aux écoles, à la planification et au pilotage de l'éducation ainsi qu'aux autres acteurs du système éducatif. Mais ces acteurs ne peuvent le faire que si l'on peut simultanément exclure des effets dommageables ou nuisibles de l'utilisation des données pour l'individu et pour le système dans son ensemble. Pour ce faire, il faut créer un cadre correspondant où il s'agit de penser ensemble la totalité des différents éléments du système (voir chapitre 3.3), de procéder à leur pondération mutuelle (afin de trouver un niveau de régulation approprié) et de les harmoniser entre eux.

Dans les lignes qui suivent, nous ferons un pas dans cette direction en expliquant des concepts clés pour l'utilisation des données dans le contexte des défis qui se posent actuellement et de leur importance en vue de l'élaboration d'une future politique d'utilisation des données dans l'éducation et la formation.

4 Discussion de certains concepts clés

Le présent chapitre explicite des bases conceptuelles d'importance déterminante en vue de l'élaboration d'une politique d'utilisation des données pour le système éducatif d'un point de vue technique. Ce faisant, nous reprenons en tant que « concepts » des conditions cadres en place et des exigences posées à l'utilisation des données, nous définissons leur contenu en les inscrivant dans l'horizon de la recherche actuelle en la matière, et nous les interprétons pour le contexte de l'éducation.

Dans le présent contexte, nous entendons par « concept » une expression conceptuelle clé qui est ancrée à la fois dans la discussion technique et dans la discussion politique autour de la notion d'utilisation des données et qui est utilisée actuellement. Dans ce sens, les concepts permettent de relier des situations de fait à des fixations d'objectifs. Ils forment une architecture conceptuelle qui offre une orientation aux intéressés, et qui justifie, ouvre et limite les marges de manœuvre qui existent pour des prises de décision – et qui, de ce fait, donne de la substance à l'élaboration déjà commencée d'une politique d'utilisation des données pour le système éducatif suisse.

Dans le domaine juridique, nous éclairons les concepts suivants :

- Autodétermination informationnelle
- Le fait de lier des données à une destination précise
- Utilisation minimale des données (ou minimisation des données)
- Portabilité des données
- Propriété des données
- Accès aux données

De plus, les notions suivantes sont aussi traitées en tant que concepts :

- Sécurité de l'information
- Interopérabilité
- *Open Data*

Les chapitres consacrés à ces concepts ont été rédigés par des spécialistes en tant qu'articles spécialisés. L'auteure et les auteurs indiqués ici assument la responsabilité des résultats de leurs recherches et de leurs conclusions, et leurs opinions ne doivent pas nécessairement coïncider avec celles qui sont défendues dans le présent rapport. educa.ch remercie Florent Thouvenin, Alfred Früh, Aurelia Tamò-Larrieux, Martin Leuthold, Hannes Lubich, Andreas Kellerhals et Matthias Stürmer pour leur collaboration.

4.1 Concepts juridiques fondamentaux

Center for Information Technology, Society, and Law (ITSL), Université de Zurich

- Prof. Dr Florent Thouvenin
- Dr Alfred Früh
- Dr Aurelia Tamò-Larrieux

Table des matières

4.1.1	Introduction et objet du mandat	48
4.1.2	Exemples d'utilisation des données	49
4.1.2.1	Contrôles de performance	49
4.1.2.2	Mindsteps	49
4.1.2.3	LehrerOffice	50
4.1.3	Contexte régulateur	50
4.1.3.1	Classification des concepts juridiques	50
4.1.3.2	Le droit de la protection des données en particulier	51
4.1.4	Analyse des concepts	53
4.1.4.1	Autodétermination informationnelle	53
4.1.4.2	Le principe de finalité	54
4.1.4.3	Minimisation des données	56
4.1.4.4	Portabilité des données	57
4.1.4.5	« Propriété des données »	57
4.1.4.6	Accès aux données	59
4.1.5	Infrastructures visant à renforcer les concepts	60
4.1.5.1	Personal Information Management Systems (PIMS) et modèles MyData	61
4.1.5.2	Systèmes d'information de l'éducation	62
4.1.6	Recommandations	62
4.1.6.1	Recommandations de principe	62
4.1.6.2	Propositions concrètes	64

4.1.1 Introduction et objet du mandat

Afin d'améliorer les processus d'enseignement et d'apprentissage, les données de l'éducation³⁸ font toujours plus souvent l'objet d'une numérisation ou d'une saisie électronique. Jusqu'ici, leur augmentation inexorable a surtout été considérée comme un risque à la lumière du droit de la protection des données, qui est également applicable dans ce domaine. Peu à peu, on constate qu'elles recèlent un immense potentiel, utilisable par divers acteurs à diverses fins. Ce point de vue a donné lieu à la présente contribution.

Si toutefois l'importance de ces données doit augmenter, il convient d'analyser les risques liés à leur utilisation, comme l'accès illicite aux données des apprenants, et de discuter des solutions possibles. Il s'agit d'harmoniser leur utilisation croissante avec d'autres objectifs comme la protection de la sphère privée et l'amélioration de la sécurité des données. On vise une politique cohérente d'utilisation des données qui respecte les intérêts en présence dans le contexte de l'éducation. Cette cohérence régulatrice est d'ailleurs également au centre de la politique des données du Conseil fédéral³⁹.

La première tâche de la présente contribution consiste à exposer les *concepts juridiques fondamentaux* déterminants dans le contexte de la politique de l'utilisation des données de l'éducation, leur contenu et leur articulation⁴⁰. Nous partons des sources scientifiques et renvoyons aux solutions techniques qui servent à les implémenter. L'accent repose sur le droit suisse et le contexte cantonal de l'éducation⁴¹. Les développements juridiques internationaux sont également pris en compte pour autant qu'ils servent à la compréhension.

Au-delà de cette première tâche, il convient également de faire des *propositions pour le développement* d'une politique d'utilisation des données cohérente. Il faut ici émettre deux réserves d'un point de vue scientifique, car il n'y a pas de définition claire ni du point de départ ni du point d'arrivée de la réflexion : premièrement, les recommandations sur l'évolution d'une politique d'utilisation des données cohérente pré-supposent la connaissance de l'état actuel de l'utilisation des données dans le paysage suisse de l'éducation. En raison de la structure fédéraliste du système éducatif, on ne voit pas clairement qui sont les acteurs⁴² qui saisissent, contrôlent, traitent et utilisent quelles données⁴³, à quel échelon administratif⁴⁴ ou quel degré scolaire⁴⁵ ; il convient donc de généraliser. Deuxièmement, les propositions en vue d'une telle politique exigent d'esquisser des objectifs concrets dans une certaine mesure. La cohérence du cadre juridique à elle seule n'est pas encore un objectif. C'est pourquoi, le présent rapport ne pourra pas (encore) répondre à la question de savoir quels buts concrets poursuit une politique d'utilisation des données et comment équilibrer raisonnablement dans ce contexte les intérêts et le besoin de protection des parties prenantes. D'abord, en connaissance des concepts juridiques analysés ici, il convient de formuler des objectifs plus exacts. On peut s'inspirer de situations réelles pour déterminer les développements nécessaires à une utilisation des données à haute valeur ajoutée⁴⁶. Vu ces deux imprécisions, la présente contribution ne peut pas éviter des généralisations, qui ne seront pas forcément pertinentes dans le cas d'espèce.

38 Voir chapitre 2, p. 15.

39 Office fédéral de la communication (OFCOM), Stratégie « Suisse numérique », Septembre 2018, <<https://www.bakom.admin.ch/bakom/fr/page-daccueil/suisse-numerique-et-internet/strategie-suisse-numerique.html>>, 30.

40 La présente contribution se fonde sur le contrat conclu le 14 décembre 2018 entre educa.ch et l'ITSL. Le rapport répond aux contenus et questions du ch. 2.1 mais les catégorise en partie autrement. En outre, il inclut les documents de référence cités sous ch. 2.2 ainsi que les recherches du Centre d'information et de documentation (IDES) du Secrétariat général de la Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP) sur les bases légales cantonales du système suisse de l'éducation.

41 En raison de la quantité de normes juridiques cantonales, la contribution qui suit examinera essentiellement les législations des cantons d'Argovie, Tessin, Soleure, Vaud et Zurich. Elle recourra à d'autres lois et ordonnances cantonales si nécessaire.

42 Par ex. les élèves et les membres du corps enseignant, des directions d'école, ainsi que du personnel administratif, d'offices statistiques et d'entreprises privées.

43 Données personnelles sur les élèves, les membres du corps enseignant et de la direction d'école, données administratives, d'application ou d'enquête.

44 Sur le plan fédéral, cantonal ou communal.

45 Degré primaire, secondaire et tertiaire, perfectionnement.

46 Des exemples de tels contenus sont l'utilisation à l'école d'outils pédagogiques de fournisseurs privés, l'utilisation d'appareils privés dans ou pour l'enseignement, la gestion des évaluations de prestation des élèves, l'échange de données avec d'autres prestataires comme les services du médecin ou du psychologue scolaire, etc.

4.1.2 Exemples d'utilisation des données

Afin toutefois de donner des points de départ suffisamment concrets pour un développement, la présente contribution se base concrètement sur les trois exemples suivants d'utilisation des données⁴⁷ en tant que représentatifs des questions qui se posent aujourd'hui :

4.1.2.1 Contrôles de performance

Les contrôles de performance (checks) sont des tests de performance exécutés dans le secteur de l'éducation de la Suisse du Nord-Ouest, c'est-à-dire dans les cantons d'Argovie, Bâle-Campagne, Bâle-Ville et Soleure, au commencement des degrés primaires 3 et 6 ainsi qu'à la moitié du 2^e et à la fin du 3^e degré secondaire I. Ils sont dirigés selon des directives standardisées par le personnel enseignant ; afin d'alléger la charge de celui-ci et de garantir l'indépendance⁴⁸, un service externe, l'Institut für Bildungsevaluation der Universität Zürich (« IBE »)⁴⁹, est chargé de leur évaluation. Associé à l'Université de Zurich et juridiquement indépendant⁵⁰, celui-ci est traité comme une entreprise privée. En ce qui concerne les contrôles, il travaille en collaboration avec divers partenaires.

Les contrôles visent trois objectifs⁵¹. Premièrement, ils rendent visible et comparable le succès scolaire individuel des élèves dans le secteur de l'éducation de la Suisse du Nord-Ouest (bilan des performances scolaires). Les autorités peuvent prendre sur cette base des mesures de promotion et des décisions d'orientation. Deuxièmement, ils forment une partie du certificat de fin d'études du degré secondaire que les apprenants peuvent joindre à leurs postulations. Troisièmement, ils offrent une aide à l'orientation pour le développement de l'enseignement et de l'école aux écoles primaires ou secondaires cantonales.

Conformément à la diversité des objectifs, divers acteurs obtiennent un accès aux données générées par les contrôles, y compris les données d'évaluation⁵² : les élèves ou les titulaires de l'autorité parentale apprennent par l'enseignant compétent compétent-e les résultats de leur test avec indication de la moyenne. En sus des résultats individuels de leurs élèves, les membres du corps enseignant obtiennent les résultats anonymisés des autres classes et écoles (moyenne des quatre cantons), à l'aide desquels ils peuvent déduire le résultat global de leurs classes et écoles. La moyenne des quatre cantons est également mise à la disposition des directions d'écoles. Les autorités de surveillance cantonales compétentes obtiennent uniquement les résultats globaux de leur propre école ainsi que des informations sur la moyenne des quatre cantons. Enfin, les résultats de toutes les classes et écoles sont agrégés et rendus accessibles au public⁵³.

4.1.2.2 Mindsteps

Développé par l'IBE sur mandat du secteur de l'éducation de la Suisse du Nord-Ouest, Mindsteps⁵⁴ est un recueil numérique de devoirs élaboré pour des classes du 3^e degré primaire au 3^e degré secondaire. Parmi un ensemble de 25 000 unités, les devoirs peuvent être soit choisis par l'enseignant et pour toute la classe ou pour des élèves individuels, soit choisis directement par un ou une élève et exécuté individuellement à l'ordinateur. L'apprenant obtient en retour une information sur ses résultats et le progrès de son apprentissage.

Mindsteps vise à montrer aux enseignants ainsi qu'aux élèves eux-mêmes l'évolution de l'apprentissage et de la performance. À partir d'un premier état des lieux des compétences acquises, ils peuvent constater les

47 Ces exemples se fondent sur des informations rendues accessibles au public.

48 Voir parag. 4 et 6 du Règlement über die kantonalen Schulleistungsprüfungen (Leistungschecks) du canton de Soleure (ci-après : Leistungschecks-Règlement SO) ; Bildungsbericht Nordwestschweiz, 2017, 156.

49 Voir IBE, Checks und Aufgabensammlung, <<https://www.ibe.uzh.ch/de/projekte/checks-aufgabensammlung.html>>.

50 Paragr. 1^{er} du « Règlement für Assoziierte Institute der Universität Zürich » du 19 décembre 2005.

51 Paragr. 3, al. 2, Leistungschecks-Règlement SO ; voir aussi Département für Bildung und Kultur Solothurn, Checks und Aufgabensammlung, <<https://www.so.ch/verwaltung/departement-fuer-bildung-und-kultur/volksschulamt/lernen-sichtbar-machen/checks/checks-und-aufgabensammlung/>>.

52 Voir parag. 7 Leistungschecks-Règlement SO.

53 Voir Ergebnisberichte der Leistungschecks, <<https://www.so.ch/verwaltung/departement-fuer-bildung-und-kultur/volksschulamt/lernen-sichtbar-machen/checks/ergebnisberichte/>>. Voir aussi parag. 8, al. 1^{er}, Leistungschecks-Règlement SO où il est expressément stipulé que « la publication de résultats de tests, qui permettent des déductions sur la situation d'élèves, de membres du corps enseignant et d'écoles » est illicite (dass die « Veröffentlichung von Testergebnissen, die Rückschlüsse auf einzelne Schüler und Schülerinnen, Lehrpersonen und Schulen ermöglichen » unzulässig ist).

54 Voir sur la conception et les objectifs de Mindsteps, Concept, <https://www.mindsteps.ch/konzept/>.

progrès de l'apprentissage sur toute la période allant du 3^e degré primaire au 3^e degré secondaire I et déterminer les étapes nécessaires à venir. Ce programme permet d'une part d'adapter l'apprentissage et le contrôle à l'élève et d'autre part de promouvoir l'apprenant d'après ses compétences.

Les licences d'utilisation de Mindsteps sont attribuées par l'IBE. La déclaration de protection des données de l'Institut indique les données et le but de la saisie⁵⁵.

4.1.2.3 LehrerOffice

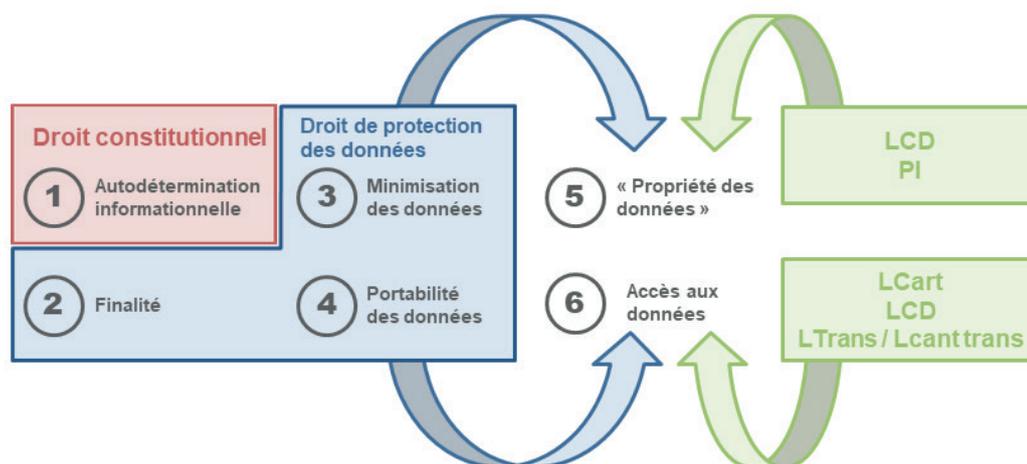
LehrerOffice est un logiciel avec lequel les membres du corps enseignants et les directions d'école organisent la gestion des notes, les travaux de routine, les plans individuels de promotion et la planification hebdomadaire⁵⁶. Il permet de concevoir les plans des matières, les questionnaires d'évaluation, les plans de promotion, les rapports d'évaluation et les contrôles ainsi que d'établir automatiquement les certificats ; il vise à soutenir les membres du corps enseignant dans leur travail quotidien via la gestion et l'enregistrement centralisés des données des élèves⁵⁷. Comme solution globale, il offre des applications tout autant mobiles que bureautiques, dont les fonctionnalités peuvent être adaptées à l'importance et aux besoins de l'école⁵⁸. Les cantons publient sur sa page Internet des instructions sur l'utilisation de la base de données (par ex. pour la saisie des certificats)⁵⁹.

LehrerOffice est un produit de l'entreprise privée Roth Soft AG, domiciliée dans le canton de Zurich. Il offre la possibilité d'enregistrer et de traiter les données soit localement, soit dans des banques de données qui permettent l'accès à plusieurs utilisateurs. Les écoles qui n'ont pas de serveur propre ou pas l'infrastructure pour des travaux indépendants dans la base de données peuvent obtenir des capacités supplémentaires sur le serveur du fournisseur⁶⁰.

4.1.3 Contexte régulateur

4.1.3.1 Classification des concepts juridiques

Le droit de la protection des données forme le pilier juridique principal de la politique d'utilisation des données cohérente visée. Mais l'objet de régulation compris dans cette dernière dépasse le cadre du premier, ce qu'on peut représenter à l'aide des six concepts suivants :



Graphique 9 : localisation normative et influence réciproque des concepts

55 Dispositions de protection des données de Mindsteps, <<https://www.mindsteps.ch/footer/datenschutz/>>.

56 LehrerOffice a été adapté pour 20 cantons alémaniques, <<https://www.lehreroffice.ch/wiki/doku.php?id=lehreroffice:desktop:kantone>>.

57 Instructions LehrerOffice, <<https://www.lehreroffice.ch/wiki/doku.php?id=lehreroffice:desktop:datenbank:datenbank>>.

58 Description LehrerOffice Desktop, <<https://www.lehreroffice.ch/lehreroffice-desktop/beschreibung/>>.

59 Voir par ex. les instructions sur les certificats avec le modèle du canton de Soleure, <<https://www.lehreroffice.ch/wiki/doku.php?id=lehreroffice:desktop:kantone:so:so>>.

60 Description LehrerOffice DB-Hosting, <<https://www.lehreroffice.ch/lehreroffice-db-hosting/beschreibung/>>.

Le concept de l'autodétermination informationnelle (1.) est d'une importance prioritaire. En général, il découle directement de la Constitution et forme le fondement conceptuel du droit de la protection des données. L'autodétermination informationnelle ne se rapporte qu'aux données personnelles et trouve son expression dans des droits plus concrets du droit de protection des données.

Les principes de la finalité (2.) et de la minimisation des données (3.) sont exclusivement inscrits dans le droit de la protection des données. C'est pourquoi, ils ne trouvent application que lorsque le traitement porte sur des données personnelles. La doctrine et la jurisprudence ne font aucune proposition pour les appliquer au-delà du droit de protection des données.

Le droit à la portabilité de données (4.) est inscrit comme tel uniquement dans la législation de l'UE. Mais on peut également déduire des droits analogues de la législation suisse. Dans le sens compris ici, il ne se rapporte qu'aux données personnelles⁶¹.

Les concepts appelés propriété des données (5.) et accès aux données (6.) n'ont pas de base légale directe bien que certains de leurs éléments découlent du droit de la protection des données. Ils sont toutefois en rapport étroit avec d'autres dispositions. Le concept de la propriété des données intervient avant tout dans la loi sur la concurrence déloyale (LCD) ainsi que dans les lois spéciales sur la propriété intellectuelle (PI), à savoir le droit d'auteur et le droit des brevets, alors que la notion d'accès aux données se trouve dans la loi sur les cartels (LCart), la loi sur la transparence de la Confédération (LTrans) et les lois sur la transparence des cantons.

4.1.3.2 Le droit de la protection des données en particulier

Comme d'autres matières juridiques, le droit de la protection des données est soumis à d'incessantes mutations technologiques et sociales. La nécessaire adaptation qui en résulte se reflète sur le plan européen par l'introduction en mai 2018 d'un règlement général sur la protection des données (RGPD) et en Suisse par la révision en cours de la loi sur la protection des données (LPD). Les deux sont étroitement liés. La révision de la LPD devra concrètement s'appuyer sur le RGPD dans le seul but d'atteindre un niveau de protection comparable⁶². Malgré ce rapprochement attendu, le RGPD n'a qu'un rôle marginal dans la présente contribution. Certes, ses normes sont importantes pour les fournisseurs privés suisses de prestations d'éducation pour autant qu'ils offrent leurs services à des personnes situées dans l'Union européenne (UE). Mais il ne joue aucun rôle en matière d'utilisation des données par les écoles primaires et secondaires communales et cantonales. C'est pourquoi, la présente contribution se limite à la situation juridique de la Suisse.

En Suisse, la LPD vise à protéger sur le plan fédéral la personnalité et les droits fondamentaux des personnes qui font l'objet d'un traitement de données (art. 1^{er} LPD). La LPD est applicable d'une part aux acteurs privés comme l'IBE⁶³, juridiquement indépendant, ou l'entreprise Roth Soft AG, et d'autre part, aux organes de la Confédération comme l'Eidgenössische Technische Hochschule (ETH) de Zurich ou l'École polytechnique fédérale de Lausanne (EPFL). La LPD n'est applicable que lorsqu'il y a traitement de données personnelles⁶⁴.

Sur le plan cantonal également il existe des lois sur l'information et la protection des données. Bien que l'accent porte ici sur le traitement des données personnelles par les organes de l'administration cantonale comme les écoles ou les universités, les lois cantonales de protection des données réglementent souvent la gestion des informations en général et comprennent expressément les données factuelles.

Les lois de protection des données de la Confédération et des cantons énoncent des règles générales pour le traitement des données personnelles. Les données sensibles et les profils de personnalité en font égale-

61 Pour autant qu'il s'agisse des droits de portabilité, il convient de classer ceux-ci sous le concept de l'accès aux données (chapitre 4.1.4.5, p. 57).

62 Dans ce sens, voir le message E-LPD, FF 2017, 6565, 6620.

63 Faute d'accomplir des tâches publiques au sens du paragr. 1^{er} en lien avec le paragr. 3, let. c, de la loi sur l'information et la protection des données du canton de Zurich (Gesetz über die Information und den Datenschutz, IDG, ci-après : IDG ZH), l'IBE est soumis à la LPD et non à l'IDG ZH cantonale ; voir à ce sujet les dispositions de protection des données, <https://www.mindsteps.ch/footer/datenschutz/>.

64 La notion de « traitement » s'entend au sens large et comprend tant au niveau cantonal que fédéral la saisie, l'enregistrement, le rattachement et la suppression de données (sur le plan fédéral, voir l'art. 3, let. e, LPD « toute opération relative à des données »).

ment partie ; c'est pourquoi ils sont en principe soumis aux mêmes règles générales⁶⁵, voire à des dispositions plus sévères. Les premières sont par exemple des informations des services du médecin ou du psychologue scolaire sur un apprenant. En soi, les notes n'en sont pas, mais leur collecte sur une certaine période constitue un profil de personnalité des apprenants qui tombe sous le coup des dispositions plus sévères. Il convient de noter que la définition des lois cantonales sur les données personnelles et le profil de personnalité peut varier⁶⁶.

Du fait que les lois de protection des données se rattachent aux données personnelles, la différence entre données personnelles et factuelles est importante, mais elle n'est pas toujours évidente. La plupart du temps, des identificateurs comme le nom d'un ou une élève, une adresse IP⁶⁷, une adresse Media-Access-Control (MAC)⁶⁸ d'un appareil ou une adresse postale sont considérées comme des données personnelles. Mais pour une donnée comme le sexe, l'âge ou le statut socio-économique, la qualification dépend de la possibilité de mettre celle-ci en relation avec d'autres données à disposition, ce qui permet de tirer des conclusions sur une personne déterminée.

Dès que les données sont anonymisées, elles sortent du domaine d'application des lois de protection des données. L'anonymisation est un processus par lequel la référence à la personne a été durablement supprimée de sorte que l'identification n'est plus possible ou seulement avec un effort disproportionné⁶⁹. La distinction entre des données personnelles et des données anonymisées nécessite toujours un examen au cas par cas ; il n'y a pas de normes juridiques déterminant le degré d'anonymisation. Mais à l'ère de Big Data, on peut douter de la permanence de l'anonymisation car la combinaison de divers supports de données permettra toujours une identification.

On peut se demander si le cryptage permet d'anonymiser les données par rapport à un tiers qui n'y a pas accès. Bien que la situation juridique ne soit pas claire à ce sujet, il faut partir du principe que les codes permettent de rétablir en tout temps le rattachement à la personne de sorte que les données personnelles persistent⁷⁰.

Le traitement de données personnelles est en principe autorisé en Suisse pour autant qu'il ne porte pas une atteinte illicite à la personnalité des intéressés. Tel est le cas des acteurs privés comme l'IBE ou l'entreprise Roth Soft AG s'ils respectent les principes généraux du traitement des données inscrits à l'art. 12, al. 2, let. a, LPD, ou qu'ils sont au bénéfice d'un motif justificatif au sens de l'art. 13 LPD. En outre, les données sensibles ou les profils de personnalité ne peuvent être communiqués à des tiers que s'il existe un motif justificatif (art. 12, al. 2, let. c, LPD). Le traitement des données personnelles et des données sensibles ainsi que des profils de personnalité par des organes fédéraux ou des unités des administrations cantonales, par ex. des écoles ou des universités, nécessite toujours une base légale⁷¹. Celle-ci doit être formulée avec suffisamment de clarté de sorte que ses destinataires puissent en déduire un comportement réel déterminé⁷².

65 Voir l'art. 3, let. a et c, LPD, ainsi que le dossier d'information, notes de bas de page 1, 9, pour une courte introduction aux notions de données personnelles, données sensibles et profil de personnalité.

66 Alors que le paragr. 6, al. 3, de la loi sur l'information et la protection des données du canton de Soleure (Informations- und Datenschutzgesetz, ci-après : InfoDG SO), l'art. 4, al. 2, de la loi sur la protection des données personnelles du canton du Tessin (legge sulla protezione dei dati personali, ci-après : LPDP TI), l'art. 4, al. 1^{er}, ch. 2, de la loi sur la protection des données personnelles du canton de Vaud (ci-après : LPrD VD), et le paragr. 3 IDG ZH s'en tiennent généralement à la lettre de la LPD, le paragr. 3, al. 1^{er}, let. k, de la loi sur l'information du public, la protection des données et les archives du canton d'Argovie (ci-après : IDAG AG) définit les données sensibles de manière toute générale comme « des données qui présentent un risque particulier d'atteinte à la personnalité en raison de leur signification, de la relation, du but ou du genre de traitement, de la catégorie de données ou d'autres circonstances (« Daten, bei denen aufgrund ihrer Bedeutung, des Zusammenhangs, Zwecks oder der Art der Bearbeitung, der Datenkategorie oder anderer Umstände eine besondere Gefahr einer Persönlichkeitsverletzung besteht »).

67 Suite de chiffres qui permettent d'identifier sans équivoque tout ordinateur dans un réseau (par ex. Internet).

68 Appelée aussi adresse physique ou adresse d'appareil. Il s'agit de l'identificateur unique d'un appareil.

69 Voir dossier d'information, notes de bas de page 1, 9 ; explications du Préposé fédéral à la protection des données et à la transparence (PFPDT), Protection des données et recherche en général, <www.edoeb.admin.ch/edoeb/fr/home/protection-des-donnees/statistik--register-und-forschung/recherche/protection-des-donnees-et-recherche-en-general.html>.

70 ROLF H. WEBER/FLORENT THOUVENIN, Gutachten zur Möglichkeit der Einführung eines Datenportabilitätsrechts im schweizerischen Recht und zur Rechtslage bei Personal Information Management Systems (PIMS), <https://www.itsl.uzh.ch/dam/jcr:26f84429-2aef-47b1-9ba1-6f6e8910c60d/180321%20BJ-Gutachten_final.pdf>, 73-74 avec renvoi à l'ATF 136 III 508, consid. 3.2.

71 Art. 17, al. 1^{er} et 2 LPD ; par ex. paragr. 8, al. 1^{er} et 2 IDAG AG, paragr. 15, al. 1^{er} et 2 InfoDG SO, art. 6, al. 1^{er}, LPDP TI, art. 5, al. 1^{er} et 2 LPrD VD, paragr. 8, al. 1^{er} et 2 IDG ZH.

72 SARAH BALLENEGGER, in : Maurer-Lambrou/Blechts (éd.), Commentaire bâlois, Datenschutzgesetz und Öffentlichkeitsgesetz, Bâle 2014, DSG Art. 17 N 7 avec renvoi à l'ATF 109 la 273, consid. 4d et ATF 124 I 40, consid. 3b.

Alors qu'en matière de collecte de données sensibles ou de profils de personnalité l'art. 14 LPD soumet les maîtres de fichiers privés à une simple obligation d'informer, cette obligation est applicable à toute collection de données personnelles par des organes fédéraux au sens de l'art. 18a, al. 1, LPD. Sur le plan cantonal, la situation juridique des obligations d'informer n'est pas uniforme. Ainsi, le canton d'Argovie prévoit une telle obligation lors de la collecte de données personnelles par des organes publics⁷³ tandis que dans le canton du Tessin, elle ne doit être respectée qu'en cas de saisie systématique⁷⁴.

4.1.4 Analyse des concepts

4.1.4.1 Autodétermination informationnelle

Par autodétermination informationnelle on comprend aujourd'hui le principe généralement reconnu selon lequel chaque personne doit pouvoir déterminer si et dans quel but les données qui la concernent peuvent être traitées et enregistrées⁷⁵.

Ce principe trouve son origine dans la jurisprudence et la doctrine allemandes. Le Tribunal constitutionnel allemand (Bundesverfassungsgericht) a défini cette notion dans son jugement dit du recensement (Volkszählungsurteil) de 1983 comme « le pouvoir de l'individu de déterminer lui-même la divulgation et l'utilisation de ses données personnelles »⁷⁶. Le Tribunal visait à faciliter le libre épanouissement de la personnalité et à maintenir la liberté de décision de l'individu en dépit de son exposition aux nouveaux processus de traitement des données⁷⁷. En Suisse, dans une décision rendue à la fin des années 80, le Tribunal fédéral a renvoyé à la jurisprudence du Tribunal constitutionnel allemand⁷⁸ et a reconnu que l'accès à des données personnelles par une autre personne pouvait mener à un sentiment de restriction qui empêche l'intéressé de se mouvoir librement, subissant ainsi une restriction de sa liberté personnelle. Au milieu des années 90, le Tribunal fédéral a déduit du droit de la protection de la personnalité du travailleur le droit d'un travailleur à consulter son dossier personnel et il a retenu que le droit à la consultation du dossier doit être compris « comme une partie du droit à l'autodétermination informationnelle »⁷⁹. Quelques années plus tard, il est parvenu à la conclusion qu'un droit à l'autodétermination informationnelle est reconnu par la doctrine et la jurisprudence⁸⁰. Aujourd'hui, ce droit est considéré comme un droit fondamental non écrit et un principe de rang constitutionnel en rapport avec le domaine de protection de l'art. 13 Cst. féd., bien que ladite notion n'ait jamais été inscrite dans la Constitution⁸¹.

Le principe de l'autodétermination informationnelle apparaît tout aussi peu dans la législation sur la protection des données, et ce, tant à l'échelon fédéral que dans les lois cantonales. Bien que la LPD entrée en vigueur en 1993 ait expressément visé le renforcement de la liberté de décision des personnes concernées⁸², il ne déploie ses effets sur le droit de la protection des données que de manière indirecte. À défaut de concrétisation à l'échelon légal, il ne figure par conséquent dans aucun objectif concret qui pourrait être directement mis en œuvre, ni dans le contexte de l'éducation ni dans d'autres domaines.

Si l'on en recherche quand même des expressions concrètes, le principe se reflète à l'échelon légal dans quatre instruments du droit de la protection des données : (a) dans le principe du consentement (ou dans

73 Paragr. 13, al. 1^{er}, IDAG AG, le paragr. 2, IDAG AG prescrivant des exceptions à l'obligation d'information.

74 Art. 9, al. 2, LPDP TI.

75 FLORENT THOUVENIN, *Wem gehören meine Daten ? Zu Sinn und Nutzen einer Erweiterung des Eigentumsbegriffs*, SJZ 113/2017, 21–32, 23 avec renvoi à l'ATF 140 I 2, E. 9.

76 BVerfGE, 65, 1.

77 STEFANIE-DANIELA WALDMEIER, *Informationelle Selbstbestimmung – ein Grundrecht im Wandel ?*, Zurich 2015, 6 ss.

78 ATF 113 Ia 1, consid. 4b.

79 ATF 120 II 118, consid. 3a.

80 ATF 122 I 153, consid. 6b.

81 THOUVENIN, n. 38, 23 avec renvoi à RAINER J. SCHWEIZER, in : Ehrenzeller/Schindler/Schweizer/Vallender (éd.), *Commentaire saint-gallois, Die Schweizerische Bundesverfassung, Zurich/Saint-Gall 2014*, Cst. art. 13 N 72 ; GIOVANNI BIAGGINI, BV, *Bundesverfassung der Schweizerischen Eidgenossenschaft, Kommentar, Zürich 2007*, Cst. art. 13 N 11 ss.

82 Message LPD FF 1988 421, 491 s. La révision en cours de la LPD paraît vouloir renforcer le principe de l'autodétermination informationnelle, comme le montre son inscription à l'art. 1^{er} du projet de loi dans le message P-LPD, FF 2017 6565. Mais la manière dont il sera repris concrètement dans la loi est encore incertaine.

le droit de le refuser), (b) dans le principe de la transparence, (c) dans diverses obligations d'informer du maître du fichier, et (d) dans le droit d'accès, qui sera examiné de manière approfondie ci-dessous⁸³.

a) Tandis qu'il agit dans la LPD comme un motif justificatif pour le traitement de données personnelles par des acteurs privés comme l'IBE ou l'entreprise Roth Soft AG, le consentement ne peut pas être présumé dans le contexte du droit public. L'interprétation stricte basée sur la lettre de l'IDG du canton de Zurich ne permet pas aux écoles cantonales, Hautes écoles ou à l'Université de s'en prévaloir pour le traitement de données personnelles⁸⁴. Mais la situation juridique n'est pas uniforme. Dans les cantons d'Argovie, de Soleure ou du Tessin, les personnes concernées peuvent le donner par exemple à des organes publics pour le traitement de leurs données⁸⁵. Même si elles ont la possibilité de décider par « oui » ou « non » de l'utilisation de leurs données par un tiers, l'ère du tout numérique fait bien vite apparaître les limites de ce moyen, destiné à renforcer leur liberté de décision. Les déclarations de protection des données ne sont en général pas écrites pour des laïcs, ce qui en gêne la lecture et la compréhension et affaiblit la valeur du consentement en tant que fondement du traitement des données⁸⁶.

b) L'on ne peut prétendre à l'autodétermination informationnelle des individus que si ceux-ci agissent en toute connaissance de cause. Dès lors, améliorer la transparence dans l'utilisation des données est une proposition régulièrement émise. Le principe de transparence a pris davantage d'importance dans l'UE depuis l'entrée en vigueur du RGPD qui le présente explicitement comme principe à la base du traitement des données (art. 5, al. 1^{er}, let. a, RGPD). Il vise à améliorer l'accessibilité, la compréhension (simplicité de la langue) et la précision des informations y relatives⁸⁷. En particulier, les tentatives de représenter les flux de données et les accès par des icônes prennent une grande importance. Celles-ci sont censées se généraliser⁸⁸. Mais ces efforts sont encore loin d'aboutir.

Le principe de transparence dans l'information et dans la communication des processus de traitement des données sert à concrétiser le principe de transparence⁸⁹. Mais aujourd'hui les exigences formulées sur le plan fédéral à l'égard de l'obligation d'informer sont rares⁹⁰, de même que sur le plan cantonal⁹¹.

c) Quant au droit d'accès, voir les explications sous Portabilité des données (chapitre 4.1.4.4, p. 57) et Accès aux données (chapitre 4.1.4.6, p. 59).

4.1.4.2 Le principe de finalité

Le principe de finalité est un principe général du droit de la protection des données. Il prescrit que les données personnelles ne doivent être traitées que dans le but qui est indiqué lors de leur collecte, qui est prévu par une loi ou qui ressort des circonstances (art. 4, al. 3, LPD). Il ne s'applique qu'aux données personnelles et ne vaut pas pour les données anonymisées.

Le principe de finalité poursuit deux objectifs. D'une part, le but du traitement des données doit être reconnaissable pour la personne concernée. D'autre part, les données ne peuvent pas être ultérieurement uti-

83 Voir Accès aux données (4.1.4.6, p. 59).

84 Paragr. 8, al. 1^{er}, IDG ZH. Celle-ci ne peut être légitimée à l'aide du consentement dans un cas d'espèce que lors d'une seconde utilisation, paragr. 9, al. 1^{er}, IDG ZH. Dans le canton de Vaud, le consentement est prévu uniquement en rapport avec le traitement de données sensibles (art. 5, al. 2, let. c, LPrD VD).

85 Paragr. 8, al. 1^{er}, let. c, IDAG AG ; paragr. 15, al. 1^{er}, let. d InfoDG SO ; toutefois, l'art. 6, al. 2, LPDP TI, ne prévoit le consentement que dans des cas d'espèce et non pour le traitement systématique des données.

86 C'est pourquoi le consentement est critiqué dans la doctrine, voir entre autres EVA MARIA BELSER, Zur rechtlichen Tragweite des Grundrechts auf Datenschutz : Missbrauchsschutz oder Schutz der informationellen Selbstbestimmung ?, in : Epiney/Fasnacht/Blaser, Forum Europarecht, Vol. N 30, Zurich 2013, 27 ss. ; FREDERIK ZUIDERVEEN BORGESIUS, Informed Consent : We Can Do Better To Defend Privacy, Privacy & Security IEEE, 13(2), 103-107, 104 ss.

87 Voir en particulier les considérants 39, 58, qui renvoient à l'utilisation d'éléments visuels, et 60 RGPD.

88 Voir par ex. Privacy Icons de Mozilla, <https://wiki.mozilla.org/Privacy_Icons>, Consent Management Icons de MyData, <<https://www.lvm.fi/documents/20181/859937/MyData-nordic-model/2e9b4eb0-68d7-463b-9460-821493449a63?version=1.0>>, 9, ou Privacy Nutrition Labels, <<https://cups.cs.cmu.edu/privacyLabel/>>.

89 Le chapitre III, al. 1^{er}, RGPD, expose les informations qui doivent impérativement être communiquées aux personnes concernées.

90 Ainsi, concernant l'obligation d'informer des acteurs privés lors de la collecte de données sensibles ou de profils de personnalité, l'art. 14, al. 2, LPD, prescrit qu'il faut communiquer à la personne concernée au moins l'identité du maître du fichier, les finalités du traitement prévu et les catégories des destinataires des données. Selon l'art. 18a, al. 2, LPD, les organes fédéraux doivent en outre informer la personne concernée du droit d'accéder aux données la concernant au sens de l'art. 8 LPD et des conséquences liées au refus de sa part de fournir les données personnelles demandées.

91 Alors que de nombreux cantons n'expliquent pas plus en détail l'obligation d'information, le paragr. 13, al. 1^{er}, IDAG AG énonce une liste du genre d'informations à communiquer aux personnes concernées. Mais l'al. 2 relativise cette obligation lorsque le traitement se fonde sur une base légale.

lisées dans un autre but (voir aussi l'art. 4, al. 4, LPD). La formulation étendue du but dans la base légale permet néanmoins de garantir la licéité du traitement. Les organes de la Confédération peuvent ainsi fonder ce dernier sur une décision démocratique du législateur⁹². Les privés comme l'IBE ou l'entreprise Roth Soft AG ne le peuvent pas.

Le principe de finalité n'existe pas seulement à l'échelon fédéral mais aussi dans les lois cantonales d'information et de protection des données⁹³. Or ces dernières le relativisent souvent en prévoyant expressément un but d'utilisation plus étendu que celui indiqué lors de la collecte de données personnelles⁹⁴. Ainsi, dans le contexte de l'éducation, l'objectif du traitement des données personnelles des élèves et des titulaires de l'autorité parentale, des membres du corps enseignant, des représentants de la direction de l'école, etc. résulte des diverses lois et ordonnances cantonales et communales qui prescrivent les missions de l'éducation. Des exemples typiques de ces missions sont les évaluations de performance des apprenants⁹⁵, la transmission d'informations concernant des caractéristiques déterminées des apprenants lors de l'orientation (changement de classe ou de degré)⁹⁶, le droit à l'information et à la consultation du dossier des parents⁹⁷, la gestion des données personnelles des élèves, des titulaires de l'autorité parentale et des membres du corps enseignant par la direction d'école aux fins d'administration, la communication avec les personnes concernées et la préparation des données à des fins de statistiques et de pilotage de la formation⁹⁸. À cet égard, les lois⁹⁹ et ordonnances¹⁰⁰ scolaires cantonales prévoient souvent que seules les données anonymisées peuvent être rendues accessibles public.

Par conséquent, lorsqu'un projet poursuit plusieurs objectifs, comme le contrôle de performance du secteur de l'éducation de la Suisse du Nord-Ouest qui forme la base des décisions de promotion et des certificats de fin d'études, et qui constitue un instrument de développement de l'enseignement et de l'école, il nécessite de la part de la législation cantonale une base légale formulée assez largement pour les saisir tous.

Il convient de noter qu'une évolution sur le plan européen relativise le principe de finalité pour autant que le traitement ultérieur vise « des fins archivistiques dans l'intérêt public, [...] des fins de recherche scientifique ou historique ou [...] des fins statistiques » (art. 5, al. 1^{er}, let. b, RGPD). Au sein de l'UE, le RGPD a notamment renforcé l'exception de l'utilisation à des fins de recherche.

En Suisse aussi, sur le plan fédéral, le traitement de données est légitime lorsqu'il vise des buts de recherche ou des évaluations statistiques et que la forme de la publication des résultats ne permet pas d'identifier les personnes concernées (art. 13, al. 2, let. e, LPD pour le traitement par des privés ; art. 22 LPD pour le traitement par des organes fédéraux). Sur le plan cantonal, les lois d'information et de protection des données autorisent le traitement des données dans des buts excluant les personnes pour la recherche, la planification et la statistique¹⁰¹. Un changement de but motivé par la recherche n'est toutefois permis que lorsque des conditions cumulatives sont remplies¹⁰². Bien que dans l'UE on doive aussi présenter des « garanties appropriées » (par ex. pour garantir la minimisation et la pseudonymisation des don-

92 URS MAURER-LAMBROU/ANDREA STEINER, in : Maurer-Lambrou/Blechts (éd.), Commentaire bâlois, Datenschutzgesetz und Öffentlichkeitsgesetz, Bâle 2014, DSG art. 4, N 15.

93 Voir paragr. 11 IDAG AG ; paragr. 16 InfoDG SO ; art. 7 LPDP TI ; art. 6 LPrD VD ; paragr. 9 IDG ZH.

94 Voir paragr. 11 IDAG AG ; paragr. 16 InfoDG SO ; art. 6 LPrD VD ; paragr. 9 IDG ZH.

95 Voir par ex. paragr. 26 ss. Schulgesetz et paragr. 26 Verordnung über die Volksschule du canton d'Argovie.

96 Voir par ex. paragr. 14 Mittelschulgesetz du canton de Soleure ; paragr. 51 Schulgesetz du canton d'Argovie ; art. 26 Landsschulkommissionsbeschluss zum Schulgesetz du canton d'Appenzell Rhodes-Intérieures.

97 Voir par ex. paragr. 21 ss. Verordnung über die Volksschule du canton d'Argovie ; art. 47 et 57 Landsschulkommissionsbeschluss zum Schulgesetz du canton d'Appenzell Rhodes intérieures ; art. 33a Landsschulkommissionsbeschluss zur Gymnasialverordnung du canton d'Appenzell Rhodes-Intérieures.

98 Voir par ex. paragr. 8 bis Schulgesetz du canton de Soleure et paragr. 18 Verordnung über die Volksschule du canton d'Argovie.

99 Voir par ex. paragr. 57c Schulgesetz du canton de Bâle-Ville et art. 103 de la loi sur la scolarité obligatoire du canton de Fribourg.

100 Voir par ex. paragr. 28 Verordnung über die Volksschule du canton d'Argovie.

101 Voir paragr. 19, al. 1^{er}, IDAG AG, paragr. 16, al. 3, InfoDG SO ; art. 15, al. 1^{er}, LPDP TI ; art. 24, al. 1^{er}, LPrD VD.

102 Le traitement ne s'effectue pas par rapport à des personnes, il se fonde sur une base légale (par ex. la loi sur la statistique fédérale [LSF] qui, selon son art. 3, al. 1^{er}, fournit également des résultats sur l'état et l'évolution de la formation), les données personnelles sont aussitôt anonymisées, la transmission à des tiers a été autorisée par l'organe fédéral et la publication s'effectue uniquement sous forme anonymisée. Voir URS MAURER-LAMBROU/SIMON KURZ, in : Maurer-Lambrou/Blechts (éd.), Commentaire bâlois, Datenschutzgesetz und Öffentlichkeitsgesetz, Bâle 2014, DSG Art. 22, N 2 ss. Les loi cantonales, comme le paragr. 19, al. 1^{er}, IDAG AG, l'art. 24, al. 1^{er} LPrD VD ou l'art. 15, al. 1^{er} LPDP TI, suivent la même conception que l'art. 22 LPD.

nées), l'exception de l'utilisation en faveur de la recherche bénéficie d'une reconnaissance élevée¹⁰³ : elle peut expressément entraîner des restrictions à l'exercice des droits individuels des personnes concernées¹⁰⁴ et faire l'objet d'élargissements dans les lois nationales des États membres qui accentuent encore son rôle.

Il convient de discuter de l'utilisation sur les plans fédéral et cantonal des données de l'éducation à des fins scientifiques et en définitive aux fins de planification et de pilotage de l'éducation ; comme pour l'exception de l'utilisation en faveur de la recherche applicable dans l'UE, on pourrait réfléchir lors de la révision de la LPD à l'opportunité d'accorder davantage d'importance à la recherche via une extension de l'exception. Le cas échéant, le message sur la LPD pourrait en outre préciser que l'exception en faveur de la recherche comprend aussi le domaine de l'éducation.

4.1.4.3 Minimisation des données

Le principe de minimisation des données¹⁰⁵ prescrit que le traitement des données personnelles ne peut intervenir que lorsqu'il est nécessaire pour atteindre son but. Il n'y a pas lieu de traiter des données inappropriées. Si possible, il convient d'utiliser des données anonymisées conformément au principe « aussi peu que possible, autant que nécessaire ». Dédit du principe de la proportionnalité (art. 4, al. 2, LPD) et inscrit dans la Constitution fédérale (art. 5 Cst. féd.), ce concept juridique se compose des trois éléments de l'aptitude, de la nécessité et de la proportionnalité au sens étroit.

Les lois cantonales de protection des données peuvent concrétiser davantage ce principe. Ainsi, selon l'IDG du canton de Bâle-Ville, il convient d'anonymiser ou de pseudonymiser les données dès que cet effort est approprié au but de protection visé¹⁰⁶. Dans le canton de Zurich, les organes publics doivent concevoir les systèmes de traitement des données de manière à saisir le moins possible de données personnelles et à supprimer, anonymiser ou pseudonymiser celles-ci dès que possible¹⁰⁷. Dans le canton d'Argovie, si elle ne précise pas la manière dont il faut réaliser le principe de minimisation, l'IDAG rattache ce principe directement au principe de proportionnalité¹⁰⁸.

Comme la conception concrète du principe de minimisation des données dépend de chaque constellation, il est difficile d'en définir préalablement des normes. Il convient d'examiner dans chaque cas d'espèce la proportionnalité du traitement car celle-ci nécessite une pesée des intérêts. L'évaluation des apprenants, par ex. dans le contexte du contrôle de performance ou de Mindsteps, exige d'autres données que le monitoring de l'éducation d'une autorité cantonale de surveillance. C'est pourquoi, il faut toujours se demander quelles données sont appropriées et exigibles pour atteindre le but et si leur traitement est approprié au sens étroit.

Le principe de minimisation des données est pertinent en particulier lorsqu'il s'agit d'une collecte de données relativement importante, comme l'établissement d'une banque de données centralisée. LehrerOffice met à disposition une telle solution qui permet l'accès simultané de plusieurs enseignants et qui stocke diverses données (plans des matières, questionnaires d'évaluation, plans de promotion, rapports et contrôles d'apprentissage, certificats, etc.). Les données individuelles sont chaque fois utilisées dans un but déterminé, comme planifier le semestre, promouvoir des élèves ou confirmer la fin d'un cycle d'études, d'où il importe de connaître les données personnelles enregistrées dans le cas d'espèce et de savoir si elles sont nécessaires pour atteindre les buts. Il serait ici judicieux d'établir des directives à l'intention des membres du corps enseignant et des directions d'école pour décrire les collectes de données permises¹⁰⁹. Ainsi, les informations relatives aux apprenants, comme le style de vêtements, les valeurs et les activités ou inactivités so-

103 Les considérants du RGPD indiquent clairement l'importance de l'utilisation des données à des fins de recherche, voir par ex. les nos 156, 157 et 159 RGPD.

104 Voir par ex. les art. 5, 9, 14, 17, 21 et 89 RGPD

105 Nommé aussi principe d'économie des données ou de prévention des données.

106 Paragr. 14, al. 2, IDG du canton de Bâle-Ville.

107 Paragr. 11, al. 2, IDG ZH ; voir aussi l'art. 11 LPrD du canton de Vaud.

108 Le paragr. 9 IDAG AG portant le titre « Proportionnalité » (« Verhältnismässigkeit ») prescrit qu'en cas de mise en œuvre de systèmes informatiques il faut respecter le principe de minimisation des données.

109 Voir par ex. les instructions d'établissement du certificat sur LehrerOffice du canton de Soleure, n. 22.

ciales qui n'ont aucune influence sur les progrès de l'apprentissage, ne peuvent pas figurer dans la banque de données. En outre, des aménagements techniques permettraient de restreindre préalablement la collecte des données par ex. par l'omission de champs non spécifiques visant la saisie d'« informations générales ».

4.1.4.4 Portabilité des données

La notion de droit à la portabilité des données est généralement comprise comme le droit d'une personne physique d'exiger la restitution des données personnelles la concernant de la part du propriétaire des données et de les faire transmettre à elle-même ou à un tiers. L'idée qui sous-tend cette notion est fortement influencée par le RGPD européen, dont l'art. 20 incarne la norme juridique la plus connue. Cette disposition exige également que les données soient éditées sous une forme lisible par ordinateur et que la restitution s'effectue gratuitement.

À proprement parler, le droit à la portabilité de données est inconnu en droit suisse, qui se trouve ainsi dans la situation de la plupart des ordres juridiques hors UE¹¹⁰. Le Conseil fédéral a évoqué la question dans son rapport de 2011 sur les médias sociaux mais sans thématiser de solution concrète¹¹¹. Alors qu'il s'exprimait encore avec réserve sur le sujet dans le rapport du 21 décembre 2016 sur l'avant-projet à la nouvelle LPD¹¹² et qu'il pointait les obstacles dans le contexte du projet du 15 septembre 2017¹¹³, il a maintenu en mai 2018 qu'il fallait examiner plus avant la question de savoir s'il convient de l'introduire¹¹⁴.

Un examen attentif de la situation juridique en Suisse montre que le droit d'accès issu du droit de la protection des données est déjà très proche du droit à la portabilité des données de l'UE décrit ci-dessus et pourrait endosser la même fonction moyennant de faibles adaptations¹¹⁵. La seule condition serait d'ajouter que le droit à la restitution se rapporte à des données électroniques dans un format structuré et que le titulaire du droit peut demander le transfert à des tiers¹¹⁶.

Les lois d'information et de protection des données cantonales connaissent aussi un droit à l'information¹¹⁷. Quant à l'objet de la restitution, seules quelques-unes de ces dispositions atteignent le degré de précision de la LPD. Selon cette dernière, toutes les données à disposition doivent être restituées à la personne concernée et le but ainsi que les bases légales doivent lui être communiqués le cas échéant (art. 8, al. 2, let. a, LPD)¹¹⁸. En ce qui concerne la forme de l'accès, quelques lois obligent les autorités à donner accès aux données sous forme compréhensible¹¹⁹. Mais cela ne garantit pas encore que le titulaire du droit d'accès puisse faire un usage ultérieur de ses données. En outre, les renseignements n'ont pas à être communiqués par écrit et encore moins sous forme électronique¹²⁰, ce qui rend extrêmement difficile la portabilité des données d'une installation à une autre.

4.1.4.5 « Propriété des données »

Sous l'aspect de leur valeur économique, la question est apparue par le passé de savoir à qui en définitive appartiennent les données. Cette discussion est menée sous la notion de « Propriété des données ». Or ladite notion est trompeuse car elle implique l'existence d'une telle institution juridique. Mais ce n'est pas

110 LUKAS HECKENDORN URSCHELER/ALBERTO ARONOVITZ/JOHN CURRAN/KAREN TOPAZ DRUCKMAN, Allgemeine Regelungen und die Berücksichtigung neuerer technischer Entwicklungen im Datenschutzrecht, Elektronische Publikationsreihe von Gutachten des Schweizerischen Instituts für Rechtsvergleichung, E-Avis ISDC 2017-07, 3 août 2016, 6 ss., 60.

111 Conseil fédéral, Cadre juridique pour les médias sociaux, Rapport du Conseil fédéral en réponse au postulat Amherd 11.3912 du 29 septembre 2011, 36-37 ; Yaniv BENHAMOU/LAURENT TRAN, Circulation des biens numériques : de la commercialisation à la portabilité des données, sic ! 2016, 571-591, 586.

112 Office fédéral de la justice, Rapport explicatif concernant l'avant-projet de loi fédérale sur la révision totale de la loi sur la protection des données et sur la modification d'autres lois fédérales, du 21 décembre 2016, 22.

113 Message E-LPD, FF 2017, 6565, 6604, 6607.

114 Voir OFCOM, communiqué du 9 mai 2018, Mesures pour une politique suisse des données tournée vers l'avenir, <https://www.bakom.admin.ch/bakom/fr/page-daccueil/l-ofcom/informations-de-l-ofcom/communiqués-de-presse.msg-id-70694.html>

115 WEBER/THOUVENIN, n. 67, 86 S.

116 Voir pour le détail WEBER/THOUVENIN, n. 67, 86 s.

117 Voir paragr. 20 IDG ZH ; paragr. 26 InfoDG SO ; paragr. 23 IDAG AG ; art. 25 LPrD VD ; art. 23 LPDP TI.

118 Pour une exception, voir paragr. 24, al. 1^{er}, IDAG AG, qui précise les mêmes obligations.

119 Paragr. 26, al. 1^{er}, InfoDG SO ; paragr. 24 IDAG AG ; art. 23, al. 2, LPDP TI.

120 Paragr. 26, al. 1^{er}, InfoDG SO ; paragr. 24 IDAG AG ; art. 26, al. 2, LPrD VD ; art. 23, al. 2, LPDP TI.

le cas. Comme les données ne sont pas de nature physique, elles ne sont pas considérées comme des choses au sens du code civil suisse (CC). Pour cette raison, il ne peut y avoir, ou, on ne peut fonder, ni propriété ni possession à leur sujet¹²¹. Cependant, elles peuvent faire sans autre l'objet de contrats par lesquels une partie aménage en général à l'autre le droit de les utiliser. Au moment de l'exécution, la partie contractante ou bien accorde un accès à sa propre base de données, ou bien en établit des copies et les transmet à l'autre partie.

Ainsi, les données sont principalement attribuées à une personne, à une entreprise ou à une autorité *en fait* mais non *en droit*¹²². La personne à qui elles sont attribuées en est désignée comme maître mais pas comme propriétaire. Il convient de souligner que malgré cette situation juridique le maître des données (de l'éducation) peut se protéger contre la transmission de celles-ci. Il peut par exemple invoquer à certaines conditions la loi sur la concurrence déloyale (LCD)¹²³ ou, pour autant qu'elles soient secrètes, se prévaloir de la protection du secret en droit civil ou pénal¹²⁴.

Mais ces considérations ne répondent pas encore à la question de savoir s'il convient de créer dans la loi une propriété sur les données. Le sujet a suscité la controverse entre la politique et le monde scientifique ces dernières années en Suisse et dans les pays européens. Tel n'est pas le cas dans la doctrine dominante et bien établie¹²⁵. Introduire un nouveau droit ne se justifie que difficilement ; il ne résoudrait que peu de problèmes et en créerait de nombreux autres, parfois imprévisibles. Abstraction faite de ce débat, la recherche n'a pas encore pu établir clairement malgré des efforts considérables comment élaborer ce droit d'un point de vue pratique, en particulier qui est censé en être le titulaire et comment définir l'objet de protection « données »¹²⁶. Le Conseil fédéral s'en est également distancé dans ses jalons d'une politique des données du printemps 2018¹²⁷.

Dans le contexte de l'éducation également, une propriété des données n'atteindrait pas son objectif. Comme vu plus haut, l'attribution de données à une entité juridique est pratiquement insoluble d'un point de vue pratique : selon certains arguments, l'apprenant qui saisit des données comme réponses à un test dans Mindsteps devrait aussi en être le propriétaire. En même temps, le producteur du logiciel dira qu'on ne peut numériser de données sans logiciel et que celui-ci détermine le bien à protéger. Et enfin, la personne qui formule les devoirs affirmera également qu'il n'y a pas de données sans devoirs. De toute évidence, il est illusoire de vouloir déterminer formellement un droit originaire dans un tel conflit d'intérêts. Ce dernier ne se résoudrait que par le biais de l'attribution à toutes les parties prenantes d'un droit ou d'une copropriété. Dans le premier cas, il en résulterait une pluralité de droit indésirable, dans le second, une situation prohibitive vu l'obligation de réunir l'unanimité de toutes les parties prenantes à chaque utilisation des données. Abstraction faite de ces difficultés pratiques, la position des éventuels propriétaires n'en serait pas vraiment améliorée car les acteurs économiquement plus forts se feraient transmettre les droits de propriété par contrat. Les premiers resteraient en définitive sans aucune protection.

Dans ce contexte, il convient de préférer la situation actuelle qui considère comme déterminant pour l'attribution le contrôle de fait sur les données de l'éducation. Par conséquent, l'école, l'établissement de formation ou l'autorité administrative est le maître des données d'administration. L'entreprise qui génère des données à l'aide du logiciel, par ex. un outil pédagogique numérique ou un logiciel d'apprentissage comme

121 Parmi beaucoup d'autres : ROLF H. WEBER/FLORENT THOUVENIN, Dateneigentum und Datenzugangsrechte – Bausteine der Informationsgesellschaft ?, ZSR 2018, 43–74, 49.

122 Dans des cas d'espèce et à d'autres conditions, les données peuvent être protégées par le droit de la propriété intellectuelle ou comme secrets d'affaire ; l'objet de la protection ne consiste dès lors pas dans les données mais dans l'invention, l'œuvre protégée par le droit d'auteur ou le savoir-faire secret.

123 Art. 5, let. c, LDA.

124 Art. 6 LDA ; art. 162 code pénal (CP).

125 Pour la Suisse, WEBER/THOUVENIN, n. 84, 72 ss. ; GIANNI FRÖHLICH-BLEULER, Eigentum an Daten ?, Jusletter du 6 mars 2017 ; DANIEL HÜRLIMANN/HERBERT ZECH, Rechte an Daten, sui generis 2016, 89 ss. ; a.M. MARTIN ECKERT, Digitale Daten als Wirtschaftsgut : digitale Daten als Sache, SJZ 2016, 245–249 ; MARTIN ECKERT, Digitale Daten als Wirtschaftsgut : Besitz und Eigentum an digitalen Daten, SJZ 2016, p. 65–274 ; MARC AMSTUTZ, Dateneigentum, Archiv für die civilistische Praxis, 218e vol., août 2018, 438–551. Pour d'autres avis et débats en Allemagne et en Europe, voir les remarques in WEBER/THOUVENIN, n. 118, n. 3 et 4.

126 FLORENT THOUVENIN/ROLF H. WEBER/ALFRED FRÜH, Data ownership : Taking stock and mapping the issues, in : Dehmer/Emmert-Streib (éd.), Frontiers in Data Science, Boca Raton 2018, 123 ss. et 129 ss. ; ANDREAS WIEBE, Protection of industrial data – a new property right for the digital economy ?, GRUR Int. 2016, 877–884, 883.

127 OFCOM, communiqué du 9 mai 2018, n. 11.

mindsteps est le maître des données d'application. Les autorités ou les organes qui mènent les enquêtes sont les maîtres des données d'enquête. De même, les acteurs qui contrôlent les métadonnées systématiques en sont les maîtres.

Au contrôle de fait des données s'ajoute alors le plan du droit de la protection des données. Le droit de la protection des données accorde aux personnes concernées, comme les élèves ou les titulaires de l'autorité parentale, des droits, en particulier le droit d'accès, qui interviennent dans ce contrôle ou qui le dominent¹²⁸. Une règle semblable est applicable lorsqu'il existe des droits de tiers, comme des droits immatériels¹²⁹ ou la protection du secret. Par conséquent, pour autant qu'il existe des données personnelles, le rattachement direct à une entité juridique est l'exception et non la règle. Il s'agit plutôt d'un réseau de rattachements de droit et de fait qu'il convient d'examiner dans chaque cas d'espèce.

4.1.4.6 Accès aux données

L'utilisation des données peut buter non seulement sur le contrôle juridique mais aussi sur le contrôle de fait, précité. C'est surtout le cas lorsque les données sont stockées dans des silos de données (propriétaires) et qu'elles sont accessibles uniquement à un cercle déterminé de personnes, à des groupes de personnes ou à des institutions. Les raisons peuvent être de nature technique, organisationnelle ou juridique. Non seulement ces restrictions pèsent sur l'utilisation des données mais encore elles entraînent dans certains cas à réitérer des enquêtes et des tâches.

Ces silos de données se trouvent par exemple dans le domaine des données d'application et d'enquête. Ainsi, ils apparaissent non seulement chez les prestataires qui excluent dans leurs conditions générales contractuelles l'utilisation des données par d'autres acteurs, mais aussi dans le domaine du pilotage de l'éducation, où elle est réservée aux chercheurs qui obtiennent, sur demande, l'accès aux données de vérification de l'atteinte des compétences fondamentales (COFO).

Par conséquent, la question se pose de savoir si des droits d'accès déterminés existent ou devraient exister dans le domaine de l'éducation. Afin d'y répondre, il convient de clarifier tout d'abord qui demande accès à quelles données, dans une grande diversité de constellations :

- a) S'il s'agit de personnes concernées qui voient traiter leurs données personnelles comme des apprenants ou les membres du corps enseignant, ils peuvent, comme expliqué plus haut, faire valoir un droit d'accès, découlant du droit de protection des données, inscrit non seulement à l'art. 8 LPD mais aussi dans les lois cantonales¹³⁰. Moyennant une petite adaptation, ce droit d'accès pourrait même devenir un droit à la portabilité¹³¹.
- b) Une base légale est nécessaire pour légitimer un flux de données allant dans le sens d'une demande d'accès, formulée par des services administratifs ou des autorités, aux données (de l'éducation) détenues par d'autres autorités¹³². Pour autant qu'il s'agisse de données sensibles ou de profils de personnalité¹³³, elle doit être contenue dans une loi au sens formel¹³⁴. Mais le contexte de ce rapport ne permet pas d'identifier ces bases légales et de déterminer si elles sont suffisantes.
- c) Les droits d'accès sont particulièrement intéressants lorsque des services administratifs ou des autorités aimeraient avoir accès à des données de l'éducation contrôlées de fait par des entreprises privées. La procédure adéquate consiste à ce que la collectivité qui confie des tâches du domaine de l'éducation à

128 Voir à ce sujet les explications sur l'autodétermination informationnelle (4.1.4.1, p. 53), la portabilité des données (4.1.4.4, p. 57) et l'accès aux données (4.1.4.6, p. 59).

129 Il s'agit là de droits visant à protéger certaines choses incorporelles (dites biens immatériels) comme les inventions, les œuvres d'art, les conceptions ou les dessins. Conformément aux objets protégés nommés ci-dessus, les droits consistent en des brevets, des droits d'auteur, des designs ou des marques.

130 Voir ci-dessus Portabilité des données (4.1.4.4, p. 57).

131 Voir également ci-dessus Portabilité des données (4.1.4.4, p. 57).

132 La règle n'est pas valable seulement sur le plan fédéral (art. 19 LPD) mais aussi sur le plan cantonal (paragr. 16, al. 2, IDG ZH ; paragr. 8 en lien avec le paragr. 14, IDAG AG ; paragr. 21, al. 3, InfoDG SO ; art. 6 en lien avec l'art. 10, LPDP TI).

133 Voir en particulier ci-dessus Protection des données (4.1.3.2, p. 51).

134 Art. 17, al. 2, DSG ; paragr. 8, al. 2, IDG ZH ; paragr. 8, al. 2, let. a, IDAG AG ; paragr. 15, al. 2, InfoDG SO ; art. 5, al. 2, let. a LPrD VD ; art. 6, al. 1^{er}, LPDP TI.

des privés reprenne les obligations d'information et de publication dans les ordonnances et les contrats de droit administratifs. Les droits d'accès de la collectivité doivent être négociés. Ainsi, la préparation des données à l'intention des requérants peut entraîner des coûts considérables ou affecter les droits de tiers¹³⁵. De toute évidence, il peut être extrêmement difficile d'obtenir ces droits lorsque les privés sont en position de force. En définitive, les autorités ne peuvent rétablir l'équilibre que par des coopérations dans le paysage suisse de l'éducation qui renforcent leur position de négociation. Des droits d'accès légaux dans l'intérêt public ne se justifieraient dans tous les cas que si la demande en matière de données de l'éducation était défavorisée de manière générale sur le plan structurel¹³⁶.

- d) La constellation inverse, dans laquelle des privés entendent accéder à des données des pouvoirs publics, est intéressante dans le domaine de l'éducation car, à la différence d'autres secteurs, les institutions publiques de l'éducation contrôlent davantage de données que les privés. Par exemple, sur le plan fédéral et cantonal, des lois sur la transparence réglementent l'accès des privés aux documents officiels qui échoient à une direction de l'éducation pour accomplir ses tâches publiques¹³⁷. Fondés sur ces lois, les prestataires privés qui fournissent des produits semblables à Mindsteps peuvent consulter les documents officiels sur place ou demander des copies¹³⁸. Par ce biais, les privés pourraient découvrir les échelles à la base des applications existantes et offrir des produits concurrents. Le droit d'accès des privés aux documents officiels est toutefois limité : l'accès est en particulier restreint ou refusé lorsque l'octroi porterait atteinte à la sphère privée de tiers¹³⁹. Ces instruments ne permettent donc pas de demander les données personnelles relevant de l'éducation.
- e) À la différence des autorités¹⁴⁰, les privés peuvent théoriquement invoquer d'autres bases légales, en particulier le droit des cartels, lorsqu'ils demandent l'accès aux données d'autres privés afin de déployer une activité économique sur un marché déterminé. Au moins en théorie, ils ont un droit d'accès découlant du droit des cartels pour autant que le maître des données dispose d'une position dominante sur le marché qui lui permette d'entraver l'introduction par le requérant d'un produit nouveau, potentiellement demandé. Mais justement dans ces constellations, l'application du droit des cartels est en pratique liée à des difficultés considérables¹⁴¹. D'une part, les procédures cartellaires sont longues et lourdes, d'autre part la nécessité de délimiter les marchés (de données) pertinents place aujourd'hui le droit des cartels devant des défis quasi insurmontables.

4.1.5 Infrastructures visant à renforcer les concepts

Les infrastructures à disposition jouent un rôle non négligeable dans la politique des données. Par exemple, lorsque les données de l'éducation sont utilisées uniquement au sein de silos de données séparés, non reliés et incompatibles, mêmes les instruments juridiques n'y peuvent grand-chose. Quant aux infrastructures techniques, il convient de distinguer deux approches : d'une part, les systèmes privés décentralisés de données et d'administration nécessitant le consentement (les « PIMS » ou les modèles MyData), et d'autre part les systèmes d'information centralisés étatiques de l'éducation.

135 ALFRED FRÜH, Datenzugangsrechte, Rechtsrahmen für einen neuen Interessenausgleich in der Datenwirtschaft, sic ! 2018, 521–539, 527 s.

136 Pour la détermination, voir FRÜH, n. 132, 528 s. La situation pourrait paraître différente si l'intérêt public exigeait d'accéder aux données contrôlées par des entreprises avec lesquelles il n'existe aucune relation contractuelle, FRÜH, *ibid.*, 529.

137 Art. 1^{er}, LTrans ; paragr. 4 ss., IDAG AG ; paragr. 12 ss., InfoDG SO ; art. 1^{er} ss., Loi sur l'information du canton de Vaud (ci-après : LInfo VD) ; art. 1^{er} ss., « Legge sull'informazione e sulla trasparenza dello Stato » du canton du Tessin (ci-après : LIT TI) ; paragr. 14, al. 2, IDG ZH.

138 Art. 6, al. 1^{er}, LTrans ; paragr. 5, al. 2, IDAG AG et paragr. 12, al. 3, InfoDG SO, permettent en outre la réception électronique des documents ; art. 9, al. 2, LIT TI ; art. 13, al. 1^{er}, LInfo VD.

139 Art. 7, al. 2, LTrans ; paragr. 4, al. 3, IDAG AG, paragr. 13, al. 1^{er}, InfoDG SO et art. 11, al. 1^{er}, LIT TI (des intérêts privés prépondérants s'opposent à l'obligation d'information) ; art. 16, al. 3, LInfo VD ; paragr. 23, al. 3, IDG ZH.

140 Les représentants de l'administration publique n'ont pas d'activité commerciale ; il leur manque la capacité concurrentielle.

141 FRÜH, n. 132, 533 ss.

4.1.5.1 Personal Information Management Systems (PIMS) et modèles MyData

Les Personal Information Management Systems (PIMS) sont des systèmes de gestion de données censés permettre à l'individu de collecter et de gérer de manière centralisée les données qui le concernent. Axés sur l'utilisateur, ces systèmes se distinguent des purs « Management Information Systems » qui permettent la gestion de données par une direction d'école ou une autorité administrative (par ex. la solution de banque de données LehrerOffice). L'utilisateur a un aperçu de son stock de données et décide en toute indépendance de l'utilisation future de ses données pour des buts propres ou autres. Un exemple en est l'utilisation scientifique des données à un moment ultérieur non déterminé¹⁴². Les systèmes personnels de gestion avec consentement connus sous le nom de MyData suivent une approche très semblable¹⁴³. Ceux-ci permettent à l'utilisateur de contrôler les consentements donnés à divers fournisseurs de services. L'approche MyData s'efforce de créer des consentements dynamiques, standardisés et lisibles par ordinateur, afin qu'ils soient faciles à comprendre par les intéressés.

Contrairement à un système d'information des données de l'éducation centralisé, (voir sous 4.1.5.2, p. 62), les structures PIMS et MyData ont pour avantage d'entraver l'accès à toutes les données en une seule attaque de pirate informatique, à la condition qu'il existe plusieurs de ces structures. Mais tout comme la centralisation, la décentralisation nécessite aussi une certaine standardisation ou uniformisation. Ainsi, les systèmes gérés par plusieurs fournisseurs ont besoin de formats de données interopérables, ce qui ne se peut que par le biais de la standardisation.

PIMS et MyData influencent certains des concepts juridiques : tant PIMS que les modèles MyData tentent de mettre les personnes en mesure de déterminer elles-mêmes le traitement et l'utilisation des données qui les concernent. Ce sont donc des instruments techniques visant à renforcer le concept de l'autodétermination informationnelle. En même temps, la connexité avec le droit à la portabilité est étroite¹⁴⁴. PIMS/MyData et le droit à la portabilité se favorisent mutuellement : le droit à la portabilité des données facilite l'utilisation des offres PIMS, car seul le droit à la portabilité permet effectivement à l'individu de rassembler des données de sources diverses dans un PIMS¹⁴⁵ ; et l'existence de PIMS/MyData peut avoir pour conséquence que les personnes concernées exercent effectivement leur droit à la portabilité.

Dans le domaine de l'éducation, des solutions PIMS pourraient permettre aux apprenants de rassembler et de gérer eux-mêmes les données de l'éducation créées durant leur cursus scolaire, voire durant leur parcours de formation et de perfectionnement. À cet effet, les apprentis pourraient collecter les données de l'éducation auprès des directions d'école et des divers degrés scolaires et les déposer dans un système de gestion centralisé. Mais on peut douter de la plus-value effective d'un tel PIMS car ces données n'échoient en règle générale qu'à un nombre restreint de collaborateurs (notamment auprès de chaque commune scolaire) et non pas à des sources innombrables. C'est pourquoi la collecte des données de l'éducation d'un ou d'une élève dans un PIMS n'est au fond intéressante qu'en cas de changement de domicile. Et même alors, il n'y aurait pas besoin de PIMS car l'échange d'informations entre autorités suffit. Du point de vue de l'individu, les solutions PIMS ne sont pas particulièrement attractives car les possibilités de gestion des données par les apprenants ou les titulaires de l'autorité parentale qui sont par exemple pertinentes pour le passage d'une classe à une autre sont nécessairement restreintes. Les solutions PIMS faciliteraient l'accès uniquement lorsque le consentement de la personne concernée est nécessaire, au moment d'une évaluation de ses données de performance et de ses progrès d'apprentissage par des acteurs privés. Pour être utiles, de telles solutions devraient toutefois présenter une plus-value d'une part pour les apprenants, c'est-à-dire des évaluations pertinentes pour eux, et/ou d'autre part pour le public, à savoir un gain de connaissances pour le pilotage de l'éducation, ce qui reste à prouver. Enfin, une masse critique est nécessaire pour imposer les solutions PIMS. La série de PIMS déjà opérationnels ne l'a pas encore atteinte¹⁴⁶.

142 WEBER/THOUVENIN, n. 67, 7.

143 En réalité, une délimitation entre PIMS et MyData est extrêmement difficile vu la complexité des configurations pratiques.

144 WEBER/THOUVENIN, n. 67, 38 s.

145 WEBER/THOUVENIN, N. 67, 7.

146 Bien qu'il existe déjà divers fournisseurs privés et publics, les services PIMS ne se sont pas imposés jusqu'à aujourd'hui. Pour un aperçu des projets PIMS, voir WEBER/THOUVENIN, n. 67, 10 ss.

4.1.5.2 Systèmes d'information de l'éducation

Les systèmes d'information de l'éducation sont des infrastructures qui centralisent et permettent d'utiliser les données de l'éducation. L'Estonie en est le précurseur. L'Estonian Education Information System (EHIS)¹⁴⁷ réunit les données de l'éducation de tous les habitants de l'Estonie. La banque de données contient des informations sur les établissements de formation, les élèves, les enseignants et les maîtres de conférence, les certificats de fin d'études, les matières et les plans d'enseignement. Une base légale a été créée à cet effet¹⁴⁸. Au paragraphe 2, elle décrit le but de l'EHIS, à savoir rassembler et organiser toutes les données nationales de l'éducation afin de pouvoir effectuer des évaluations statistiques. L'EHIS est conçu en divers sous-registres qui contiennent des informations sur les membres du corps enseignant, les apprenants, les établissements de formation ou les plans d'enseignement (paragr. 5). Tandis que quelques registres sont publics comme ceux sur les plans d'enseignement et les établissements de formation, d'autres ne peuvent être consultés que par les personnes concernées, par exemple les enseignants ou les apprenants (paragr. 7).

L'EHIS se fonde sur une base légale qui détermine les principes de traitement (paragr. 4 ss.) et définit préalablement les données personnelles (paragr. 16 et 18 sur les données du corps enseignant) ainsi que les buts présidant à leur utilisation (paragr. 15 sur les mêmes données) et la durée de conservation des données (paragr. 8 en lien avec les catégories de données). Sous la surveillance d'un service responsable, il est soumis à l'autorité de protection des données d'Estonie chargée de veiller au respect du droit de la protection des données et à la légalité de son entretien (paragr. 43). La centralisation de ce système lui permet d'être largement standardisé, ce qui facilite grandement l'utilisation et l'évaluation des données. De plus, il est ainsi particulièrement bien défendu contre le piratage des données.

Un système d'information de l'éducation comme celui de l'Estonie entrave la création de silos de données privés. Le système est particulièrement intéressant du point de vue du concept de la portabilité des données, comme le droit d'accès étendu du droit de protection des données, ou du droit d'accès : les demandes d'accès aux données pourraient toujours se faire auprès du service de gestion, dans le premier cas par les personnes concernées sur la base de leur droit à la portabilité ou d'accès, dans le second cas par des privés ou des chercheurs sur la base des droits accordés par les lois sur la transparence.

Comme le montre l'exemple de l'EHIS, le besoin de coordination et de gestion d'un système centralisé est élevé. En sus des défis technologiques, il faudrait établir en Suisse des unités administratives et des bases légales ad hoc.

4.1.6 Recommandations

Pour poursuivre le développement d'une politique d'utilisation des données dans le domaine de l'éducation en Suisse, et sur la base des concepts juridiques examinés, nous émettons quatre recommandations de principe et quelques recommandations concrètes.

4.1.6.1 Recommandations de principe

a) État des lieux

Dans un premier temps, il paraît indispensable d'établir un état des lieux systématique de l'utilisation des données dans le système suisse de l'éducation. Il n'a pas encore été fait car la structure fédérative du système de l'éducation complique considérablement la tâche. À défaut, dans le cadre imparti, la présente contribution s'est uniquement concentrée sur des exemples comme le contrôle de performance, Mindsteps ou LehrerOffice, et n'a pu livrer qu'un aperçu de la situation juridique qui prévaut aujourd'hui.

D'une part, cet état des lieux exige que toutes les données fassent l'objet d'une saisie systématique, à savoir les données personnelles des élèves, des membres du corps enseignant, des directions d'école ainsi que les données d'administration, d'application ou d'enquête, qui sont saisies, contrôlées, traitées et utilisées sur un plan administratif fédéral, cantonal et communal par divers acteurs comme les membres du corps enseignant, les directions d'école, les services administratifs, les offices statistiques et les entreprises

147 Pour des informations sur le système de l'éducation en Estonie, voir <<https://e-estonia.com/solutions/education/>>.

148 La base légale de l'EHIS est téléchargeable sous <<https://www.riigiteataja.ee/akt/129122011185?leiaKehtiv>>. Les explications suivantes se basent sur une traduction (NdT : en allemand) de Google Translate.

privées. À cette condition, l'analyse de cas concrets pourrait démontrer par quels acteurs les données sont contrôlées, dans quels buts d'utilisation, si elles sont échangées et dans quel sens. L'état des lieux fournirait un aperçu des silos de données existants qu'il faudrait éventuellement déverrouiller.

D'autre part, cet état des lieux implique également d'établir un aperçu de toutes les lois cantonales applicables dans le domaine de l'éducation, comme les lois et ordonnances scolaires cantonales, et en matière de protection des données, c'est-à-dire les lois cantonales sur l'information et la protection des données.

b) Buts d'une politique d'utilisation des données cohérente

Après un examen plus attentif, la politique d'utilisation des données cohérente visée dans l'éducation n'est pas un but en soi. Bien sûr, il convient de rechercher dans ce domaine une symétrie raisonnable avec le droit de protection des données et la sécurité des données tout en veillant à la cohérence. Mais les buts qu'une telle symétrie doit viser n'ont encore jamais été précisés, même dans la stratégie numérique de la CDIP. Les « buts » évoqués dans ce document représentent uniquement des visions possibles pour le paysage de l'éducation dont la portée et l'articulation restent vagues¹⁴⁹. Seule la formulation de buts précis permettra de projeter les développements nécessaires à l'aide de faits réels tels que la mise en œuvre à l'école d'outils pédagogiques numériques de fournisseurs privés, l'utilisation d'ordinateurs privés dans ou pour l'enseignement, la gestion des évaluations de performance des élèves, l'échange de données avec d'autres prestataires comme les services du médecin ou du psychologue scolaire, etc.

c) Standardisation

L'utilisation des données de l'éducation dépend de l'interopérabilité des systèmes. Sans un certain degré de standardisation, toute utilisation est impossible. Sinon, les coûts de transaction sont trop élevés. A priori, même les instruments juridiques n'y peuvent pas grand-chose. Un droit à la portabilité ne peut pas remplir sa fonction si les données personnelles requises sont livrées dans un format propriétaire ou sous forme imprimée. Par conséquent, les prescriptions du droit d'accès ou du droit à la portabilité doivent être accompagnés de mécanismes tels que des normes, des standards industriels ou des bonnes pratiques qui garantissent que la norme juridique atteigne son but. Le succès du droit à la portabilité prévu par l'UE à l'art. 20 RGPD dépend par exemple en première ligne de l'interprétation pratique qui sera faite de la formulation légale « format structuré, couramment utilisé et lisible par machine ».

Les infrastructures peuvent favoriser la standardisation technologique¹⁵⁰, surtout via la centralisation des systèmes. Dans tous les autres cas, même lors de l'utilisation des PIMS, il est nécessaire de faire un effort dans ce sens. Même s'il ne s'agit pas d'une tâche étatique¹⁵¹, l'État devrait s'y atteler et s'entendre avec les parties prenantes pour utiliser des standards existants.

Il conviendrait de discuter de la standardisation des contenus. En effet, la question se pose de savoir s'il faut standardiser les données saisies dans les cantons, les lieux de leur(s) enregistrement(s), les organes qui les contrôlent et les gèrent, etc. L'exemple estonien a montré jusqu'où peuvent aller les efforts d'harmonisation. À l'instar de la loi sur l'harmonisation de registres (LHR) qui prescrit à son art. 6 le contenu minimal des registres des habitants, il faudrait déterminer les types de données nécessaires sur les apprenants, le corps enseignant, ou tout autre personne chargée de tâches de la formation. En d'autres termes, les institutions cantonales de l'éducation devraient systématiquement identifier leur besoin de données et s'entendre pour déterminer celles qui doivent figurer dans leurs systèmes.

Évidemment, les tâches technologiques et matérielles de standardisation exigent un effort de coordination considérable. On peut se demander si cet effort est justifié. Ces processus pourraient tout d'abord se limiter aux cantons, qui pourraient dans cette optique recourir judicieusement aux concordats intercantonaux.

d) Protection de l'individu

Les recommandations qui suivent et qui précèdent servent à faciliter l'utilisation des données dans le domaine de l'éducation. Mais ces simplifications sont potentiellement à la charge de l'individu. C'est

149 Voir la stratégie numérique de la CDIP du 21 juin 2018 pour la gestion de la transition numérique dans le domaine de l'éducation, ch. 1, qui expose notamment que les possibilités d'utilisation des données du système éducatif sont pleinement exploitées (ch. 1.2) et qu'il faudrait prévoir une identité numérique reconnue pour chaque élève et pour chaque personne au sein des établissements de formation (ch. 1.3).

150 Voir ci-dessus Système d'information de l'éducation (chap. 4.1.5.2, p. 62.)

151 La standardisation ou la normalisation technique est effectuée par des organisations de standardisation nationales et internationales privées.

pourquelles apprenants, notamment, doivent faire l'objet d'une attention soutenue. Des mesures sont nécessaires dès l'apparition de menaces concrètes, telles que les atteintes à la sécurité des données appelées *data breaches*. Ce principe est applicable même sans définition détaillée des buts de la politique d'utilisation des données.

4.1.6.2 Propositions concrètes

a) Transparence

La question de l'existence et de l'étendue de l'autonomie dont jouit l'individu par rapport aux données de l'éducation qui le concernent ou de l'étendue de son autodétermination informationnelle dépend dans une large mesure de la transparence en matière d'utilisation des données. Ici, le RGPD semble aller plus loin que la LPD et les dispositions cantonales. Le but du principe de transparence en vigueur au sein de l'UE vise à rendre les informations sur le traitement des données plus accessibles suivant le groupe cible, tels les élèves du degré primaire. Les déclarations de protection des données sont essentiellement rédigées avec simplicité et concision et les processus de traitement sont expliqués au moyen d'illustrations. De telles approches peuvent se combiner avec les obligations d'information. Le législateur jouit ici d'une certaine marge de manœuvre pour concrétiser et standardiser les obligations d'information et de communication des entités privées et publiques qui traitent des données sur le plan fédéral et cantonal¹⁵². Mais on peut se demander si l'accroissement de la densité normative et la création de nouvelles obligations d'information sont les bonnes solutions ; par exemple, les personnes concernées ne prennent généralement pas connaissance des informations contenues dans les déclarations de protection des données. C'est pourquoi un surcroît d'information n'amène pas nécessairement plus de transparence. C'est même plutôt l'inverse. Car la vraie transparence sur l'utilisation des données de l'éducation peut aussi apparaître d'une tout autre façon, à travers une culture d'ouverture et de simplicité dans les écoles, au sein des autorités et à l'égard des « laïcs », avec le soutien éventuel de médias visuels ou audiovisuels.

b) Bases légales

Tant qu'elle contrôle l'échange et l'utilisation des données de l'éducation, la collectivité comme les écoles et les autorités de surveillance a en principe le pouvoir de créer les bases nécessaires. Mais les progrès en sont singulièrement à leurs balbutiements ; en tout cas, il n'y a selon toute évidence aucune disposition dans les législations cantonales qui prévoient, au-delà de phénomènes concrets, l'utilisation et l'échange des données de l'éducation. Les potentialités sont élevées. Sous réserve du principe de finalité, de nouvelles bases légales comprenant la description d'un objectif plus général ciblé¹⁵³ sur les buts de la politique des données pourraient simplifier l'utilisation des données. Lors de la promulgation de nouvelles lois ou de la révision de dispositions existantes, il convient de conseiller aux pouvoirs législatifs et exécutifs cantonaux de rédiger les buts aussi largement que possible eu égard audit principe, afin de permettre d'autres utilisations.

c) Exception d'utilisation au bénéfice de la recherche

La proposition d'accorder une importance accrue à l'exception au bénéfice de la recherche est étroitement liée à l'incitation formulée ci-dessus¹⁵⁴. Dans le domaine des données de l'éducation, on peut se demander par exemple s'il faudrait renoncer à appliquer certaines normes de protection des données lors du traitement de données personnelles à des fins de recherche. L'exception d'utilisation à des fins de recherche devrait faire l'objet d'une description claire et standardisée dans les législations cantonales et fédérale. La notion de recherche devrait y être interprétée largement et inclure les utilisations des données de l'éducation dans un but d'évaluation et de planification.

d) Minimisation des données en tant que tâche administrative

Les données collectées par les membres du corps enseignant et les directions d'école ainsi que la durée de leur enregistrement doivent être connues. Sinon, comme le montre l'exemple des solutions individuelles telle LehrerOffice, cette collecte risque de dépasser le but du traitement. En outre, des restrictions peuvent lui être préalablement apportées via la configuration technique de la banque de données¹⁵⁵. La minimisation

152 C'est déjà le cas au sein de l'UE. Voir art. 14 et 15 RGPD.

153 Voir ci-dessus Buts d'une politique d'utilisation des données cohérente (4.1.6.1, p. 62).

154 Voir ci-dessus les explications sur le principe de finalité (4.1.4.2, p. 54). En outre, l'exception d'utilisation à des fins de recherche autre que celle prévalant au sein de l'UE pourrait empêcher la réalisation de projets de recherche prévus avec des établissements situés dans l'UE.

155 Voir par ex. les instructions et le modèle de certificat du canton de Soleure, n. 56.

bien comprise des données est une tâche incombant aux autorités (de surveillance) cantonales et communales. Celles-ci sont prédestinées à donner aux écoles des instructions ou des directives claires sur l'usage d'instruments comme LehrerOffice.

e) Développement du droit d'accès selon le droit de la protection des données

Inscrit dans la LPD et les législations cantonales de protection des données, le droit d'accès est un instrument important. Mais il n'a jamais été adapté au monde numérique¹⁵⁶ et, partant, n'a pas de valeur dans le contexte de l'utilisation des données. Modernisé, il simplifierait sensiblement l'utilisation des données et conduirait à la portabilité des données. Son extension vers un droit à la portabilité pourrait encourager l'utilisation des PIMS ou de modèles MyData, leur effet dans ce domaine étant toutefois sujet à des réserves¹⁵⁷. Des propositions de développement ont déjà été faites : les données doivent être transposées gratuitement dans un format courant et transmises au choix à la personne concernée ou à un tiers désigné, par exemple une entreprise privée, une autre autorité ou un administrateur de données.

f) Données de l'éducation et privés

Les autorités qui concluent des contrats avec des prestataires privés devraient faire preuve de prudence : elles devraient non seulement se ménager des droits d'utilisation sur les données mais aussi prendre conscience du fait que les données de l'éducation échappent à la sphère d'influence de l'État et à l'application de la législation cantonale. Elles doivent par conséquent respecter certains principes. Il conviendrait de formuler ceux-ci à leur intention sous forme générale et abstraite.

Même si elles connaissent les problèmes liés à ces constellations, elles risquent de ne pas pouvoir s'imposer face à des fournisseurs économiquement plus forts. En conséquence, il s'agit de renforcer le pouvoir de négociation des établissements de formation plus faibles via des contrats-cadres portant sur l'utilisation de prestations spécifiques de l'éducation. Le succès du contrat-cadre négocié par educa.ch¹⁵⁸ avec Microsoft sur l'utilisation de Microsoft Office 365¹⁵⁹ en est un exemple. Ces contrats pourraient se négocier pour d'autres prestations et garantir que les fournisseurs privés remplissent les obligations d'information et de divulgation.

Si on ne peut pas rétablir l'équilibre des forces dans le paysage suisse de l'éducation par des contrats-cadres et d'autres coopérations, il faudrait en définitive réfléchir à la possibilité d'inscrire dans la loi un droit général d'accès aux données de l'éducation détenues par des privés.

156 Voir ci-dessus Portabilité des données (4.1.4.4, p. 57) ; dans de nombreux cantons les informations ne doivent pas être fournies par écrit.

157 Voir ci-dessus sur les PIMS (4.1.5.1, p. 61).

158 Voir les contrats-cadres d'educa.ch, <<https://www.educa.ch/fr/contrats-cadre/microsoft>>. Le contrat-cadre prévoit par exemple que le for est à Berne et que les seuls serveurs cloud que Microsoft peut choisir sont ceux situés dans l'UE. Les établissements de formation de Suisse peuvent adhérer à ce contrat-cadre et configurer les services d'Office 365 pour leurs propres écoles.

159 Microsoft Office 365 (« Office 365 ») est une version basée sur le web de Microsoft Office et permet l'accès à des produits Microsoft comme Word, Excel, Powerpoint, OneNote, SharePoint, Teams et OneDrive. Les membres du corps enseignant ou les élèves peuvent s'annoncer à Microsoft avec un compte d'utilisateur et utiliser les services directement via le cloud.

4.2 Sécurité de l'information

Division Security & Network, SWITCH

- Martin Leuthold

Institut pour les Systèmes Mobiles et Distribués, FHNW

- Prof. Dr. Hannes Lubich

Table des matières

4.2.1	Contexte	67
4.2.2	Situation de la menace	68
4.2.2.1	Facteurs de risque	68
4.2.2.2	Besoin de protection qui en résulte	69
4.2.3	Solutions	71
4.2.3.1	Aspects conceptuels	73
4.2.3.2	Aspects techniques	75
4.2.3.3	Aspects juridiques et réglementaires	80
4.2.3.4	Aspects économiques	81
4.2.4	Remarques finales	82

4.2.1 Contexte

La Suisse est devenue un centre de prestations de services (banques, assurances, formation initiale et continue, etc.) et de produits (construction de machines, pharmacie, alimentation, etc.) établi dans le monde entier. Dans ces domaines, outre la qualité traditionnellement élevée de prestations et de produits en Suisse, l'avance en termes de savoir et donc la capacité innovative jouent un rôle important dans un environnement global et compétitif. À l'instar des grandes entreprises établies sur le marché, les petites et moyennes entreprises spécialisées jouent un rôle essentiel dans la recherche, l'expérimentation, la fabrication et la diffusion de produits et prestations de services innovateurs. Parallèlement à d'autres avantages inhérents à la situation de la Suisse, tels que la stabilité économique, financière, politique, juridique et sociale, le bon niveau de la formation et un climat général propice à l'économie et à l'investissement, ce « mélange » constitué de petites et de grandes entreprises, de production et de prestations de services, ainsi que de tradition et d'innovation contribue à ce que la Suisse soit considérée comme une place économique attrayante, sûre et résistante aux crises.

Toutefois, cette efficacité économique élevée a pour effet que le système économique, politique et social, ainsi que la propension générale au risque sont plutôt conservateurs – dans un monde de plus en plus globalisé et disruptif en raison de changements économiques, sociaux et politiques rapides, la Suisse est donc confrontée à un défi de taille : réagir rapidement et convenablement à ces développements ou pouvoir même y contribuer activement. Étant donné que la connaissance représente l'unique matière première suffisante et disponible à long terme en Suisse, la formation initiale et continue de haute qualité et bien établie à tous les degrés de formation en vue d'une qualification durable sur le marché du travail, la recherche fondamentale et appliquée jouent un rôle décisif pour la force innovatrice et la capacité à développer des produits et des services attrayants. Par conséquent, les institutions d'enseignement et de recherche suisses, les entreprises de toutes envergures et orientations, ainsi que la politique sur les plans cantonaux et fédéraux se penchent de manière approfondie sur la numérisation et sur de nombreuses technologies qui se situent en aval (apprentissage numérique, « *Internet of Things / Industrie 4.0* », « *Big Data* », « *Block-chain* », intelligence artificielle, etc.), qui ont le potentiel d'enrichir de façon décisive les formes d'apprentissage et de formation existantes ainsi que les produits et prestations de services, ou de les remplacer par de nouvelles offres. Toutefois, cet élan vers la numérisation a lieu malgré des initiatives centrales de pilotage (cf. OFCOM [2018] : Stratégie « *Suisse numérique* » ; SEFRI [2017] : *Défis de la numérisation pour la formation et la recherche en Suisse*, CDIP [2018] : *Stratégie de numérisation. Stratégie de la CDIP du 21 juin 2018 pour la gestion de la transition numérique dans le domaine de l'éducation*) souvent sous une forme fortement fragmentée et influencée par le fédéralisme. Actuellement, à un niveau supérieur, on discute par ailleurs de manière approfondie de la « cyber-souveraineté » de la Suisse, de concepts de « cyberdéfense » et de leur intégration dans des domaines supérieurs de la protection de la population, de la protection des infrastructures critiques et de l'approvisionnement économique général du pays. La *Stratégie nationale de protection de la Suisse contre les cyberrisques 2018-2022* constitue un résultat significatif de ces efforts.

Tous ces nouveaux développements et technologies sous-jacentes présentent un élément en commun : la protection de la performance et de la capacité d'innovation de la place industrielle de la Suisse ; par ailleurs, la protection des données et de la propriété intellectuelle qui y sont présentes constituent un thème d'importance primordiale compte tenu des menaces physiques et virtuelles en constante augmentation. Par ailleurs, dans le domaine de la formation scolaire assistée numériquement, un important accent est mis sur la protection des données à caractère personnel (notamment de mineurs) et la mise à disposition de savoir-faire et de suffisamment de ressources pour le fonctionnement sécurisé des infrastructures d'information et de communication nécessaires à cet effet, y compris la prise en compte d'appareils privés et l'utilisation de services de communication externes (WhatsApp) ou des stockages centraux de fichiers (Dropbox) entre les écoles, les apprenants et, le cas échéant, leurs tuteurs légaux.

Les domaines de la formation et de la recherche sont particulièrement concernés par le problème de la protection suffisante des données personnelles ou des intérêts personnels, économiques et nationaux qui y sont liés. D'une part, par le biais de la formation et de la recherche, la Suisse élabore et finance les bases essentielles de l'innovation en tant que pilier de la place industrielle suisse actuelle et à venir, et doit donc veiller à une protection adaptée à la menace respective. D'autre part, la formation et la recherche sont précisément fondamentalement tributaires d'un libre accès aux données pertinentes d'un point de vue de la

recherche et de la qualité, d'une collaboration internationale, d'un échange professionnel ainsi que d'une liberté généralement grande en matière d'enseignement et de recherche.

Cette formation et cette recherche délibérément ouvertes sont donc largement défendues par les institutions impliquées, entrant cependant partiellement en contradiction avec les exigences de l'autodétermination informationnelle et de la protection des données, ainsi que de l'utilisation commerciale avec des exigences de sécurité élevées des résultats élaborés.

Les parties concernées sont pleinement conscientes qu'il convient d'empêcher autant que possible les attaques portées contre les infrastructures correspondantes, la falsification ou la suppression de données ainsi que l'utilisation non intentionnelle de données et de la propriété intellectuelle. Toutefois, compte tenu de la complexité, de l'organisation fédérale et des intérêts fortement divergentes des organisations impliquées, il n'est ni possible, ni judicieux d'élaborer et d'imposer une solution uniforme pour la protection de toutes les institutions d'enseignement et de recherche en Suisse. Bien que des institutions centrales telles que SWITCH¹⁶⁰ et des projets tels que FIDES participant à la mise sur pied de la Fédération de services d'identités pour l'espace suisse de formation¹⁶¹ peuvent, sur la base de leur activité antérieure, assumer une fonction de coordination et tirer profit de nouveaux développements technologiques tels que le « *Big Data* » (mégadonnées) dans le domaine de la détection des menaces, les institutions scolaires de formation, d'enseignement et de recherche académiques ainsi que leurs mandants et utilisateurs respectifs doivent jouer un rôle essentiel et structurant pour définir et appliquer des mesures de protection appropriées.

4.2.2 Situation de la menace

4.2.2.1 Facteurs de risque

Les menaces qui pèsent sur les institutions de formation scolaire et d'enseignement et de recherche académiques découlent, d'une part, de changements sociaux, organisationnels, économiques et techniques échappant à l'influence directe des institutions concernées, et, d'autre part, de lacunes dans l'aménagement, le fonctionnement et l'adaptation continue des dispositifs de protection concernés. Les points énumérés ci-dessous sont notamment considérés comme facteurs de risque pertinents du secteur de la formation et de la recherche :

- le transfert toujours plus important de valeurs immatérielles significatives ou de l'accès à des valeurs matérielles (installations, bâtiments, systèmes de pilotage, etc.) dans l'espace virtuel, y compris des accès possibles à distance couplés à des privilèges élevés et d'autres interfaces souvent complexes et difficilement contrôlables,
- la dépendance croissante de l'économie, la science, la formation, l'administration et des particuliers (et donc au final d'une société numérique et mise en réseau au-delà des frontières nationales) face à la qualité suffisante, à la disponibilité, à la performance et à la sécurité de systèmes virtuels toujours plus cyber-physiques (par ex. infrastructures critiques d'approvisionnement et d'urgence basées sur l'informatique, etc.) et à la quantité en forte croissance de données recueillies, enregistrées et traitées,
- un environnement compétitif sur les plans politique et économique, dans lequel tant les États, les associations d'États que les entreprises sont toujours plus prêts à réaliser ou à faire réaliser un travail éducatif et économique agressif afin de préserver leur avantage concurrentiel respectif,
- la diffusion du crime organisé dans le cyberspace, puisque
 - les données qui y sont présentes ont une valeur de vente et d'utilisation attrayante pour les criminels ;
 - dans le cyberspace, en comparaison avec la criminalité organisée dans d'autres domaines tels que le commerce des drogues, la traite humaine ou le trafic des armes, il est possible d'atteindre des gains plus élevés moyennant un besoin d'investissement préalable moindre, et
 - le risque de détection par les autorités compétentes est moindre en raison de l'absence physique de l'agresseur « sur place »,

160 <https://www.switch.ch>

161 <https://fides.educa.ch/>

- l'élargissement du champ d'action des infrastructures de l'information et de la communication et de leurs valeurs généralement virtuelles et donc les dommages virtuels causés au monde physique via l'« Internet des objets », dans lequel les dommages prennent une forme physique et, le cas échéant, ne sont pas réversibles, ainsi que
- l'asymétrie des moyens entre les agresseurs et les agressés, c'est-à-dire la très large palette de motivation des agresseurs (allant des collaborateurs frustrés aux organisations d'information commerciales et étrangères, en passant par des criminels ou des organisations isolés) comportant des ressources considérables se trouve face aux moyens limités de reconnaissance, de prévention et d'élimination des risques¹⁶².

4.2.2.2 Besoin de protection qui en résulte

Ces facteurs de risque entraînent un besoin considérable de protection qu'il faut couvrir par des moyens organisationnels, techniques et juridiques appropriés, applicables de manière réaliste et exploitables. Les besoins de protection pertinents pour le secteur de la formation et de la recherche se caractérisent de la manière suivante :

- Protection contre l'intrusion à la vie privée, l'utilisation non intentionnelle (« *Mobbing* », « *Sexting* », etc.), la transmission, la falsification ou la suppression de données administratives pertinentes pour le fonctionnement ou particulièrement dignes de protection (données concernant les personnes [personnel enseignant, collaborateurs, élèves, étudiants, clients, par ex. leur adresse de domicile, leurs numéros de téléphone portable, etc. ainsi que toutes les données des services des ressources humaines relatives aux employés, aux chargés de cours, etc.], les comptes / le trafic des paiements de / à l'intention de collaborateurs, d'étudiants, de chargés de cours, de partenaires de recherche, etc., les informations relatives aux cartes de crédit [cartes de crédit d'instituts ou données relatives aux cartes de crédit de clients de la formation des adultes, etc.], les examens et les corrigés, les notes / expertises, etc.). Dans ces domaines, le respect des règles légales du devoir de diligence, de la responsabilité non transmissible des organes de direction en ce qui concerne la gouvernance nécessaire, de la loi sur la protection des données ainsi que des ordonnances applicables à cet effet revêt une importance significative.
- Protection de la propriété intellectuelle sous la forme de données d'application et d'enquête (connaissance auto-développée ainsi que propriété intellectuelle confiée à des tiers, par ex. dans le cadre d'une coopération de recherche avec tierces parties) contre l'utilisation illégale, la falsification, le retard, la communication anticipée, etc. Cette protection acquiert une importance croissante compte tenu des menaces à la hausse – les facteurs pertinents dans ce domaine sont :
 - la nécessité de protéger et de renforcer la Suisse en tant que place de formation, de production et de services dans un monde toujours plus compétitif, d'éviter les dommages correspondants par des opportunités manquées et d'éviter les dommages consécutifs (responsabilité envers des tiers, reproche de négligence, etc.) ;
 - la dépendance accrue de la force innovatrice de la Suisse face aux petites et moyennes entreprises, qui ne peuvent pas compenser suffisamment rapidement la perte de savoir pertinent en termes d'un avantage compétitif ou d'une caractéristique unique par des gains dans d'autres secteurs d'activité. Ces entreprises, précisément, coopèrent souvent avec les établissements de formation et de recherche ou sont des spin-offs qui présentent une forte implication des hautes écoles ;
 - la frontière toujours plus diffuse entre les tâches classiques des services de renseignements étrangers (protection de l'État, etc.) et la transmission d'informations recueillies en termes d'espionnage économique afin de promouvoir son propre pôle économique ;
 - l'utilisation secondaire observable d'outils matériels ou logiciels, développés initialement pour la reconnaissance militaire ou étatique ou pour la poursuite pénale, par des criminels isolés ou des organisations (chevaux de Troie dirigés contre l'État, outils de décryptage ou de déverrouillage de téléphones portables, etc.) jusqu'aux outils d'attaque qui peuvent être également acquis et utilisés par des particuliers motivés en la matière.

¹⁶² La plupart du temps, l'agresseur privilégie l'efficacité à l'élégance – il suit donc la voie du moindre effort. Par ailleurs, une faille unique inconnue de l'agressé est suffisante à l'agresseur, alors que l'agressé doit trouver et combler toutes les lacunes. Enfin, en raison du marché désormais établi d'achat ou de location de failles, l'agresseur dispose également d'un avantage face aux agressés qui agissent souvent isolément.

- Protection des interfaces techniques et organisationnelles avec les prestataires de services et les organisations partenaires (données d'application et d'enquête et métadonnées) dans de propres environnements TIC souvent très complexes et au vu d'une organisation opérationnelle ou de projet souvent temporaire, mettant en œuvre des employés à temps partiel et de fréquents changements de personnel dans le secteur informatique.
- Protection contre des interruptions d'activité dans les établissements d'enseignement et de recherche (par ex. « *Denial of Service* » – attaque par des élèves ou des étudiants lors d'un examen ou suite à un mauvais résultat d'examen) et contre le chantage (fréquemment par le biais d'attaques « *CryptoLocker* » ou « *DDoS* ») ou les attaques à mobile idéologique (par ex. dans le cas d'une utilisation de nouvelles formes d'apprentissage et d'enseignement, lors de thèmes socialement controversés de l'enseignement scolaire et lors de travaux et de coopérations de recherche tels que la recherche comportementale ou génétique, etc.).
- Protection des infrastructures informatiques de l'enseignement et de la recherche, très performantes, bénéficiant d'un bon réseau mais pas toujours gérées et surveillées de façon centrale, contre des abus causés par des outils d'attaque et de délits commis par des internes ou des tiers (envoi de « *Spam* », piratage de « *Botnet* », piratage de logiciels, enregistrement et distribution de données illégales, offre de services criminels ou commerciaux non autorisés, utilisation de logiciels illégaux, falsifiés ou sans licence, etc.).
- Protection de la bonne réputation des établissements de l'enseignement et de la recherche comme instances fiables et suffisamment sûres face à la société, aux organismes de financement étatiques et à l'économie privée. Compte tenu de la pression financière croissante, une bonne ou mauvaise réputation peut être décisive à cet égard pour les acceptations ou refus correspondants de ressources financières et pour l'attrait de l'environnement respectif.

Ce catalogue de menaces peut donner l'impression que seul un système de sécurité et une gestion des risques rigides, conçus, gérés et strictement surveillés selon des directives centrales, soient appropriés pour protéger convenablement la place suisse de la formation et de la recherche ainsi que les intérêts économiques et étatiques qui en découlent. Toutefois, deux facteurs pertinents entrent en contradiction directe avec une telle approche :

- Les contextes de l'enseignement et de la recherche vivent dans une large mesure de leurs habitudes d'ouverture – depuis un accès à peine contrôlable durant les heures d'ouverture des bâtiments, jusqu'aux interfaces de projets et de données temporaires à fonctionnement bilatéral, aux flux de logiciels et de données non contrôlables par le biais d'appareils privés, de supports de données mobiles, d'espaces de stockage externes en nuage (« *Cloud* »), en passant par l'utilisation de terminaux privés par les élèves, les étudiants et les collaborateurs, par la liberté de choix lors de la sélection/l'acquisition (souvent avec des privilèges élevés auto-administrés) d'infrastructures de serveurs, de systèmes d'exploitation, d'environnements de développement, de concepts d'exploitation et de maintenance, etc. par les collaborateurs. Ces habitudes d'ouverture représentent un véritable défi pour les responsables informatiques respectifs, mais elles sont souvent défendues avec véhémence par les utilisateurs.
- Les ressources et la force d'impact des services informatiques locaux respectifs sont limitées et contraignent les responsables informatiques à se concentrer sur un nombre restreint de domaines de travail maîtrisables (assistance de premier et de deuxième niveau, fonctionnement, surveillance et maintenance des environnements TIC centraux pour la formation et l'administration, définition et mise en œuvre d'un concept d'identité, d'accès et de rôles / droits, protection des parties administratives de l'infrastructure en interne et en externe par un concept de zones basé sur firewall / DMZ, etc., service de piquet limité et traitement des « *Trouble Tickets* », incidents / défaillances, etc.). Par ailleurs, les services opérationnels sont de plus en plus externalisés, ce qui permet généralement aux prestataires de services respectifs d'avoir un accès la plupart du temps étendu à l'infrastructure ainsi qu'aux données et métadonnées (fichiers journaux, etc.) qui y sont enregistrées, transférées ou traitées. Dans ce domaine, les établissements de l'enseignement et de la recherche bénéficient aujourd'hui déjà de services centraux (SWITCH-CERT, stockage de fichiers et échange de données sécurisés via les services SWITCH,

etc.) – une utilisation encore plus large de tels outils centralisés serait tout à fait possible d'un point de vue technique, mais nécessiterait toutefois aussi bien des investissements supplémentaires considérables dans la mise à disposition d'environnements d'exploitation informatiques suffisamment performants, qu'un large consensus en ce qui concerne les « règles du jeu » opérationnelles et les axes stratégiques pour les développements à venir.

Pour définir et faire fonctionner un dispositif adéquat de sécurité de l'information permettant d'utiliser des données dans le paysage de la formation et de la recherche en Suisse, il est essentiel de conserver la liberté de l'enseignement et de la recherche et d'examiner dans le détail la diversité des menaces, la situation en matière de ressources et l'exploitation de mesures de sécurité correspondant aux différents degrés de formation en Suisse – il ne serait donc guère difficilement applicable à une structure fortement fédéraliste et particulière un système de gouvernance et de contrôle central et rigide.

4.2.3 Solutions

Sur la base des réflexions ci-avant relatives à la situation de la menace et au besoin de protection, les solutions pour gérer la sécurité de l'information dans la formation et la recherche doivent satisfaire les conditions-cadres suivantes :

- La situation de la menace et le besoin de protection présentent de grandes différences entre les degrés primaire / secondaire et le degré tertiaire.
- Le système éducatif organisé de façon fédéraliste et marqué prioritairement par les cantons continue, dans tous les aspects, à être très diversifié – ce qui restera également ainsi.
- La taille et la complexité des organisations de formation autonomes sont très diverses et l'utilisation et l'intensité de l'utilisation d'outils informatiques et, ce qui en découle, la complexité des paysages informatiques existants présentent donc une très importante variabilité.
- La disponibilité du personnel spécialisé en sécurité informatique va d'une absence de disponibilité ou d'un rôle à temps partiel pour les enseignants des petites écoles primaires communales, jusqu'à de grandes divisions informatiques hautement professionnalisées, disposant de propres centres de compétences en matière de sécurité informatique, dans les grandes institutions universitaires et les organisations de recherches d'envergure.
- La liberté de l'enseignement et de la recherche revêt une importance élevée et se trouve en contradiction, ou pour le moins en concurrence, avec les exigences de la sécurité de l'information et des bonnes pratiques (« *Good Practice* ») actuelles.
- Actuellement, les élèves et étudiants sont pratiquement tous des natifs de l'ère numérique, ce qui conduit à une accélération de la numérisation de l'enseignement. Il en résulte une délocalisation toujours plus rapide des activités principales dans l'espace virtuel, délocalisation qui a, aujourd'hui également, atteint le niveau primaire. Par conséquent, la dépendance de l'enseignement face aux outils informatiques est toujours plus devient de plus en plus forte ; la sécurité et la protection des données deviennent rapidement plus importantes.
- La recherche est aujourd'hui déjà fortement numérisée. Ici, l'utilisation de technologies de pointe, comme par ex. le « *High Performance Computing* » (HPC), ou l'analyse, le traitement et le transport de très grandes quantités de données (« *Big Data* »), mais aussi l'utilisation de capteurs toujours plus performants, revêtent une importance croissante. L'évolution fut et demeure fulgurante.
- À l'avenir, les fonds destinés à l'enseignement et à la recherche n'augmenteront guère, les dépenses liées à l'agrandissement et au fonctionnement d'outils informatiques poussés par la numérisation s'accroîtront cependant significativement, raison pour laquelle les mêmes fonds doivent être utilisés de manière très sélective et appropriée – des coopérations sont essentielles pour permettre une utilisation optimale des ressources financières.
- La densité de la réglementation augmente rapidement au niveau de la sécurité de l'information et de la protection des données.

Afin de développer des approches de solutions, les conditions cadres énumérées ci-dessus, l'évolution de la situation de la menace et le besoin de protection constaté permettent de tirer les conclusions suivantes :

- Les fortes différences dans la situation de la menace et dans le besoin de protection doivent être prises en considération par le biais d'une procédure fondamentalement basée sur le risque. Des organisations similaires seront très probablement exposées à des situations de menaces similaires et présenteront des besoins similaires. Par conséquent, il est également possible de réduire de tels risques par le biais de mesures du même type et de synergies correspondantes sur le plan organisationnel et technique.
- En dépit de la diversité fédérale, des différents systèmes scolaires présentant différentes structures organisationnelles et différents états d'avancement de la numérisation, ainsi que des paysages de systèmes informatiques établis, il faut s'attendre à ce que l'on trouve des exigences comparables en termes de sécurité de l'information et de sécurité informatique¹⁶³ avec un potentiel élevé de synergie dans de nombreux domaines thématiques à un même degré de formation.
- Une approche centralisée et uniforme de la gestion de la sécurité de l'information dans l'enseignement et la recherche n'est ni judicieuse ni réalisable. Toutefois, l'élaboration de ce que l'on appelle les bonnes pratiques pour la gestion de la sécurité de l'information et l'architecture de la sécurité informatique pour des catégories d'organisations est tout à fait envisageable. De telles bonnes pratiques doivent pouvoir être mises en place de façon modulable, afin de pouvoir être adaptées simplement à différentes conditions cadres.
- Les organisations qui ne possèdent peu ou pas de savoir-faire en informatique et en sécurité informatique devront collaborer ou trouver un fournisseur approprié de services informatiques. Dans une large mesure, ce cas de figure est valable pour la sécurité informatique, car les spécialistes possédant cette spécialisation sont une denrée rare, ce qui a comme conséquence un recrutement difficile ou impossible, mais implique également des revendications salariales élevées. Une détente du marché du travail n'est pas prévisible dans ce domaine.
- En tant que part de leur responsabilité de gestion non déléguable, les directions d'écoles devront, à tous les niveaux, décider comment concilier, d'une part, la liberté de l'enseignement et de la recherche, et, d'autre part, une sécurité suffisante. En conséquence, elles devront supporter les risques résiduels qui en résultent. Par ailleurs, cette pesée des intérêts devra prendre en compte la convivialité d'utilisation ainsi que les aspects économiques et opérationnels. L'élaboration et la mise en œuvre des bonnes pratiques en matière d'enseignement et de recherche mentionnées ci-dessus constitueront un fil conducteur primordial à cet égard.
- La numérisation est déjà bien avancée dans le domaine de la recherche. Ici, les défis résideront dans le fait que les quantités de données augmentent massivement, avant tout dans le domaine technique et dans celui des sciences naturelles. La mise à l'échelle nécessaire des systèmes informatiques (capacités de calcul, de transfert, mémoire, etc.) pourra être financable uniquement sur une base commune. En ce qui concerne la sécurité informatique, de nouveaux concepts, architectures et modèles opérationnels plus évolutifs et plus financables sont également nécessaires à cet effet.
- L'augmentation des besoins d'investissement et des frais d'exploitation en matière d'informatique et de sécurité informatique, combinée à la stagnation du budget issu des pouvoirs publics consacré à l'enseignement et à la recherche, constituent un défi. Ce dernier peut être géré uniquement par une augmentation de l'efficacité, une collaboration sur le plan national à tous les niveaux de formation ainsi qu'une standardisation de l'informatique de base¹⁶⁴. Dans une large mesure, cette dernière comprend également la sécurité informatique.

¹⁶³ Il est nécessaire de clairement différencier la « sécurité de l'information » de la « sécurité informatique ». La sécurité de l'information comprend l'ensemble des mesures techniques et organisationnelles à des fins de protection d'informations sous toutes leurs formes (par ex. imprimées, orales, enregistrées, etc.) au sein d'une organisation, alors que la sécurité informatique, en tant que sous-domaine de la sécurité de l'information, se limite aux mesures techniques de protection des données électroniques et des plateformes de traitement. Étant donné que l'efficacité des mesures de sécurité dépend notamment de la connaissance de leur contexte de création et d'utilisation et que cette connaissance est plutôt possible pour les informations que pour les données brutes, conformément au modèle consistant à affiner les données brutes en informations, nous parlons par ailleurs systématiquement de « sécurité de l'information » et non de « sécurité des données ».

¹⁶⁴ Par informatique de base, nous entendons l'infrastructure informatique de base que chaque organisation exploite et qui ne contribue pas à la distinction entre les institutions et les écoles. La sécurité informatique fait principalement partie de cette informatique de base.

- Des exigences accrues au niveau légal, comme par ex. celles découlant du nouveau règlement de base sur la protection des données de l'UE, de la révision en cours de la législation relative à la protection des données en Suisse, mais aussi d'autres réglementations qui se dessinent dans le domaine de la sécurité de l'information et de celle de l'informatique, accroîtront de façon accélérée la charge de travail nécessaire pour élargir et faire fonctionner la sécurité informatique (en comparaison avec l'informatique en général).

Sur la base des conditions-cadres formulées et des conclusions qui en sont tirées, nous aborderons ci-dessous plus en détail quelques aspects partiels de l'ébauche de solution proposée.

4.2.3.1 Aspects conceptuels

Dans l'enseignement, il convient de prendre en compte séparément les organisations des degrés primaire et secondaire et celles du degré tertiaire, en raison du profil de risque relativement différent entre elles. Cela conduira à deux structures qui n'ont en partie pas tout à fait la même configuration ou la même pondération,¹⁶⁵ mais qui ont toutefois une base commune¹⁶⁶.

La définition d'une approche conceptuelle pour gérer la sécurité de l'information et la sécurité informatique dans l'enseignement et la recherche ne devrait pas être abordée sur une base entièrement nouvelle. Les conditions-cadres et la situation initiale présentent des similitudes avec l'économie. Par conséquent, des concepts éprouvés peuvent être repris dans le secteur de l'enseignement et de la recherche.

Une approche établie et pragmatique consiste en une procédure subdivisée en trois niveaux :

1. Niveau 1 : hygiène de base

S'applique aux petites et moyennes organisations comportant un besoin de protection normal, un traitement moins complexe de l'information et sans propre savoir-faire spécialisé dans la sécurité informatique. Mise en œuvre de 10 à 20 mesures de protection fondamentales au maximum, nécessaires pour toute organisation afin d'atteindre une « hygiène de base » dans la sécurité de l'information – le cas échéant, échelonnée selon des critères tels que l'âge des apprenants, le degré de maturité de l'établissement d'enseignement, etc. Un administrateur informatique qui possède des connaissances normales en matière de sécurité doit pouvoir mettre en œuvre ces mesures. Il ne faut pas réinventer une telle structure ; celle-ci existe déjà, par ex. sous la forme de guide pour PME auprès de MELANI¹⁶⁷, ou comme programme en 10 plus 10 points d'InfoSurance sur le site Internet d'ISSS¹⁶⁸. La mise en œuvre et l'application systématique de ces mesures portent la sécurité de telles organisations à un niveau qui ne les rend pas attrayantes en tant que cibles.

2. Niveau 2 : hygiène de base élargie

Ce niveau se base sur le niveau 1 et ajoute, à l'aide des résultats d'un catalogue de questions dans des domaines thématiques spécifiques, des mesures de protection supplémentaires par rapport à l'hygiène de base. Ce niveau entre en ligne de compte lorsqu'un ou plusieurs des attributs qui suivent s'appliquent : organisation à plusieurs niveaux et, par conséquent, processus et interfaces plus complexes, besoin accru de protection dans des domaines définis, traitement plus complexe de l'information et environnement des systèmes plus sophistiqué. Certains concepts existent déjà à ce niveau, par ex. en tant que résultats issus des travaux d'un sous-groupe de travail de la Commission « Avenir du traitement et de la sécurité des données en Suisse » du Département fédéral des finances (DFF)¹⁶⁹.

¹⁶⁵ Structure : collection de concepts se basant sur un savoir et une expérience récents, applicables pour la sécurité de l'information ainsi que la sécurité informatique d'un groupe cible défini d'organisations, y compris les modèles pour la mise en œuvre ainsi que la documentation.

¹⁶⁶ La numérisation est plus avancée dans la recherche que dans l'enseignement. Ici aussi, de grandes différences résident en termes de complexité et d'exigences envers la sécurité informatique. Les plus grands défis se trouvent aujourd'hui déjà dans les champs de recherche techniques et des sciences naturelles, cette situation se renforcera encore en raison du développement fulgurant des connaissances et de la technique. De ce fait, il convient de considérer séparément la recherche technique et celle dans le domaine des sciences naturelles des autres champs de recherche.

¹⁶⁷ Lien : <https://www.melani.admin.ch/melani/fr/home/documentation/listes-de-contrôle-et-instructions/secure-informatique--aide-memoire-pour-les-pme.html> (aide-mémoire de MELANI relatif à la sécurité de l'information pour les PME).

¹⁶⁸ Lien : https://www.kmu.admin.ch/dam/kmu/fr/dokumente/savoir-pratique/Informatique-et-IT/InfoSurance_10_Points_Programme_FR.pdf.download.pdf/InfoSurance_10_Points_Programme_FR.pdf (programme InfoSurance en 10 points).

¹⁶⁹ Groupe d'experts relatif à l'avenir du traitement et de la sécurité des données en Suisse (travaux clôturés en 2018 et groupe dissous par la suite) : les informations abordées ne sont actuellement pas accessibles au public.

3. Niveau 3 : approche basée sur les risques et système de management concernant la sécurité de l'information (SMSI)

S'applique davantage aux grandes organisations multinationales, mais aussi aux organisations dont la création de valeur est fortement dépendante du traitement de l'information, qui traitent systématiquement les données critiques ou sensibles de clients et/ou qui présentent un traitement complexe de l'information ainsi qu'un environnement sophistiqué des systèmes informatiques. De telles organisations doivent pleinement mettre en œuvre les mesures des niveaux 1 et 2 et déduire par la suite, sur la base d'une analyse de risques et de besoin de protection, quelles mesures supplémentaires il est nécessaire de mettre encore en œuvre. Elles doivent systématiquement traiter les risques de sécurité de l'information dans le cadre d'un système de management de la sécurité de l'information (SMSI) et améliorer continuellement ce dernier. Ici aussi, il existe des approches disponibles publiquement, comme par ex. ce qu'on appelle la norme minimale élaborée par l'Office fédéral pour l'approvisionnement économique du pays (OFAE)¹⁷⁰. Avec une charge de travail maîtrisable, il est possible d'adapter cette norme aux conditions-cadres spécifiques de l'enseignement. ISO 27001 est une norme acceptée et praticable internationalement pour mettre en place et faire fonctionner un SMSI, qu'il est possible d'adapter à toute taille d'organisation et à des conditions cadres spécifiques¹⁷¹.

De manière générale, la décision d'adopter une procédure appropriée doit se prendre sur la base d'une évaluation échelonnée et complète du besoin de protection. La question se pose avant tout de déterminer si un besoin élevé de protection est nécessaire, et, le cas échéant, où celui-ci s'opère. Les questions qui suivent sont importantes à cet égard :

- À quel point la création de valeur dépend-elle d'informations (correctes) et de leur disponibilité ?
- L'organisation traite-t-elle et/ou stocke-t-elle systématiquement des informations critiques provenant de tiers (par ex. des données personnelles sensibles), ou a-t-elle accès à de telles données ?

Les facteurs pertinents qui permettent de décider du niveau de protection à appliquer sont mentionnés ci-dessous :

- Taille de l'organisation
- Complexité du traitement de l'information et de l'environnement des systèmes informatiques
- Ressources et savoir-faire en matière de sécurité de l'information et de sécurité informatique
- Ressources financières

Le management de la sécurité de l'information est une tâche de gestion. La mise en place et l'exploitation nécessitent un mandant / sponsor issu des cadres supérieurs de l'administration de l'éducation. Une mise en œuvre couronnée de succès doit avant tout être dirigée par un cadre. Par conséquent, le responsable du processus (préposé à la sécurité ou responsable de la sécurité des systèmes informatiques, RSSI) doit avoir un poste de dirigeant suffisamment important. Les conflits d'intérêt, par ex. en subordonnant le RSSI à l'informatique, doivent être évités. Il faut séparer la création des prescriptions (directives, etc.) et les activités d'audit de l'ingénierie et du fonctionnement de la sécurité informatique.

En raison du nombre important d'organisations comparables pour les degrés primaire et secondaire, nous recommandons l'élaboration d'une structure et d'une procédure spécifiques pour ce secteur partiel sur la base des travaux préliminaires mentionnés ci-dessus. De ce fait, une approche spécifique aux branches basée sur les bonnes pratiques est instaurée, qui facilite la mise en œuvre avec succès et accroît la qualité des résultats. Par ailleurs, une telle collaboration permet une charge de travail globale significativement moindre (sur le plan national) et accroît la comparabilité des SMSI. En outre, un moyen éprouvé réside en la mise en place de groupes d'échange d'expériences dans le domaine de la sécurité de l'information et de la sécurité informatique avec des organisations intéressées issues des degrés primaire et secondaire.

¹⁷⁰ Lien : https://www.bwl.admin.ch/bwl/fr/home/themen/ikt/ikt_minimalstandard.html (Office fédéral pour l'approvisionnement économique du pays (OFAE), norme minimale pour les TIC).

¹⁷¹ Nous recommandons une certification du SMSI selon ISO 27001 uniquement lorsque elle est exigée administrativement ou légalement, ou lorsqu'il est possible d'obtenir une plus-value. Dans le cas contraire, il résulte un surcroît de travail sans création de plus-value.

L'échange d'expérience à un niveau de confiance élevé est très efficace dans de tels organes. Des écoles motivées peuvent également s'entraider lors d'événements difficiles dans le sens d'une « aide accordée aux voisins ». À cet égard, educa.ch pourrait prendre en charge la plateforme de confiance ainsi que la médiation. De telles structures existent déjà au degré tertiaire. SWITCH mène avec grand succès un « *Security Working Group* » pour spécialistes en sécurité informatique ainsi qu'un « *ISMS Working Group* » à l'intention des RSSI. Ce dernier dispose d'un sous-groupe de travail qui se consacre entièrement à la thématique de la sensibilisation (« *Awareness* »).

Les organisations qui, après analyse de leur situation initiale et de la classification, se voient incapables d'effectuer elles-mêmes les travaux nécessaires, doivent chercher un soutien externe en faisant par ex. appel à des conseillers spécialisés pour un projet relatif à la sécurité, ou à un RSSI¹⁷² « *On Time* » à temps partiel, qui assume la responsabilité du fonctionnement de la sécurité de l'information et de la sécurité informatique. À cet égard, si la demande est suffisante, un groupe de spécialistes peut être constitué de manière centralisée pour la branche.

Alternativement, dans de tels cas, il est vivement recommandé de vérifier l'externalisation (« *Outsourcing* ») de la sécurité informatique ou de l'informatique comme variantes de solution. Afin de choisir le partenaire adéquat, il faut faire appel à un soutien externe qui dispose des connaissances et de l'expérience requises. Par ailleurs, l'aspect décisif déterminant le succès d'une externalisation réside en ce que suffisamment de savoir-faire soit à disposition en interne afin de contrôler et de piloter le ou les fournisseurs de services (« *Service Provider* »). Toutefois, la responsabilité du maintien d'une sécurité appropriée de l'information ne peut être déléguée aux instances dirigeantes de l'organisation en question.

Ces réflexions s'appliquent également au degré tertiaire. Ici, des efforts communs ont déjà commencé pour définir une approche selon les bonnes pratiques pour la mise en place de SMSI et créer ensemble des modèles de lignes directrices (« *Policy Templates* »). Les spécialistes de la sécurité informatique et les RSSI d'organisations de plus grande envergure échangent régulièrement des informations lors de manifestations privées. SWITCH sert de plateforme de confiance pour de telles activités. Dans le domaine de la sécurité informatique, SWITCH fournit également des prestations de services spécifiques à la communauté (CERT¹⁷³, « *Firewall DNS* »¹⁷⁴, « *Identity Management* » [gestion de l'identité]). Cette collaboration étroite entre les hautes écoles dans le domaine de la sécurité a fait ses preuves avec succès. À l'avenir, elle continuera d'être élargie afin de gérer des menaces toujours plus complexes.

4.2.3.2 Aspects techniques

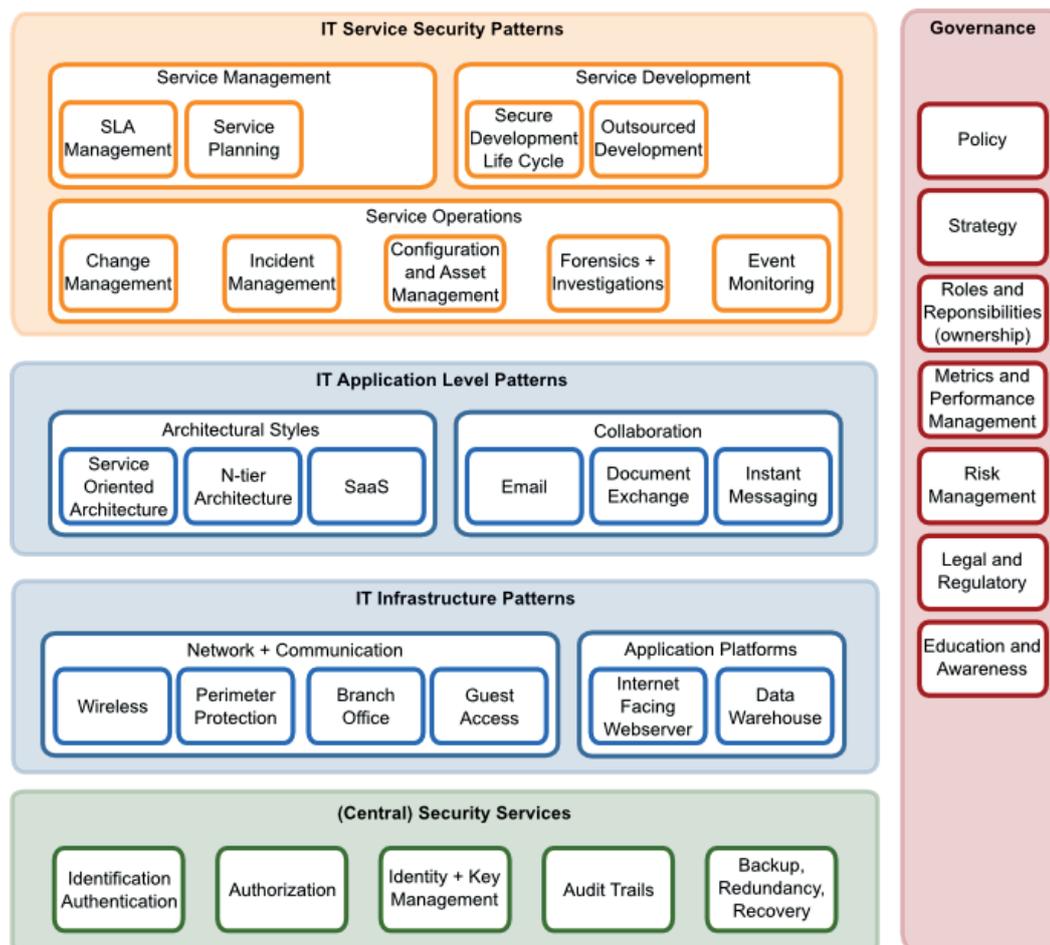
Afin de pouvoir effectuer un examen systématique des thèmes techniques, nous nous basons sur un modèle d'architecture de sécurité informatique de référence et nous nous prononçons de manière fondamentale sur les principaux composants. Les aspects techniques dépendent dans une très large mesure du modèle opérationnel choisi – avec des variantes allant de l'exploitation de la sécurité informatique entièrement en interne, à l'externalisation de l'ensemble de celle-ci. Vous trouverez des explications relatives à ce sujet au chapitre « Aspects conceptuels ».

Les explications du présent chapitre se basent sur le « *Security Architecture Landscape* » de l'« *Open Security Alliance* », présenté dans le graphique ci-dessous.

172 L'abréviation CISO signifie « *Chief Information Security Officer* », ou, en français, responsable de la sécurité des systèmes informatiques (RSSI) d'une organisation. Le RSSI est propriétaire des processus de la gestion de la sécurité de l'information.

173 « *Computer Emergency Response Team* »

174 DNS est l'abréviation anglaise de « *Domain Name System* » et désigne un service de base indispensable à l'exploitation de réseaux ainsi qu'Internet. On appelle communément « *Firewall DNS* » une technologie basée sur DNS RPZ qui permet de bloquer l'accès à des noms de domaine utilisés abusivement.



Graphique 10 : Security Architecture Landscape de l'Open Security Alliance (OSA)¹⁷⁵

Les champs thématiques gouvernance (« *Governance* ») et modèles de sécurité des services informatiques (« *IT Service Security Patterns* ») sont, en premier lieu, orientés vers les processus et les organisations et se concentrent sur le pilotage, la planification et le fonctionnement de la sécurité informatique. Les deux champs offrent un grand potentiel pour une collaboration dans le cadre des degrés primaire et secondaire ainsi que tertiaire. Les trois autres champs thématiques techniques offrent beaucoup d'opportunités dans le domaine de l'infrastructure, avec toutefois moins de potentiel de synergie pour les applications et prestations de services, car les paysages actuels sont très différents entre eux dans le domaine des applications. Dans le domaine des « *IT Infrastructure Patterns* » et des « *(Central) Security Services* », il existe des domaines thématiques où une coordination transversale entre les degrés et une standardisation et/ou une exploitation commune sont judicieuses ou nécessaires.

Governance

Même si la « *Governance* » n'a pas d'action directe sur la technique, elle est toutefois d'une grande pertinence pour la sécurité informatique, car elle pilote sa conception et son fonctionnement, raison pour laquelle nous abordons ici brièvement ce champ thématique. La « *Governance* » est représentée comme une fonction transversale, car elle pilote tous les niveaux de l'architecture de la sécurité informatique. Le chapitre 4.2.3.1. Aspects conceptuels est consacré à quelques thèmes traités dans ce champ.

Une collaboration entre les institutions en vaut la peine, même si ces thèmes doivent être mis en œuvre séparément en ce qui concerne les conditions cadres locales spécifiques et en prenant en compte la propre culture de l'organisation concernée. La collaboration établie des RSSI d'organismes de l'éducation du degré tertiaire le montre dans le « *ISMS Working Group* » géré par SWITCH. Les sept thèmes de ce champ thématique présentent de nombreuses similitudes avec des organisations comparables. Voilà pourquoi il

¹⁷⁵ Source : <http://www.opensecurityarchitecture.org/cms/foundations/osa-landscape>

vaut la peine d'élaborer en commun une « bonne pratique » et des modèles de documents et d'outils associés, qui peuvent alors être adaptés par l'organisation individuelle, moyennant une faible charge de travail et en tenant compte des conditions cadres, des pratiques et des spécificités organisationnelles locales. Ainsi, la charge de travail est réduite pour l'institution individuelle et la structure mise en place à l'aide d'une intelligence commune mène à une meilleure qualité de la conception et des documents.

Dans le cas d'une externalisation de la sécurité informatique également, l'organisation concernée doit s'assurer qu'elle puisse garantir la « *Governance* » par ses propres moyens. C'est seulement ainsi qu'elle est en mesure de piloter le prestataire de services externe et de surveiller son activité.

Pour les degrés primaire et secondaire, il vaut donc la peine d'examiner une collaboration sur le plan national dans ce domaine.

La nécessité de mettre en place et de maintenir une sensibilisation appropriée à la problématique mais aussi aux limitations de la sécurité de l'information et de la protection des données (« *Security Awareness* ») est commune à tous les degrés de formation et à toutes les personnes impliquées. Alors que les campagnes et les activités correspondantes du personnel enseignant et des chercheurs se concentrent non seulement sur les aspects techniques, mais aussi sur les éléments sociaux, tels que la fonction de modèle ou de supérieur hiérarchique et la sensibilisation continue des apprenants à l'aide de cas pratiques récents, des initiatives d'« *Awareness* » ludiques et bien intégrées dans les structures d'âge et les structures sociales doivent transmettre, en fonction du degré de formation, la pertinence du sujet et la responsabilité personnelle des personnes concernées. Dans ce domaine également, il existe un potentiel élevé de synergie pour la planification, la mise en œuvre, l'évaluation et l'amélioration conjointes de campagnes de ce type.

Modèles de sécurité des services informatiques (« IT Service Security Patterns »)

Dans ce champ thématique, les mêmes considérations s'appliquent en principe comme pour la gouvernance (« *Governance* ») (chapitre précédent) : la collaboration en vaut la peine aussi bien aux degrés primaire / secondaire que tertiaire.

Toutefois, il convient de faire des déclarations supplémentaires sur des thèmes individuels.

- Fonctions de gestion des services (« *Service Management* ») : dans le cas d'une externalisation de la sécurité informatique, l'organisation doit s'assurer qu'elle peut piloter la gestion des services par ses propres moyens. C'est uniquement de cette manière qu'elle est en mesure d'assurer la qualité des services ainsi qu'une planification suffisante des capacités.
- Fonctions de développement des services (« *Service Development* ») : dans le cas d'une externalisation, ces fonctions sont de la responsabilité du fournisseur de service (« *Service Provider* »). Dans le processus d'évaluation, les organisations qui effectuent l'externalisation devraient toutefois demander des renseignements sur la manière dont le fournisseur de services a régulé ces processus.
- Fonctions d'opérations des services (« *Service Operations* ») : le thème « criminalistique + investigation » (« *Forensics + Investigation* ») nécessite des experts hautement spécialisés. Développer de telles capacités n'en vaut la peine pour aucune organisation à quelque degré du système éducatif que ce soit. Une variante consiste à mettre cette fonction à disposition de façon centrale, tel que l'a déjà fait SWITCH-CERT au degré tertiaire. Alternativement, on achète cette fonction sur le marché auprès d'un fournisseur de services spécialisé. Les quatre autres fonctions sont d'une importance capitale pour une gestion couronnée de succès de la sécurité informatique et méritent une priorité élevée ainsi qu'une attention particulière. Ceci s'applique avant tout à la configuration et à la gestion des actifs (« *Configuration and Asset Management* »), en deuxième lieu à la gestion des changements et des incidents (« *Change and Incident Management* ») et, par la suite, à la surveillance des événements (« *Event Monitoring* »). Dans le cas d'une externalisation (« *Outsourcing* »), la gestion des changements et des incidents (« *Change and Incident Management* ») dispose au moins d'interfaces avec le fournisseur de service. Dans de nombreux cas, ce dernier nécessitera une collaboration intense afin de gérer les processus correspondants dans la qualité requise, parce qu'il a besoin à cet égard d'informations sur le contexte spécifique de l'organisation. Pour toutes les institutions qui gèrent par elles-mêmes l'informatique et la sécurité informatique, une collaboration dans le domaine de la surveillance des événements (« *Event*

Monitoring ») en vaut en tous les cas la peine. Cette collaboration peut débuter lors de l'identification et de la mise en œuvre de cas d'utilisation (« *Use Cases* ») intéressants et poursuivre son évolution jusqu'à l'exploitation d'une infrastructure centrale commune permettant de gérer les données de journal (« *Log Data* ») et leur analyse.

Modèles des niveaux d'applications informatiques (« IT Application Level Patterns »)

Comme mentionné ci-dessus dans le résumé de l'évaluation, nous attendons dans ce champ thématique la plus grande diversité (historique) des paysages. Par conséquent, les améliorations et gains d'efficacité réalisables à court terme restent limités à ce niveau. Toutefois, il est d'autant plus passionnant de ne pas perdre de vue ce qui se passe dans le paysage des applications et l'architecture de sécurité dans la communauté. Il est prévu que l'on puisse exploiter d'intéressants potentiels à moyen et à long terme, par ex. avec l'élaboration d'approches de bonnes pratiques dans des thèmes spécifiques. Les services de collaboration (« *Collaboration* »), mentionnés explicitement à la figure et séparément en raison de leur importance, peuvent constituer un tel thème.

Modèles des niveaux d'infrastructure informatique (« IT Infrastructure Level Patterns »)

Ce domaine thématique recèle quelques thèmes très intéressants, notamment dans les réseaux et la communication (« *Network + Communication* »). Nous voyons moins de potentiel dans les plateformes d'applications, bien qu'un échange entre spécialistes relatif aux expériences vécues et aux bonnes pratiques peut y être intéressant. Pour ces composantes d'infrastructure toutefois, les bonnes pratiques sont la plupart du temps définies par les fournisseurs.

Pour les thèmes regroupés sous « réseau + communication » (« *Network + Communication* »), la situation se présente toutefois d'une manière significativement différente. Notamment pour les degrés primaire et secondaire, il convient de développer dans les thèmes « protection du périmètre » (« *Perimeter Protection* ») et « sans fil » (« *Wireless* ») des concepts communs de bonnes pratiques. Par ex., il n'est pas clair pour quelle raison il devrait y avoir des approches de conception fondamentalement différentes pour la segmentation du réseau. Il en va de même pour le thème « bureau de succursale » (« *Branch Office* »), qui, dans la recherche et l'enseignement, s'illustre plutôt comme une mise en réseau de diverses écoles et campus par le biais d'un réseau étendu (« *Wide Area Network (WAN)* ») ou d'un réseau métropolitain (« *Metropolitan Area Network [MAN]* »). À cet effet également, il convient d'élaborer une recommandation commune de bonnes pratiques. Le degré tertiaire le prouve avec le Réseau national de la recherche et de l'enseignement (« *National Research and Education Network* » [NREN]) exploité en commun.

Comme les expériences du degré tertiaire (prestation de services établie internationalement sous le nom « *eduroam* ») permettent de le reconnaître facilement, il vaut la peine pour le thème « accès invité » (« *Guest Access* ») de faire un pas supplémentaire et d'établir une solution que l'on peut fédérer sur le plan national. Parmi les hautes écoles et les organisations de recherche au niveau mondial, même la Fédération internationale des fédérations (d'écoles nationales) a rencontré un franc succès. La convivialité d'utilisation de sa solution en est la raison. Avec celle-ci, l'architecture de sécurité correspondante est automatiquement standardisée. Les degrés primaire et secondaire ne sont pas concernés par cette dimension internationale. Toutefois, une fédération nationale, telle qu'elle est actuellement élaborée dans le projet FIDES, apporte clairement une valeur ajoutée.

(Central) Security Services

Les services de sécurité centraux (« *Central Security Services* ») définissent l'infrastructure de base de la sécurité informatique et mettent à disposition une fonctionnalité centrale. Une collaboration dans ces différents domaines est nécessaire et importante, bien qu'elle se fait selon différents aspects :

- Les modules « identification », « authentification », « autorisation » et « gestion de l'identité » (« *Identification, Authentication, Authorization and Identity Management* ») réglementent ensemble ce qu'on appelle la gestion des identités et des accès (« *Identity and Access Management* », [IAM]). À l'heure de la numérisation, où les « limites » des systèmes disparaissent toujours plus et où la protection des périmètres perd de son efficacité et de son importance, cet IAM devient toujours plus essentiel, car une identification sans équivoque des utilisateurs ainsi qu'un contrôle étroit des droits d'accès de ceux-ci deviennent le facteur décisif de sécurité. Les composantes de la gestion des accès doivent impérativement rester entre les mains des organisations individuelles, alors que, dans le cas de la gestion des identités, il est

important que les approches choisies soient compatibles et qu'une continuité de l'identité puisse donc être garantie à tous les degrés scolaires et dans la formation continue ultérieure de l'individu.

L'identité « AAI », actuellement remplacée par le SWITCH edu-ID lié à l'individu et à usage perpétuel, établie sur l'ensemble du territoire au degré tertiaire et liée aux institutions, indique à quel point une gestion de l'identité bien coordonnée peut être précieuse. Pour l'AAI (et, par la suite également le SWITCH edu-ID), une fédération établie dans toute la Suisse est en fonction et, avec de nombreux autres États ou leurs hautes écoles, constitue une fédération transfrontalière des fédérations. Cette identité académique utilisable sur l'ensemble du territoire est aussi la base de services à valeur ajoutée, comme par ex. l'eduroam mentionné ci-dessus, et celle d'une collaboration internationale simplifiée dans l'enseignement et la recherche. Un IAM très bien adapté et cohérent dans l'ensemble du secteur de l'enseignement et de la recherche apporte une très grande valeur ajoutée, tout en étant indispensable au vu de la numérisation. Cela est déjà très bien établi dans le degré tertiaire. Il est important que les degrés primaire et secondaire intègrent le système existant. Une solution de ce type est momentanément élaborée dans le projet FIDES.

- La gestion des clés (« *Key Management* ») constitue le processus critique pour mettre en œuvre les procédures cryptographiques. Ces dernières jouent également un rôle essentiel dans la sécurité informatique, ici aussi avec une importance à la hausse au vu d'une criminalité sur internet qui se développe de façon fulgurante. Dans la cryptologie, l'utilisation de procédures et d'algorithmes connus publiquement et classés comme sûrs est nécessaire. Les bonnes pratiques correspondantes doivent être développées en commun aux différents degrés.
- La sauvegarde, la redondance et la récupération (« *Backup, Redundancy and Recovery* ») sont des thèmes supplémentaires dans lesquels une étroite collaboration est indiquée. Il n'est pas pertinent de développer ici des concepts sur une base complètement nouvelle ; les infrastructures nécessaires et onéreuses, du moins dans le cas d'une sauvegarde (« *Backup* »), peuvent être exploitées et utilisées en commun. Pour gérer de graves accidents, une collaboration régionale d'un groupe d'institutions au sens d'une « aide accordée aux voisins » est tout à fait pertinente.
- Enfin et surtout, les « audit trails » revêtent une importance cruciale. Ces derniers sont soumis à des exigences élevées de sécurité, notamment en matière d'intégrité des informations et de réglementation claire de l'accès. Ici aussi, la recommandation consiste à définir en commun une approche de bonnes pratiques aux différents degrés.

Différences entre formation et enseignement d'une part et recherche d'autre part

La formation scolaire et l'enseignement académique d'une part et la recherche d'autre part présentent de grandes différences en termes de niveau de développement et de degré de maturation des solutions informatiques et de la sécurité informatique. La numérisation bien avancée dans la recherche et l'accroissement rapide des volumes de données à traiter en sont les causes essentielles. Cela conduit à ce que, aujourd'hui déjà, il faut chercher des solutions innovantes dans la recherche pour le traitement, le stockage et le transport de très grandes quantités de données. La sécurité informatique est également prise en compte dans ce processus. La « science DMZ », une nouvelle approche pour la transmission évolutive, sûre et à très haut débit de grandes quantités de données, constitue un exemple d'un tel concept.

Nouvelles technologies pertinentes

Dans le domaine du soutien technologique de la sécurité de l'information, il existe une large offre de produits et de services opérationnels ou conceptuels, difficiles à évaluer par les profanes. Il paraît donc sensé de prévoir les activités suivantes dans le cadre des efforts communs :

- Identification des technologies de base de la sécurité de l'information, telles que la gestion des identités numériques, l'identification des vulnérabilités actuelles et des contre-mesures possibles, la collecte et l'évaluation des journaux système en ce qui concerne des menaces identifiables, les architectures de réseau avec différentes zones de protection et voies d'accès, le renforcement des systèmes contre les attaques, la sécurité des appareils mobiles, etc. à l'aide d'une architecture de référence plutôt générique en sécurité informatique. À cet égard, il faut alors différencier pour chaque « élément de construction » quelles fonctions sont interdépendantes et de quelles manières, quels éléments sont nécessaires ou facultatifs pour quel secteur, etc. Un accent particulier est mis sur la compatibilité et la cohérence (par ex.

lors de l'utilisation d'une identité numérique uniforme à tous les degrés de formation) des systèmes mis en œuvre et de leurs stocks de données.

- La mise à disposition d'environnements informatiques à usage collectif pour l'enseignement et la recherche, qui satisfont les exigences de sécurité des données et de la protection des données définis ci-avant et qui peuvent être utilisés sans examen approfondi au cas par cas (l'activité « science DMZ » déjà citée, qui met à disposition des accès sécurisés aux ordinateurs de haute puissance pour des projets de recherche, en est un exemple récent).
- Évaluations de la technologie en termes de fonctionnalité, de pertinence, d'intégration, etc. (la technologie « *Block-chain* » peut-elle être par ex. utilisée pour associer sans possibilité de falsification tous les diplômes et certifications obtenus au cours de la formation initiale et continue à l'identité numérique d'un apprenant ou d'une apprenante ?).

4.2.3.3 Aspects juridiques et réglementaires

En raison de l'importance croissante d'informations dans tous les domaines de la vie et de la numérisation en rapide évolution dans la société, la formation et l'économie, l'attractivité et le potentiel d'abus de données augmentent également. La situation de la menace qui évolue rapidement (cf. explications aux chapitres précédents) le confirme de manière flagrante.

Par conséquent, les législateurs de tous pays se penchent également sur ces phénomènes concomitants négatifs et la densité de la réglementation augmente continuellement. Cette tendance devrait se poursuivre de façon ininterrompue. Elle livre des modèles successivement plus strictes pour la sécurité de l'information.

Avec le Règlement sur la protection des données (RGPD) récemment introduit et contraignant au sein de l'Union Européenne (UE) qui exerce une action sur les territoires autres que celui de la Suisse, la protection des données des personnes physiques dans l'espace européen est un exemple d'une telle évolution. Aux degrés primaire et secondaire, le RGPD ne devrait avoir qu'une importance mineure, car les écoles desservent une zone géographique limitée et non transfrontalière et les élèves sont en principe domiciliés en Suisse. Il convient d'observer que les organisations des degrés primaire et secondaire sont en possession et traitent de grands stocks de données personnelles de mineurs, particulièrement dignes de protection. Dans les régions frontalières, les enseignants dont le domicile se trouve dans un pays limitrophe et qui travaillent en Suisse (frontaliers) peuvent être pertinents. Le RGPD est toutefois très important pour le degré tertiaire, qui doit être ouvert internationalement si la Suisse veut conserver sa place de leader dans les domaines de l'enseignement et de la recherche. Cela est valable aussi bien pour l'enseignement que pour la recherche. La révision en cours de la loi fédérale sur la protection des données (LPD), qui sera adaptée au RGPD, doit être prise en compte à tous les degrés. Par ailleurs, il faut continuer de prendre en compte en Suisse les variations cantonales en matière de protection des données et les interprétations des autorités cantonales chargées de la protection des données.

Dans l'enseignement à tous les degrés, la numérisation conduit à des défis spécifiques supplémentaires. Les examens effectués par le biais de plateformes électroniques et leur évaluation, ou des travaux rédigés, soumis et évalués par voie électronique, posent des exigences élevées en matière de disponibilité des systèmes durant les périodes d'examen, de confidentialité des informations stockées, de protection contre une manipulation ultérieure des travaux et leur évaluation, ainsi qu'en matière de traçabilité des activités et des modifications en conséquence. Si ces exigences ne sont pas satisfaites, alors il faut s'attendre à sur une plus grande proportion d'oppositions contre la légalité des examens, l'évaluation de ces derniers et des travaux et donc contre les notes et les décisions de promotion. Dans ce contexte, il est également nécessaire de réglementer les possibilités de consulter, de faire opposition et recours par rapport aux informations stockées, mais aussi les résultats des processus algorithmiques et les prévisions ou autres mesures organisationnelles.

Par ailleurs, au degré tertiaire, il faut prendre en compte d'autres réglementations spécifiques selon le domaine de spécialisation. Certaines réglementations, par ex. le contrôle des exportations ou l'utilisation de marchandises dangereuses, sont applicables dans divers domaines spécialisés, alors que d'autres réglementations ne concernent que d'autres domaines spécifiques, comme par ex. dans un contexte médical ou

pharmaceutique, dans le domaine de la chimie ou dans celui du nucléaire. On a souvent créé ces réglementations pour des biens et processus physiques ; toutefois, elles sont aujourd'hui toujours plus applicables aux informations et au traitement de l'information et sont donc pertinentes pour la sécurité de l'information et la sécurité informatique.

En règle générale, quel que soit le degré de formation, il faut veiller à ce que les personnes impliquées ou leurs représentants légaux aient la possibilité de consulter, de faire opposition et recours en ce qui concerne les informations stockées et traitées. Par analogie, cela est également valable pour les résultats dynamiques de processus algorithmiques, de prévision, etc. car, sinon, des effets non autorisés ou discriminants sont possibles. Toutefois, il faut analyser plus en détail comment le contrôle institutionnel de la sécurité de l'information et des algorithmes assurant le traitement des données doit être mis en place, notamment si ces algorithmes ne sont pas librement disponibles et testables, mais représentent une propriété intellectuelle protégée du fabricant ou de l'exploitant, ou si les algorithmes ou leurs paramètres de contrôle se modifient automatiquement dans les applications d'autoapprentissage et d'autoconfiguration et dynamiquement sans possibilité d'intervention et de contrôle manuel. Une méthodologie ainsi qu'une mise en œuvre centrales seraient souhaitables ; il faut toutefois s'attendre à ce que toutes les réglementations ne soient pas suffisamment univoques dans ce domaine pour pouvoir être mises en œuvre directement et sans conflit d'objectif – parallèlement à la réglementation prospective de ces questions, on s'attendra donc à ce que la sécurité juridique nécessaire se constituera seulement par le biais d'un jugement judiciaire fondé sur des litiges concrets.

4.2.3.4 Aspects économiques

Les aspects économiques de la sécurité de l'information et de la sécurité informatique ont déjà été abordés en introduction à ce chapitre et sous 4.2.3.1 Aspects conceptuels. Nous souhaitons relever brièvement les points suivants :

- Le développement de la situation de la menace, poussé par la numérisation et la professionnalisation de la criminalité sur Internet, conduit à des attaques toujours plus professionnelles et complexes, avec un potentiel de dommages croissant rapidement. La sécurité informatique doit donc développer continuellement des mesures supplémentaires, ce qui conduit dans ce domaine à une charge de travail conceptuelle et opérationnelle à la hausse.
- À l'heure où les budgets alloués à l'enseignement et à la recherche stagnent, et, parallèlement, où les besoins d'investissement dans l'informatique et la sécurité informatique sont fortement à la hausse du fait de la numérisation, et où les frais d'exploitation augmentent en conséquence, il faut réaliser des gains d'efficacité.
- Comme nous l'avons montré, il existe beaucoup de potentiel de collaboration, notamment dans la sécurité informatique, qui permet de réaliser des économies de frais et, par ailleurs également, une amélioration de la qualité des solutions définies conjointement et partiellement réalisées.
- Les spécialistes de la sécurité informatique constituent, aujourd'hui encore, une denrée rare sur le marché du travail. De ce fait, pour des raisons financières mais aussi qualitatives, il est tout à fait judicieux d'examiner dans quels thèmes il faut privilégier les centres de compétence par rapport aux actions individuelles non coordonnées.

Actuellement, il n'est pas possible de se prononcer sur les besoins financiers concrets en qualité et en quantité suffisantes – il sera donc nécessaire de prélever, lors d'une phase de travail à venir, suffisamment de données sur les coûts d'exploitation récurrents et sur les coûts d'investissement nécessaires à l'aide de projets pilotes représentatifs et pertinents. En raison des caractéristiques très différentes et de la progression non linéaire attendue de l'utilisation des synergies entre les organisations éducatives très diverses, il ne suffira cependant pas d'appliquer des modèles d'extrapolation simplistes – sans pilotage et sans prélèvement empirique des données issues de la pratique de mise en œuvre, il n'est donc guère possible, à l'heure actuelle, de faire des déclarations concernant le besoin financier. Toutefois, une « plateforme nationale de la formation » au sens décrit ci-dessus entraînera dans tous les cas une réduction des coûts, pour autant que le financement du concept et un fonctionnement sûr, performant et stable puissent être garantis à moyen terme.

4.2.4 Remarques finales

Les considérations exposées dans cette étude sur la sécurité appropriée de l'information dans le « paysage suisse de l'enseignement et de la recherche » se résument de la manière suivante :

1. Une solution appropriée et flexible face aux défis à venir pour l'utilisation professionnelle des données tout en garantissant une sécurité de l'information et une protection des données suffisantes est nécessaire, mais aussi possible – sans pour autant être gratuite.
2. L'orientation fédérale fondamentale des différents groupes d'intérêt subsiste et il convient d'en tenir suffisamment compte – les concepts, les outils, etc. peuvent bénéficier d'effets globaux de synergie et d'externalisation pour limiter les dépenses initiales et récurrentes, la mise en œuvre et la « Governance » restent toutefois largement du ressort des organisations individuelles.
3. Un modèle de protection échelonné selon les besoins des groupes d'intérêt respectifs paraît être plus judicieux et plus simple à mettre en œuvre qu'une solution centrale globale – ce modèle échelonné nécessite toutefois une participation active et un soutien à la mise en œuvre de la part des organisations individuelles concernées.
4. Afin d'obtenir en quantité suffisante des données qualitatives et quantitatives sur la structure exacte et sur les charges d'exploitation, des clarifications supplémentaires sont nécessaires, pouvant aller jusqu'à un pilotage représentatif d'éléments de solution dans le cadre d'une phase de projet ultérieure structurée, financée et coordonnée en conséquence.
5. Au vu de la complexité croissante des moyens informatiques et de communication mis en œuvre et à venir, ainsi que de la situation de la menace existante et constamment à la hausse, les étapes esquissées qui suivent doivent être abordées rapidement.

4.3 Open Data

Andreas Kellerhals, ancien responsable OGD

Table des matières

4.3.1	Données et données ouvertes	84
4.3.1.1	Définition	84
4.3.1.2	Stratégies suisses	86
4.3.1.3	Stratégie spécifique aux données ouvertes dans le domaine de la formation ?	87
4.3.1.4	Portails	87
4.3.2	Données de formation ouvertes : utilisation, disponibilité, potentiels	89
4.3.2.1	Rapports sur l'éducation : données utilisées	89
4.3.2.2	Données de formation disponibles	89
4.3.2.3	Potentiel : coordination de l'offre et de la demande de données	90
4.3.3	Conclusions et perspectives	92
4.3.4	Bibliographie	92

4.3.1 Données et données ouvertes

Actuellement, les données sont souvent décrites au choix comme l'« or » ou l'« huile » de la société de l'information. Il vaut toutefois mieux de les considérer comme une « infrastructure ». Les données peuvent être utilisées sous une forme non rivale comme contribution aux objectifs les plus variés et aux intentions les plus diverses (OCDE 2015:177-206). Dans le domaine de l'éducation, la stratégie de numérisation de la CDIP (2018) vise à « exploiter ouvertement les possibilités d'utilisation des données », tant pour organiser administrativement et politiquement l'instruction publique que pour la vie scolaire quotidienne, pour observer le développement de l'instruction publique, pour saisir la réalisation des objectifs et pour inciter aux innovations pédagogiques. Au sens de l'OCDE, les données sont donc comprises comme une infrastructure à multi-usage. Elles alimentent en même temps un cycle de décisions (conception de la politique de formation) et d'analyses (contrôle de l'efficacité, « concrétisation de la reconstitution » Mastronardi 1991:9), qui, de leur côté, serviront de base à d'autres décisions. Ce cycle de décisions et de contrôles devrait être ouvert à la participation démocratique, ce qui signifie que les bases de la planification, de la conception et du contrôle de l'efficacité doivent être également ouvertes dans le cadre de ce qui est légalement possible.

Actuellement, les données ne sont pas encore obligatoirement des données ouvertes : comment se comportent les données par rapport aux données ouvertes ? Comment les données ouvertes peuvent-elles contribuer à la politique de formation ? Qui est intéressé à acquérir des données et à quelles fins ? Dans ce qui suit, on mentionne tout d'abord ce que sont les données ouvertes (définition), dans quel contexte ce concept est mis en œuvre (stratégies) et quels sont les résultats principaux (portails). Dans une seconde partie, on traite les données disponibles et utilisées dans le domaine de l'éducation – examinées plus en détail dans l'exemple des rapports sur l'éducation 2014 et 2018 – et la question d'éventuelles données lacunaires. Les conclusions font une brève synthèse de l'ensemble de la matière abordée dans ce chapitre.

4.3.1.1 Définition

Les données ouvertes sont des « données qui peuvent être librement utilisées, réutilisées et redistribuées par quiconque » (opendatahandbook.org, 12.2018).¹⁷⁶ Cette définition est également valable pour les données ouvertes spécifiques, telles que les données ouvertes d'administration, les données ouvertes de recherche¹⁷⁷, mais aussi les données ouvertes professionnelles. Les limites de l'ouverture sont fixées par des intérêts de protection légitimes et prépondérants (par ex. protection des données, protection de l'information, protection du secret d'affaire et du secret professionnel, éventuellement aussi les droits d'auteurs), ce qui présuppose une pesée minutieuse des intérêts.¹⁷⁸

Cette définition des données ouvertes semble apparemment claire, mais, en y regardant de plus près, elle nécessite des explications sur plusieurs points. En premier lieu, on n'est jamais tout à fait au clair sur ce que sont en fait les données. Dans la hiérarchie du savoir, les données se situent entre les sigles et les informations. Des informations peuvent être obtenues des données et, à partir de là, du savoir peut à son tour être généré. Afin d'explicitier cette considération, faisons appel à une paraphrase : les données ne sont pas seulement des données statistiques, financières ou de mesures (par ex. données météorologiques). Les informations géographiques ainsi que les textes (listes, textes juridiques, procès-verbaux, etc.) sont également considérés comme données – en résumé : tout ce qui peut être traité comme données est considéré comme données (Conseil fédéral 2018:2).¹⁷⁹ Actuellement, une définition aussi large est notamment indiquée, car, grâce aux applications de l'intelligence artificielle, de toutes nouvelles possibilités de préparation et d'analyse de données structurées ou non structurées sont à disposition. Ainsi, si l'on parle de données ouvertes, il faut le comprendre dans un sens large ; toutefois, la discussion tourne toujours autour des don-

176 Jusqu'ici, l'UE a plutôt utilisé la notion de « *Public Sector Information* ».

177 [Déclaration de principe du Fonds national suisse sur le libre accès aux données de la recherche \(« Open Research Data »\)](#) ; il ne faut pas confondre les données ouvertes de la recherche avec l'« Open Access », le libre accès aux publications scientifiques. Sur le plan fédéral, le Département fédéral de l'économie, de la formation et de la recherche (DEFR), ainsi que le Secrétariat d'État à la formation, à la recherche et à l'innovation (SEFRI) sont chargés d'examiner, en collaboration avec les acteurs de la recherche et des hautes écoles, les mesures appropriées pour l'ouverture des données de recherche, afin de réaliser une politique des données satisfaisant les exigences actuelles.

178 Par exemple, dans le domaine de la formation, cela se manifeste par la discussion sur la conformité d'une publication de notes avec les règles en matière de protection des données (cf. par ex. Rudin 2019).

179 Dans la première stratégie de libre accès aux données publiques (« Open Government Data », OGD), le Conseil fédéral a défini les données de manière relativement étroite et uniquement dans une note de bas de page (Conseil fédéral 2014:3596), ce qui a conduit à des ambiguïtés constamment récurrentes (cf. BAR 2017:11).

nées déjà existantes – en d'autres mots, l'appel à des données ouvertes est le souhait d'un libre accès aux données existantes, et non une exigence de prélever ou de saisir de nouvelles données. Le fait qu'il faille également collecter de nouvelles données dans certaines circonstances n'est naturellement nullement remis en question. L'exigence d'un accès ouvert à de telles données nouvellement saisies ne l'est pas non plus¹⁸⁰ pour des questions éducatives et de formation ; l'utilité des données résulte uniquement des interrogations et des décisions méthodologiques. C'est précisément quand on conçoit l'éducation au sens large que de nombreux éléments se transforment en données de formation potentielles. Si toutes ces données sont ouvertes, alors elles peuvent être librement utilisées dans le contexte éducatif.

Sont ouvertes les données qui peuvent être librement utilisées aussi bien d'un point de vue juridique, financier que technique (opendatahandbook.org, okfn.org/opendata). Des restrictions juridiques sur l'utilisation des données (en raison par ex. de la protection des données) sont nécessaires, mais lorsqu'aucune raison de ce type ne va à l'encontre d'une utilisation ultérieure ouverte, une restriction de l'utilisation ultérieure n'est par exemple pas autorisée (uniquement pour les établissements scolaires). Les données ouvertes devraient être utiles sous licences libres (par ex. « Creative Commons ») pour tous et à toutes fins. Toutefois, l'utilisation ultérieure peut être soumise à certaines conditions (par ex. obligation d'indiquer les noms, transfert dans les mêmes conditions). L'utilisation doit être possible gratuitement.¹⁸¹ Par ailleurs, les données doivent exister dans un format adéquat, de sorte que le traitement ultérieur et la connexion avec d'autres données soient aisés (lecture automatisée, formats ouverts¹⁸², interopérabilité sémantique). De plus, la disponibilité de ces données devrait être assurée à long terme (vérifiabilité, traçabilité). En outre, les données ouvertes devraient être complètes, actuelles et crédibles. Elles doivent être décrites de telle sorte que l'on reconnaisse clairement ce qu'elles représentent (objet, moment ou période de la saisie, couverture géographique, etc.) et comment elles ont été créées (méthode de relevé ou de la mesure, sources). Il s'agit là d'une condition méthodologique pour chaque traitement ultérieur critique des données publiées et leur liaison avec d'autres données.¹⁸³ C'est uniquement dans ces conditions que les données ouvertes permettent de trouver des réponses aux questions, de prendre des décisions – pour soi-même, pour la communauté, pour une entreprise ou une organisation – ou de mieux utiliser les offres existantes de prestations de services.

La première définition des données ouvertes (« Open Data ») de l'Open Knowledge Foundation date de 2006. Le premier portail de libre accès aux données (publiques) a été mis en service en 2009 déjà par le gouvernement américain.¹⁸⁴ Depuis lors, ce concept a rapidement été largement approuvé par les gouvernements nationaux et les organisations internationales (G8, OCDE, Banque mondiale). Au cours de ces quelques années, de nombreuses stratégies ont été adoptées, fondamentalement similaires entre elles : pratiquement tout le monde s'accorde à dire que les objectifs principaux ou la motivation la plus importante d'une politique d'ouverture des données (gouvernementales) sont l'amélioration de la transparence, l'augmentation de la participation (gouvernement ouvert), l'encouragement de l'innovation ainsi que l'efficacité et l'efficacité.¹⁸⁵

180 Dans le cadre de la discussion de ce rapport, il a été fait référence, par exemple, à la collection audio en suisse-allemand du domaine public pour enfants, y c. la transcription pour les enseignants, grâce à laquelle des assistants linguistiques et des outils de traduction permettent aux enfants de comprendre le suisse-allemand et d'apprendre à le parler. Il s'agit là d'un exemple, d'autres pourraient s'y ajouter. De son côté, le Fonds national suisse opte pour des données de recherche ouvertes, ce qui est très utile ici (cf. note 177).

181 La gratuité de l'utilisation ne s'applique pas sans restriction, aussi bien opendatahandbook.org que okfn.org/opendata relèvent que « rien de plus que des frais raisonnables de reproduction » peuvent être réclamés. La première stratégie OGD se limite aux cas où les données peuvent être publiées « sans charge supplémentaire spécifique » (Conseil fédéral 2014:3499).

182 Le modèle à cinq étoiles définit les niveaux de qualité des publications de données à partir des fichiers PDF, disponibles sur le net : une étoile, via les formats de données propriétaires et ouverts, deux ou trois étoiles, jusqu'au format RDF (« Resource Description Format ») ou pour les données ouvertes liées, quatre et cinq étoiles (données ouvertes 5 étoiles).

183 À propos de l'histoire : selon Wikipédia, le concept des données ouvertes remonte à l'Année géophysique internationale 1957/1958 et a poursuivi l'objectif « de faciliter l'échange et l'utilisation de données scientifiques par la mise en place de centres d'échange de données et la standardisation de métadonnées », la notion est apparue pour la première fois en 1995, à nouveau dans le contexte de la géophysique et des sciences environnementales (Chignard 2013). Une première ébauche de définition date de 2005 (Définition du Savoir Libre). Sur le plan politique, cette évolution a été favorisée de façon décisive par la directive programmatique gouvernementale ouverte (2009) (« *Open Government Directive* ») de Barack Obama, ou son partenariat gouvernemental ouvert (2011) (« *Open Government Partnership* »), suivis peu après par la Charte du G8 pour l'ouverture des données gouvernementales au public (2013) (« *G8 Open Government Data Charter* »). L'émergence du thème peut être retracée par la suite (Kellerhals 2018).

184 Le premier portail était data.gov, data.gov.uk a suivi au printemps 2010 en Angleterre.

185 Le Fonds national suisse souligne le caractère vérifiable des résultats de la recherche (FNS 2017), les CFF encouragent quant à eux l'innovation (Trachsel 2018).

4.3.1.2 Stratégies suisses

Dans ce paysage stratégique diversifié, la Suisse s'intègre avec entre-temps deux stratégies OGD nationales (Conseil fédéral 2014, 2018). La spécificité de la Suisse consiste à accorder une relativement faible importance aux objectifs de politique étatique, ce qui surprend peu « compte tenu du fait que la transparence et la participation ont déjà atteint un niveau satisfaisant » (Neuroni, Riedl & Brugger 2013:1912). L'innovation et la croissance économique sont sur la liste des objectifs de la première stratégie OGD, situées avant la transparence et la participation (Conseil fédéral 2014). Dans la stratégie 2019-2023 (Conseil fédéral 2018:2), l'aspect de la participation est plus fortement pondéré, conformément à la compréhension internationale (Conseil fédéral 2018) ; malgré tous les droits démocratiques, il ne faut pas non plus sous-estimer cet aspect en Suisse (Kellerhals 2018).

Selon la structure fédérale, les cantons et les (plus grandes) communes disposent également de propres stratégies partielles (par ex. Genève, Thurgovie ; la Ville de Zurich) ou du moins des décisions gouvernementales (par ex. le canton de Zurich), qui définissent le cadre de l'ouverture des données. À cet égard, certains cantons individuels intègrent leurs activités dans la stratégie nationale de la Confédération¹⁸⁶, alors que d'autres conservent plus d'autonomie.¹⁸⁷ À leur tour, d'autres institutions travaillent avec des déclarations de principe (FNS) ou publient leurs données conformément à une stratégie spécifique propre au groupe (par ex. CFF¹⁸⁸).

Le mandat général d'information de la part des autorités (Constitution fédérale, LOGA 1997), qui autorise par principe la publication, mais ne l'impose pas (Wiedmer, Seibert 2015), constitue le fondement juridique d'une politique d'ouverture en matière de données (de l'administration) dans le domaine public (Confédération). Selon l'interprétation, la loi fédérale sur le principe de la transparence dans l'administration (LTrans 2004) peut également être comprise comme un mandat d'information active (comme étant une obligation) (Nuspliger 2006) – la discussion qui tourne autour des bases légales, y compris la question de savoir si l'on a besoin d'une loi OGD spécifique, ne devrait pas encore être achevée.¹⁸⁹ En outre, dans le domaine de la formation, il convient encore de clarifier la question concernant la manière dont il faut gérer les données individuelles de formation (dossier éducatif électronique analogue au dossier électronique du patient ?). Il existe un double intérêt (public et privé) pour de nombreuses données cruciales pour le fonctionnement du système éducatif. Similairement au domaine des archives, on peut constater que rien ne s'oppose juridiquement à une protection à long terme de telles données à caractère personnel dans les institutions publiques, pour autant qu'elles aient été collectées légalement.¹⁹⁰ Par conséquent, leur disponibilité pour des analyses est théoriquement remplie ; dans la pratique toutefois, les données de la vie scolaire quotidienne (au niveau communal) sont rarement susceptibles d'être collectées de façon centralisée et d'être par la suite archivées. Afin de modifier cette situation, on devrait chercher la discussion à l'aide des services compétents.

186 De la même manière, le canton de Thurgovie poursuit une propre politique OGD, l'intégrant toutefois clairement dans la stratégie nationale : « Les OGD du canton de Thurgovie sont disponibles via la plateforme OGD de la Confédération (opendata.swiss). Certains liens directs sont également disponibles sur les sites Internet des offices cantonaux respectifs, qui redirigent vers la plateforme nationale ou vers les [OGD thurgoviennes](#). »

187 Alternativement, le canton de Genève met par ex. à disposition sa [Stratégie en matière de libre accès aux données publiques de l'administration genevoise \(Open Data\)](#).

188 Stratégie du groupe adoptée le 15.12.2015 (Trachsel 2018).

189 La stratégie pour des données administratives ouvertes 2019-2023 prévoit jusqu'à mi-2020 un examen des bases légales, afin de « créer des conditions-cadres juridiques optimales » (Conseil fédéral 2018:3, 6).

190 Cet examen s'applique au niveau fédéral lui-même pour conserver les données personnelles particulièrement sensibles. L'affaire des fiches permet de tirer la leçon suivante : même les données prélevées illicitement doivent être archivées, car elles ont servi de base à des décisions qui, autrement, ne peuvent plus être comprises ; il faut organiser l'accès à de telles données de façon particulièrement sensible.

Il faudrait clarifier la question de savoir si et, le cas échéant, quelles données de l'instruction publique font partie des données personnelles particulièrement sensibles. Au niveau fédéral, cette catégorie de données personnelles est énumérée de façon exhaustive dans la loi fédérale sur la protection des données (LPD 1992). Sur le plan cantonal, de telles énumérations sont considérées comme exemplaires et non exhaustives (par ex. <http://dsb.bs.ch/datenschutz/was-sind-personendaten.html>). Vous trouverez des solutions pour garantir l'autodétermination informationnelle selon une perspective juridique chez Weber, Thouvenin (2017) ou SIR (2018).

4.3.1.3 Stratégie spécifique aux données ouvertes dans le domaine de la formation ?

La question à discuter ouvertement est de savoir de combien d'autres stratégies a-t-on actuellement encore besoin. Certes, la seconde stratégie OGD nationale est uniquement contraignante pour l'administration fédérale centrale, mais elle se comprend comme une invitation et offre de soutien aux cantons et communes (Conseil fédéral 2018). Dans le cadre de sa politique de propriétaire, la Confédération souhaite également faire progresser la publication de données d'entreprises proches de l'État.¹⁹¹

La stratégie OGD nationale constitue un cadre conceptuel suffisant, bien que l'instruction publique soit généralement soumise à une compétence cantonale et connaisse sa réalisation au niveau communal avec la participation de différents acteurs privés et proches de l'État. L'obligation de publier des données administratives devrait être également statuée dans les territoires cantonaux individuels. En ce qui concerne les acteurs proches de l'État (par exemple les maisons d'édition spécialisées dans le matériel pédagogique ou la centrale intercantonale de matériel pédagogique), les cantons peuvent également obtenir les publications nécessaires de données dans le cadre de leur politique de propriétaire. Par ailleurs, les cantons et les communes pourraient également imposer aux personnes physiques de droit public ou privé qui effectuent des tâches d'exécution déléguées, la publication des données dans le cadre de leur mandat.

À première vue, il n'est donc pas possible de discerner la valeur ajoutée d'une stratégie spécifique de données de formation ouvertes. Il paraît plus opportun de viser tout d'abord à des clarifications pragmatiques et de définir des lignes directrices pour le traitement des données et d'assurer les conditions infrastructurelles nécessaires pour un apport fiable des données de haute qualité.¹⁹² À cet égard, ces critères exigent une volonté claire d'ouverture des données, une organisation opérationnelle et une définition des compétences en matière de publication des données ouvertes, y compris la mise à disposition des ressources nécessaires. Les chances de succès sont faibles sans ressources pour la préparation soignée des données, le soin et l'élargissement continu de l'offre en données de formation, ainsi que le soutien de parties intéressées avec des informations supplémentaires et des assistances (réseautage, organisation de hackathon ou d'autres événements, etc.). Il peut même s'avérer judicieux de proposer une infrastructure technique (par ex. pour l'hébergement de données pour les petites entités produisant des données). Des gains d'efficacité sous forme d'économies peuvent être attendus uniquement suite à un investissement initial (cf. le problème du manque de ressources dans la mise en œuvre de la stratégie OGD de la Confédération CDF 2018:4, 23).

4.3.1.4 Portails

Une plateforme de données constitue généralement l'élément principal de tous les efforts de publication de données ouvertes. Aujourd'hui, il existe plus de 2500 portails de ce type dans le monde, dont 16 rien qu'en Suisse ([Open Data Soft](#)¹⁹³), qui remplissent tous une fonction d'orientation pure et simple et qui facilitent l'accès central aux données ouvertes sans centraliser les données elles-mêmes. Les plateformes sectorielles (par ex. [geo.admin.ch](#), [opentransportdata.swiss](#)) ou les plateformes institutionnelles (par ex. [data.stadt-zuerich.ch](#)) agissent de façon complémentaire. L'accès aux données ouvertes est toujours possible par divers biais (plateforme centrale, portails professionnels spécifiques ainsi que directement via les offres Internet des propriétaires des données). Cette redondance n'est pas négative en soi, elle accroît la chance que des profanes trouvent un accès aisé à des données de haute qualité sur le thème de la formation. Toutefois, la question se pose de savoir si le développement de portails supplémentaires (spécialisés sur le plan thématique) et, de ce fait, une hausse supplémentaire de la redondance, en vaut la peine en termes de dépenses et de coûts. En premier lieu, nous considérons qu'il convient d'élargir les portails existants adéquats qui présentent peu d'investissements avec des fonctionnalités supplémentaires (par ex. des fonctions de recherche et de tri), au lieu de réaliser des portails entièrement nouveaux avec des fonctionnalités largement redondantes.

¹⁹¹ En 2014 déjà, l'OFT a mandaté les CFF pour concevoir une propre plateforme pour les données de l'ensemble des transports publics et la publication de données personnelles. Les données d'[opentransportdata.swiss](#) sont également disponibles via [opendata.swiss](#).

¹⁹² En ce qui concerne le besoin de directives, cf. la stratégie de numérisation de la CDIP (2018), point 1, utilisation de données dans l'instruction publique.

¹⁹³ Les 16 portails nationaux sont ceux d'organisations suisses ou établies en Suisse.

Le nombre de données publiées par portail est divers, le potentiel de données publiables est toutefois encore loin d'être épuisé. Les portails ne lient pas simplement les prestataires aux utilisateurs de données (au sens d'une chaîne linéaire de création de valeur), mais sont des nœuds importants dans les réseaux, dans lesquels les prestataires de données sont également des utilisateurs et réciproquement. Dans le dialogue entre les acteurs en réseau, il faut clarifier quelles données il convient de publier à l'avenir ; ce dialogue doit aider à concilier le plus judicieusement possible l'offre et la demande.¹⁹⁴

Sélection de portails OGD ¹⁹⁵	Jeux de données	concernant la formation ¹⁹⁶		concernant l'école ¹⁹⁷	
		N	%	N	%
CH, opendata.swiss	6'568	490	7.5	365	5.6
USA, data.gov	301'207	428	0.1	-	-
UK, data.gov.uk	46'987	1400	3	857	1.8
D, govdata.de	21'697	2161	10	999	4.6
A, data.gov.at	20'743	131	0.6	123	0.6
F, data.gouv.fr	37'302	655	1.8	293	0.8
I, dati, gov.it	23'297	761	3.3	375	1.6
IE, data.gov.ie	8799	178	2	157	1.8
ES, datos.gob.es	21268	1'766	8.3	112	0.5
CY, data.gov.cy	2666	66	2.5	0	0
S, oppnadata.se	1132	202	17.8	12	1.1
SF, avoindata.fi	1321	37	2.8	97	7.3
Korea, data.go.kr	28415	1'402	4.9	698	2.5
CAN, open.canada.ca	81426	231	0.3	490	0.6

Tableau 1 : données administratives publiées et ouvertes en général et sur le thème de la formation / de l'école en particulier

Les données sur ces plateformes sont presque exclusivement publiées par des administrations publiques. Pour le traitement de questions de politique de la formation, les données de formation sont toujours à disposition sur de tels portails en relation avec les données de contexte. Il serait attrayant et possible de rendre accessible sur opendata.swiss des données issues d'autres acteurs provenant de tous les échelons fédéraux, y compris du domaine para-étatique ou même de particuliers. Vu sous cet angle, un portail spécifique aux données de formation ne semble apporter aucune réelle valeur ajoutée par rapport aux possibilités spécifiques d'accès aux données et à une description standardisée et harmonisée de ces données, mais pourrait plutôt produire un effet de caisse de résonance.¹⁹⁸ La personnalisation (« *Customizing* ») correspondante d'une offre de données de formation via les portails existants constitue déjà un défi et occasionne une charge de travail à ne pas sous-estimer.

194 En ce qui concerne la publication, la stratégie 2019-2023 précise que les services administratifs de la Confédération publieront en principe leurs nouvelles données sous la forme d'OGD (ouvert par défaut, « Open by Default ») à partir de 2020 ; pour des raisons de ressources et pour les données existantes qui ne sont toutefois pas encore publiées en libre accès, il convient de procéder par ordre de priorité selon les domaines politiques (les domaines de l'environnement, de la mobilité et de la santé y sont mentionnés). À ce stade, par exemple, il serait nécessaire d'insister lors d'une table ronde pour que l'on accorde la priorité aux données de formation (Conseil fédéral 2018:6).

195 Justification du choix des pays : l'Irlande, l'Espagne, la France, l'Italie et Chypre sont les précurseurs actuels conformément au rapport de l'UE, l'Angleterre et les États-Unis sont les pays historiques ; par ailleurs, les autres pays limitrophes de la Suisse ont encore été sélectionnés (Commission européenne 2018). La Suisse fait partie de ce que l'on appelle les suiveurs (« Followers ») (rang 22 sur 31, dans la troisième catégorie sur un total de quatre). Les classements d'opendatabarometer.org (CAN, UK, Australie, F, Corée du Sud), index.okfn.org (Taiwan, Australie, UK, SF) ou de l'OCDE arrivent à des résultats complètement différents, raison pour laquelle la Suède, la Finlande, le Canada, la Corée et Taïwan ont encore été pris en compte dans le tableau.

196 Pour la Suisse, la catégorie correspondante est « Éducation, science » ; pour les autres pays, elle varie d'« Éducation (Higher Education, Postsecondary Education) » à « Éducation, culture, sports » en passant par « Bildung, Forschung » resp. « Éducation, recherche (éducation, recherche, scolarité, enseignement, lycée, école, université) » ou « Istruzione, cultura, sport ».

197 Comme complément, on a ajouté le 10.1.2019 la colonne « ensembles de données relatifs aux écoles » (« School », « Escuela », « Skola », « Koulu ») (recherche d'ensembles de données correspondants sur les portails individuels). Pour la Corée, la recherche a été réalisée en anglais. Le site des États-Unis était hors service en raison du « Shutdown ».

198 Le canton de Thurgovie montre l'exemple avec son [portail cantonal](#) mais encore sans aucune donnée de formation.

4.3.2 Données de formation ouvertes : utilisation, disponibilité, potentiels

Aussi saturés en matériel ou en données que soient par exemple les rapports sur l'éducation du CSRE ou de l'OCDE, la question de l'exhaustivité de la base des données continue toutefois à se poser. La relation entre les données et les interrogations et analyses est complexe. Les interrogations et les objectifs de connaissance définissent des mesures, des catégorisations, des classifications et celles-ci ont des répercussions sur la formation d'hypothèses et sur le plan d'enquête. Il n'y a pas ici de bonne ou mauvaise procédure ; il faut toutefois toujours réfléchir à ces interdépendances et les rendre transparentes.

4.3.2.1 Rapports sur l'éducation : données utilisées

Comme exemples d'une utilisation réelle de données dans le domaine de la formation, on a examiné plus en détail les rapports 2014 et 2018 :

Rapport de formation	2014	2018
Visualisation ¹⁹⁹	277	351
- basée sur des sources (bibliographie)	49	25
- basée sur des données	227	325
Provenance des données	26	34
Nombre des sources de données	244	344
Visualisations basées sur plus d'une source de données	20	45
Données liées à des calculs supplémentaires	74	170

Tableau 2 : utilisation des données dans les rapports 2014 et 2018

Les données utilisées proviennent de 26 ou 34 différentes sources de données, et pour les deux rapports regroupés, de 46 : la source la plus importante est l'OFS (63 ou 67 % des données utilisées), suivie par l'OCDE (10 ou 8 %) et, uniquement en 2018, par la CDIP (6 %) – les autres organisations produisant des données fournissent toutes moins de 5 % des données utilisées. Si l'on structure les sources de données selon la typologie du chapitre 2 (cf. graph. 4, p. 15), alors 14 (36 %) font partie de la catégorie « *Administration de l'éducation* », 9 (23 %) appartiennent aux catégories « *Consortiums établis* » et « *Recherche* », 5 (13 %) à la catégorie « *Agences spécialisées* » et 2 (5 %) à celle des « *Autres acteurs* ». Les prestataires de services et les médias ne sont pas (encore ?) représentés ici.²⁰⁰ Il faut vérifier les accès alternatifs et les approches méthodologiques, en particulier pour évaluer les données des médias.²⁰¹

4.3.2.2 Données de formation disponibles

Si nous déplaçons notre attention de l'utilisation des données à leur offre, nous constatons alors que l'OFS a publié 4054 tableaux ou 159 ensembles de données interactifs sur le thème de la formation et de la science. Sur la plateforme opendata.swiss, 491 ensembles de données sont répertoriés sur le même thème. Ceux-ci proviennent principalement d'offices fédéraux (430), notamment de l'Office fédéral de la statistique (407) et en proportions réduites de la Bibliothèque nationale suisse (12), de la Bibliothèque am Guisanplatz (4), de l'Office fédéral de l'énergie (3), des Archives fédérales (2), de la Commission indépendante d'experts internements administratifs (1) et du Musée national suisse (1). D'autres ensembles de données proviennent des cantons (Zurich et Bâle-Ville, 29), des villes (Berne, 4), ainsi que de tiers tels que l'« OpenGLAM Working Group » (14), de la bibliothèque de l'EPF de Zurich (11), de Swissbib (2) et du Fonds national suisse (1).

¹⁹⁹ Pour une visualisation dans chaque rapport, ni une source, ni une base de données ne sont indiquées.

Les indications sur toutes les visualisations sont systématiques et se différencient entre les sources (référence à la bibliographie) et les données (indication de la source des données), complétées dans certains cas par la référence à ses propres calculs. Les indications des sources des données restent générales, ce qui ne rend pas leur identification aisée auprès de l'OFS.

²⁰⁰ Plusieurs raisons expliquent les données lacunaires visibles ici : les données issues de la vie scolaire quotidienne ou des écoles individuelles ne sont pas systématiquement saisies de façon centrale ni conservées à long terme. Pour le moins dans le canton de Berne, aucune obligation de conserver ou d'archiver n'est valable au niveau des communes (scolaires) (cf. ODArch communes 2014, ODArch communes, annexe 2018). Il en va probablement de même pour les données dites d'application (cf. graph. 4), pour lesquelles, le cas échéant, une protection des intérêts commerciaux pourrait encore être invoquée.

²⁰¹ Pour la période 2013-2018, la banque de données des médias Swissdix référence près de 300 000 articles relatifs au critère de recherche « École », plus de 2500 articles relatifs au système scolaire, plus de 5500 articles ayant trait à l'instruction publique et près de 4000 articles concernant la politique de formation (les réponses multiples n'ont pas été déterminées). Sur le plan méthodologique, un accès comparable aux analyses du système politique par l'Année politique suisse semble digne d'être examiné ; celle-ci comprend déjà des thèmes de politique de formation (<https://anneepolitique.swiss/articles?chapters%5B%5D=1.8.1.>).

L'hétérogénéité des données accessibles est élevée.²⁰² À l'aide de l'évaluation des offres de données relatives aux critères de recherche *formation / éducation* ou *école*, il n'est que partiellement possible de déterminer ce qui est véritablement pertinent pour le domaine de la formation. Certains des champs thématiques ainsi saisis se focalisent également sur la culture et le sport, d'autres se concentrent sur la science, et d'autres encore prélèvent des données identiques relatives à différentes collectivités locales. Il faudrait analyser ces résultats de façon détaillée. Mais il est clair que les données scolaires et de formation ne sont pas encore largement représentées dans le contexte des données ouvertes. Mais, en même temps, il devient également clair qu'une quantité plus importante de données est disponible que ce que l'ont évalué les deux rapports sur l'éducation ; il n'est pas aisé de déterminer où et par qui ces données sont utilisées.

4.3.2.3 Potentiel : coordination de l'offre et de la demande de données

D'une part, malgré la multiplicité des données utilisées, des données lacunaires sont déplorées dans les deux rapports sur l'éducation, qui ont malheureusement empêché des « évaluations bien fondées ou des planifications de mesures » (CSRE 2014:6, 2018:6). D'autre part, le rapport révèle également que l'évaluation des données disponibles ne conduit pas simplement à la découverte de causalités logiquement contraignantes (par ex. le thème de la connaissance et du comportement qui influence l'environnement, CSRE 2018:317). Exiger d'autres données afin de combler les données lacunaires semble a priori compréhensible, mais soulève des questions lorsque seule une petite proportion des données disponibles est utilisée. Il convient d'éviter une course aux armements lors de la publication des données. Une discussion sur la politique de collecte des données s'impose en premier lieu, afin que la numérisation devienne une véritable occasion « d'enrichir considérablement le compte-rendu sur l'éducation au cours des années à venir » et pas seulement de combler les données lacunaires incriminées, mais de constituer la base de données nécessaire à un monitoring méticuleux et à des décisions ultérieures (CSRE 2018:7 et CDIP 2018, objectifs 5 et 7.1).

Les questions suivantes restent ouvertes : quelles données font véritablement défaut ? Les données font-elles défaut ou seul l'accès aux données est-il difficile ? Qu'en est-il des nombreuses données de formation qui ne sont pas du tout utilisées dans le monitoring de l'éducation ?

1. À quoi devrait ressembler un programme de collecte des données qui ne présente aucune donnée lacunaire ? Jusqu'en 2019, l'OFS s'est fixé l'objectif de « proposer de nouveaux produits qui permettront de mieux rendre compte du fonctionnement et de l'efficacité du système éducatif. Il s'agit en particulier d'effectuer des croisements entre les programmes de mesure des compétences [...] et l'observation des parcours de formation » (OFS 2016:22) et le programme annuel 2018 fait de ces « nouveaux produits pour le monitoring de l'éducation en Suisse » un objectif annuel avec une « observation globale et intégrée de l'ensemble des degrés de formation » (OFS 2018:3). Toutefois, s'il subsiste des lacunes, il faut impérativement discuter de la conception du programme de suivi.
2. Il convient d'examiner pour quelles raisons de nombreuses données ne sont pas utilisées pour le monitoring de l'éducation. Quels sont les avantages de ces données ? Pourrait-on éventuellement renoncer à des collectes individuelles ?
3. Si le problème réside dans l'accès aux données existantes, alors il conviendrait d'encourager une publication plus rapide des données correspondantes (tant que les organes fédéraux sont concernés). Dans le cas des cantons et des communes, les conditions préalables devraient d'abord être régulièrement créées à cet égard. La politique des données administratives ouvertes est absolument ouverte à de telles discussions, l'articulation de la demande est fondamentale pour un développement ultérieur du programme de publications axé de manière pragmatique sur les performances, par ex. dans le cadre du Forum Administrations publiques, de la Table ronde ou dans un groupe de travail thématique à mettre en place (Conseil fédéral 2018:5).

202 Deux exemples explicitent pour diversité de données : collection historique de photographies « Service actif Première Guerre mondiale », banque de données de recherche du FNS. Le problème des doubles catégorisations n'a pas pu être abordé systématiquement dans le cadre de ce mandat : parmi les données sur la population stockées sur opendata.swiss (645 ensembles de données), 16 sont par ailleurs liées à la catégorie « Formation, science » ; parmi les 109 ensembles de données de la catégorie « Finance », deux sont liés à « Formation » ou à « Formation et science ». Toujours est-il que de tels ensembles de données semblent plus directement liés aux questions de l'éducation et de la formation que si l'on entend par données de formation toutes les données « qui sont générées dans le cadre du système éducatif, qui permettent des déclarations à ce sujet ou d'en tirer des conclusions » (chap. 2).

4. Nous sommes confrontés à un problème plus fondamental si les données provenant d'autres acteurs font défaut dans le domaine de l'éducation (par ex. institutions éducatives, prestataires de services ; cf. Graphique 4). En premier lieu, il s'agit ici de créer la disposition et les conditions préalables à une publication des données.

Pour inciter à de nouvelles publications de données, il est important d'accorder une attention pragmatique au présent et à l'avenir. Les données nouvellement collectées doivent être destinées à être publiées dès le départ. Ainsi, le processus de collecte, de traitement et d'évaluation des données peut être axé dès le début sur une publication de données ouvertes. Cela a des conséquences positives sur la description des données, le choix des formats de données, etc. En revanche, la publication de données qui existent, mais qui ne sont pas encore librement disponibles, requiert une procédure différente et fastidieuse. La publication de données devrait se dérouler dans le cadre de projets clairement définis et adaptés à une demande concrète. Cela garantit que l'effort conduit véritablement à des résultats utilisés.²⁰³

Cette orientation de la publication des données vers une demande d'utilisation est importante, car l'utilisation des données administratives ouvertes, y compris dans le domaine de l'éducation, est, jusqu'ici, restée modeste en Suisse. Sur la quarantaine d'utilisations attestées sur opendata.swiss, l'une seule d'entre elles traite des aspects de la politique de formation (l'application [Eckdaten](#)) et encore ne le fait-elle sans s'appuyer sur des données de politique de formation au sens strict.²⁰⁴ Toutefois, on peut imaginer qu'il soit par ex. possible de développer des applications d'évaluation de « bonnes » écoles²⁰⁵ – de manière analogue aux « bons » hôpitaux ou aux « bonnes » maternités – et que ces applications soient utilisées. Cependant, cela soulève non seulement des questions claires sur le droit de la protection des données, qui exigent une pesée minutieuse des intérêts (Rudin 2019), mais également la question de savoir si de tels instruments de classement sont dans l'intérêt d'une politique démocratique de formation. Toujours est-il que de telles offres rendent des informations largement disponibles qui, autrement, seraient généralement uniquement accessibles à un cercle restreint de parties intéressées (à fort pouvoir d'achat).

Toutefois, le principal potentiel des données ouvertes dans le domaine de la formation ne semble pas uniquement résider dans une nouvelle publication de données pas encore accessibles, mais aussi dans l'ouverture de nouvelles possibilités de participation et de discussion. En théorie, les données ouvertes permettent à toutes les parties intéressées de se faire leur propre avis sur l'instruction publique et de le relier à des objectifs – à la fois concurrentiels. Les données ouvertes permettent de nouvelles approches créatives de résolution de problèmes²⁰⁶ et offrent également de nouvelles connaissances aux responsables de l'instruction publique (Kellerhals 2018). Un second potentiel important réside dans l'accroissement de l'efficacité administrative, car il « faut en outre partir du principe que le libre accès aux données publiques (' O[pen] G[overnment] D[ata] ') peut avoir un effet positif non négligeable sur les flux de données, par exemple dans les domaines de la recherche et de l'acquisition d'informations », ce qui est certainement le cas également pour les données non gouvernementales (« Non-Government Data ») (groupe d'experts 2018:137). Dans ce contexte administratif, rappelons ici aussi le principe « Once Only » (principe Une fois seulement) que l'UE a adopté dans la déclaration de Tallinn fin 2017 et auquel la Suisse a également adhéré. Ce principe vise à identifier les charges administratives redondantes dans les services publics grâce à la collaboration et à l'échange de données entre nos administrations aux niveaux national, régional et local ainsi qu'avec d'autres pays pour les services publics numériques transfrontaliers (« to identify redundant

203 Si les données existantes devaient être traitées de façon plus systématique, il y aurait par ex. des enseignements à tirer du projet FNS L'éducation en chiffres (« *Bildung in Zahlen* »). Mise en valeur historique de données statistiques sur l'éducation en Suisse (FNS, banque de données de recherche : <http://p3.snf.ch/Project-130398> ou <http://p3.snf.ch/Project-146239>).

204 Cette liste se base sur des entrées volontaires, c'est-à-dire, en d'autres termes : l'utilisation des données ouvertes est significativement plus importante que ce que le portail le démontre.

205 À cet égard, « bon » reste une notion très ouverte, qui peut signifier aussi bien des écoles présentant des taux élevés de succès lors du passage vers les écoles supérieures, que des écoles avec des offres spécifiques (enseignement musical, offres sportives spécifiques, cours de soutien, etc.), des écoles comportant une faible proportion d'étrangers ou celles accessibles aux personnes à mobilité réduite. À titre d'exemple, on se référera aux applications scolaires suivantes : [Hoe kies ik een school](#), [RomaScuola](#) ou The School Portal ; pour les applications relatives aux crèches, nous vous renvoyons à [KitaSuche](#) (qui sont toutes disponibles sur <https://www.europeandataportal.eu/en/using-data/use-cases>).

206 L'application Viadi d'Ubique constitue un exemple précoce et intéressant qui a significativement modifié le mode d'utilisation des horaires et qui a fourni le modèle des « horaires tactiles » polyvalents des CFF ; les idées venant du secteur privé ont donc permis une innovation manifeste. Dans un milieu plus proche de la politique, on suppose que, dans le cas des données ouvertes, la discussion sur la discrimination fiscale des couples mariés aurait pu avoir lieu immédiatement sur la base d'un vote plutôt que deux ans plus tard (Chancellerie fédérale 2016, Schöchli 2018a, b). En principe, la statistique (officielle) a revendiqué à un stade précoce une puissance transformative et un pouvoir explicatif général, qui sont en principe simplement démocratisés ici (Desrosières 2005).

administrative burden in public services by collaboration and data exchange across our administrations at national, regional and local level as well as with other countries for cross-border digital public services ») (Conseil européen 2017:4).

4.3.3 Conclusions et perspectives

Gérer la politique, l'administration et le monitoring de l'éducation et de la formation en se basant sur les données est judicieux. Cela ne nécessite pas une stratégie spécifique à cet égard – les fondements des données ouvertes (« *Open Data* ») sont suffisamment connus et reconnus – mais plutôt des mesures concrètes permettant l'application de ce principe chez tous les acteurs concernés (des institutions publiques au secteur privé).

Le succès d'une politique des données de formation ouvertes dépend du libre accès d'un grand nombre de données (ou du moins celles qui ont été demandées). Afin qu'elles soient utilisées, elles doivent être accessibles dans des formats lisibles par machine et avec suffisamment de descriptions des données (métadonnées), afin d'assurer une connectivité tant technique que sémantique. Des efforts sont donc nécessaires afin de préparer les données disponibles. De manière pragmatique, cela devrait s'appliquer en premier lieu aux données collectées et saisies à partir de maintenant ; ce qui permet d'orienter toute la production de données vers une disponibilité ouverte.

Les institutions actives dans l'instruction publique ainsi que les tiers, les parties intéressées commerciales ou non commerciales tirent profit des données ouvertes. C'est la demande et les ressources disponibles qui déterminent jusqu'à quelle date on remonte pour traiter les données déjà existantes, mais disponibles sous une forme peu compréhensible ou non accessible numériquement.

Dans ce contexte, collecter de nouvelles données pour combler d'éventuelles données lacunaires constitue une autre tâche. Le concept des données ouvertes part du principe que seules les données déjà existantes sont rendues disponibles. S'il semble judicieux d'exploiter d'autres sources de données et ce, jusqu'au niveau des établissements individuels de formation (écoles), alors il faudrait chercher le dialogue avec les acteurs concernés. Cela nécessiterait ici une politique claire de l'homogénéisation et de la standardisation de la collecte des données et de leur description. À cet égard, il faut si possible résoudre pour tous, de façon uniforme et contraignante, toutes les questions sensibles au niveau de la protection des données (approches PIM, dossier éducatif électronique pour tous) afin de faire concorder l'autodétermination informationnelle individuelle et l'intérêt public dans l'évaluation des données. Il convient donc d'effectuer une pesée minutieuse des intérêts afin de déterminer quelles données individuelles sont réellement nécessaires pour la politique de formation et sous quelle forme, et si cet intérêt était éventuellement prépondérant en comparaison aux intérêts individuels de protection ou si les mesures d'anonymisation peuvent suffisamment prendre en compte les intérêts légitimes de protection.

Toutefois, le grand avantage d'une politique des données ouvertes de formation réside non seulement dans la possibilité de réaliser des analyses encore plus nombreuses et détaillées, mais de susciter de nouvelles idées, d'inciter des analyses créatives et d'utiliser le savoir qui a fait éventuellement défaut jusqu'à présent – en bref, d'arriver à de nouvelles perspectives de manière démocratique et transparente, sans rendre les utilisateurs dépendants d'algorithmes opaques de recherche et d'évaluation ou faire en sorte que les résultats soient accessibles uniquement à un cercle restreint de parties intéressées.

4.3.4 Bibliographie²⁰⁷

AFS [Archives fédérales suisses] (2017). Rapport « Dateninventarisierung Bund », 2. Inventarisierungsrunde, Berne (https://www.egov.ch/index.php/download_file/force/1417/3337/).

Année politique suisse, une entreprise de l'Académie suisse des sciences humaines et sociales, <https://anneepolitique.swiss> (données à partir de 1966).

207 Tous les sites Internet ont été vérifiés entre le 5 et le 15 janvier 2019.

- ArchDV Gemeinden (2014). Direktionsverordnung über die Verwaltung und Archivierung der Unterlagen von öffentlich-rechtlichen Körperschaften nach Gemeindegesetz und deren Anstalten BSG 170.711. Bern (<https://www.belex.sites.be.ch/frontend/versions/1458>).
- CDF [Contrôle fédéral des finances] (2018). Mise en œuvre de la stratégie Open Government Data Suisse à la Confédération (audit transversal) – Archives fédérales suisses. Berne (https://www.efk.admin.ch/images/stories/efk_dokumente/publikationen/wirtschaft_und_verwaltung/allgemeine_verwaltung/17491/17491BE_Endg%C3%BCltige_Fassung_V04.docx.pdf).
- CDIP [Conférence suisse des directeurs cantonaux de l'instruction publique] (2018). Stratégie de numérisation. Stratégie de la CDIP du 21 juin 2018 pour la gestion de la transition numérique dans l'instruction publique (https://edudoc.ch/record/131562/files/pb_digi-strategie_f.pdf).
- Chancellerie fédérale (2016). Initiative populaire « Pour le couple et la famille – Non à la pénalisation du mariage » (brochure en vue des votations). Berne (https://www.admin.ch/dam/gov/fr/Dokumentation/Abs-timmungen/Februar16/Erlaeuterungen_FR_bf.pdf.download.pdf/Erlaeuterungen_FR_bf.pdf).
- Chignard, Simon (2013). Une brève histoire de l'Open Data, Paris (<http://parisinnovationreview.com/article/une-breve-histoire-de-lopen-data>).
- Commission européenne (1989). Lignes directrices pour améliorer la synergie entre secteur public et secteur privé sur le marché de l'information. Luxembourg : Office des publications officielles des Communautés européennes (<https://publications.europa.eu/fr/publication-detail/-/publication/7c37bbee-4363-4ec7-91ff-b6848142ec97>).
- Commission européenne (2018). Open Data Maturity in Europe 2018. Report. Bruxelles (https://www.europendataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf).
- Conseil de l'UE (2017). Tallinn Declaration on eGovernment at the ministerial meeting during Estonian Presidency of the Council of the EU on 6 October 2017. Tallinn (https://ec.europa.eu/newsroom/document.cfm?doc_id=47559).
- Conseil fédéral (2014). Stratégie en matière de libre accès aux données publiques en Suisse pour les années 2014 à 2018. In : Feuille fédérale 2014:3347-3360 (<https://www.admin.ch/opc/fr/federal-gazette/2014/3347.pdf>).
- Conseil fédéral (2018). Stratégie pour des données administratives ouvertes en Suisse 2019–2023 (Stratégie en matière de libre accès aux données publiques, stratégie OGD). In : Feuille fédérale (en cours de parution).
- CSRE [Centre suisse de coordination pour la recherche en éducation] (2014). L'éducation en Suisse, rapport. Aarau (https://www.skbf-csre.ch/fileadmin/files/pdfs/bildungsberichte/2014/bildungsbericht_14_f.pdf).
- CSRE [Centre suisse de coordination pour la recherche en éducation] (2018). L'éducation en Suisse, rapport. Aarau.
- Desrosières, Alain (2005). La politique des grands nombres. Histoire de la raison statistique. La Découverte (1993).
- FNS [Fonds national suisse] (2017). Déclaration de principe du FNS sur le libre accès aux données de la recherche (« Open Research Data »). Berne (http://www.snf.ch/fr/leFNS/points-de-vue-politique-de-recherche/open_research_data/Pages/default.aspx).
- G8 (2013) Policy paper : G8 Open Data Charter and Technical Annex. Londres : Cabinet Office (<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>).
- Groupe d'experts (2018). Rapport du groupe d'experts concernant le traitement et la sécurité des données. Berne (<https://www.news.admin.ch/news/message/attachments/55754.pdf>).

- ISDC [Institut suisse de droit comparé] (2018). Gutachten zur Datenportabilität sowie zu Regelungen betreffend die Wiederverwendung von Daten. Lausanne (<https://www.bakom.admin.ch/dam/bakom/de/dokumente/informationsgesellschaft/datenpolitik/Endbericht%20Datenpolitik%202018.pdf.download.pdf/Endbericht%20Datenpolitik%202018.pdf>)
- Kellerhals, Andreas (2018). Open Government Data – ein Beitrag zur Digitalisierung von Demokratie und öffentlicher Verwaltung. In : Yearbook of Swiss Administrative Sciences (<https://ssas-yearbook.com/articles/10.5334/ssas.120>).
- LOGA (1997). Loi sur l'organisation du gouvernement et de l'administration, RS 172.010 (<https://www.admin.ch/opc/fr/classified-compilation/19970118/index.html>).
- LPD (1992). Loi fédérale sur la protection des données, RS 235.1 (<https://www.admin.ch/opc/fr/classified-compilation/19920153/201401010000/235.1.pdf>).
- LTrans (2004). Loi fédérale sur le principe de la transparence dans l'administration, RS 152.3 (<https://www.admin.ch/opc/fr/classified-compilation/20022540/index.html>).
- Mastronardi, Philippe (1991). Kriterien der demokratischen Verwaltungskontrolle. Analyse und Konzept der parlamentarischen Oberaufsicht im Bund. Basel, Frankfurt/M : Helbling & Lichtenhahn.
- Neuroni, A. C., Riedl, R., & Brugger, J. (2013). Swiss Executive Authorities on Open Government Data – Policy Making Beyond Transparency and Participation. 2013 46th Hawaii International Conference on System Sciences, 1911–1920. DOI : <https://doi.org/10.1109/HICSS.2013.19>.
- Obama, Barack (2009). Directive gouvernementale ouverte (« *Open Government Directive* »). Washington : The White House (<https://obamawhitehouse.archives.gov/open/documents/open-government-directive>).
- Obama, Barack (2011). Partenariat gouvernemental ouvert (« *Open Government Partnership* »). Washington : The White House (<https://www.opengovpartnership.org>).
- OCDE [Organisation de coopération et de développement économiques] (2015). Data-Driven Innovation for Growth and Well-Being. Paris : OCDE (<https://doi.org/10.1787/9789264229358-en>).
- ODArch communes, annexe (2018). Délais de conservation minimaux destinés aux communes municipales, aux communes mixtes, aux syndicats de communes, aux sections de commune, aux corporations de digue et à leurs établissements, RSB 170.711–A1. Berne (https://www.belex.sites.be.ch/frontend/structured_documents/4083/download_pdf_annex.pdf).
- OFS [Office fédéral de la statistique] (2016). Programme pluriannuel de la statistique fédérale 2016–2019. Rapport, Neuchâtel (<https://www.bfs.admin.ch/bfs/fr/home/ofis/statistique-publique/systeme-statistique-publique/programme-pluriannuel.assetdetail.5453187.html>).
- OFS [Office fédéral de la statistique] (2018). Programme statistique annuel de la Confédération 2018, Neuchâtel (<https://www.bfs.admin.ch/bfsstatic/dam/assets/4282920/master>).
- Parlement européen (2003). Directive concernant la réutilisation des informations du secteur public. Bruxelles : Journal officiel de l'Union européenne (<https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32003L0098&from=DE>).
- Rudin, Beat [Préposé à la protection des données du canton de Bâle-Ville] (2019). Evaluationen im Bildungsbereich. (<http://www.dsb.bs.ch/taetigkeitsbereiche/bildung/evaluationen-im-bildungsbereich.html>).
- Schöchli, Hansueli (2018a). Bund publizierte jahrelang falsche Zahlen zur Heiratsstrafe – CVP prüft Abstimmungsbeschwerde. Zurich : NZZ (<https://www.nzz.ch/schweiz/bund-publizierte-jahrelang-falsche-zahlen-zur-heiratsstrafe-cvp-prueft-abstimmungsbeschwerde-ld.1395677?reduced=true>).

Schöchli, Hansueli (2018b). Initiative populaire « Pour le couple et la famille – Non à la pénalisation du mariage » (« *Volksinitiative* « Für Ehe und Familie – gegen die Heiratsstrafe » »). Zurich : NZZ (<https://www.nzz.ch/schweiz/die-steuervorlage-zur-heiratsstrafe-ist-sistiert-ld.1396268>).

Trachsel, Christian (2018). SBB Open Data. Berne (http://www.digitale-nachhaltigkeit.unibe.ch/unibe/portal/fak_wiso/a_bwl/inst_wi/abt_digital/content/e90958/e490158/e636040/e652610/e682224/OpenData-Vorlesung2018_Termin08_Gastreferat_SBB_Trachsel_ger.pdf).

Van Eechoud, Mireille (2016). Open Data Values, Calculating and Monitoring the Benefits of Public Sector Information Re-use. In : Studien zur Informationsfreiheit 3, 107–142. Baden-Baden : Nomos.

Weber, Rolf H., Thouvenin, Florent (2017). Gutachten zur Möglichkeit der Einführung eines Datenportabilitätsrechts im schweizerischen Recht und zur Rechtslage bei Personal Information Management Systems (PIMS). Zürich (https://www.bakom.admin.ch/dam/bakom/de/dokumente/informationsgesellschaft/datenpolitik/180321%20BJ-Gutachten_final.pdf.download.pdf/180321%20BJ-Gutachten_final.pdf).

Wiedmer, Anne et Seiberth, Corinna (2015). Konzept : Rechtliche Rahmenbedingungen zur Publikation von Daten als Open Government Data (OGD). Bern (https://www.egovernment.ch/index.php/download_file/force/473/3337/).

Sites Internet :

<http://afla.md>

<http://lab.evodevo.it/romascuola/viewer>

<http://opendefinition.org/od/2.1/en/> ou <https://opendefinition.org/history>

<https://eckdaten.ch/>

<https://5stardata.info/de/>

https://de.wikipedia.org/wiki/Open_Data

<https://hoekiesikeenschool.nl>

<https://ogd.tg.ch>

<https://okfn.org/opendata>

<https://opendatahandbook.org>

<https://opendefinition.org/history>

<https://www.ge.ch/document/strategie-matiere-libre-acces-aux-donnees-publiques-administration-genevoise-open-data>

<https://opendatainception.io/#?q=Switzerland>

4.4 Interopérabilité

Centre de recherche Durabilité Numérique (« Digitale Nachhaltigkeit »), Université de Berne

- Dr Matthias Stürmer

Table des matières

4.4.1	Définition de l'interopérabilité	97
4.4.1.1	Conditions préalables pour garantir l'interopérabilité	97
4.4.1.2	Niveaux de l'interopérabilité	97
4.4.1.3	Résultats scolaires, un exemple tiré du contexte éducatif	98
4.4.2	Interopérabilité dans d'autres secteurs	98
4.4.3	Avantages et défis de l'interopérabilité	99
4.4.4	Normes internationales d'interopérabilité dans le secteur de la formation	100
4.4.4.1	Cadre d'interopérabilité des écoles / des systèmes (« <i>Systems / School Interoperability Framework SIF</i> »)	100
4.4.4.2	« Common Education Data Standards » (CEDS)	100
4.4.4.3	« IMS Global Learning Consortium »	100
4.4.4.4	« Ed-Fi Alliance » et « Project Unicorn »	101
4.4.5	État actuel en Suisse	101
4.4.6	Recommandations d'action	101
4.4.7	Références	102

4.4.1 Définition de l'interopérabilité

L'interopérabilité est la possibilité d'échanger et d'utiliser des informations provenant de deux ou plusieurs systèmes ou composants (IEEE 1991). Transférée dans le contexte des données, l'interopérabilité crée la compatibilité et permet ainsi la portabilité des données entre les modèles de données et les langages de requêtes, même si les données proprement dites se trouvent dans différents formats sur des systèmes distribués (Chen et al. 2008). De ce fait, cela permet par ex. de trouver et de partager des données, bien que celles-ci soient stockées sur différents serveurs, dans différentes banques de données ou sur plusieurs systèmes de stockage. Contrairement à la notion d'« intégration », où les systèmes sont étroitement couplés entre eux, l'« interopérabilité » permet la coexistence de différents systèmes informatiques autonomes et de solutions logicielles que l'on a probablement implémentées sur différents systèmes d'exploitation et langages de programmation.

Les données différemment structurées (données hétérogènes) et les incompatibilités de formats des données qui en résultent empêchent l'interopérabilité des données. Par le biais de normes communes de données et de métadonnées cohérentes, il est possible de formater des informations d'une manière uniforme, de sorte qu'elles soient compatibles entre elles. Les normes sont définies de manière très précise à l'aide de spécifications étendues. Si la spécification d'une norme est accessible publiquement et gratuitement, il est possible d'implémenter cette norme par le biais d'une solution logicielle sans frais de licence ou autres charges similaires et s'il existe par ailleurs une implémentation « *Open Source* » de la norme, alors on peut désigner celle-ci comme étant une « norme ouverte » (Tiemann 2006).

4.4.1.1 Conditions préalables pour garantir l'interopérabilité

Il est possible d'atteindre l'interopérabilité des systèmes lorsque les conditions préalables suivantes sont remplies (GridWise 2009 ; Ise 2014) :

1. **Standardisation** : les formats de données et autres propriétés des informations sont définies de façon uniforme.
2. **Interchangeabilité** : il est possible d'échanger des parties du système alors que le système global continue de fonctionner.
3. **Adoption des normes** : les normes de branches déjà existantes sont adoptées.
4. **Architecture ouverte des systèmes** : des interfaces diffusées et neutres vis-à-vis des fabricants sont appliquées, que tous les développeurs de logiciels peuvent implémenter.

4.4.1.2 Niveaux de l'interopérabilité

Les systèmes et leurs données peuvent être interopérables sur quatre niveaux différents (Sheth 1999 ; Nagarajan et al. 2006) :

1. **L'interopérabilité des systèmes** signifie qu'il est possible d'échanger physiquement des données sur des supports de données ou à l'aide de protocoles de communication (Internet). → Limitation : il est certes possible d'échanger des données entre différents systèmes, mais comme elles sont probablement enregistrées dans différents formats techniques (par ex. US-ASCII vs UTF-8), elles ne sont pas compatibles entre elles.
2. **L'interopérabilité syntaxique** comprend des formats de données compatibles entre eux en ce qui concerne les types de caractères et leurs agencements, de sorte qu'il est possible d'importer sans pertes des données d'un système vers un autre. → Limitation : certes, les données peuvent être interopérables d'un point de vue technique, de sorte qu'elles peuvent être regroupées par différents systèmes. Toutefois, elles ne sont pas nécessairement comparables entre elles, puisqu'elles ont été saisies dans les différents systèmes par exemple avec des échelles différentes.
3. **L'interopérabilité structurelle** est garantie lorsque les structures de données (par ex. tous les champs de banques de données) de deux systèmes sont totalement compatibles entre elles et, de ce fait, lorsqu'aucune unité de données n'est perdue lors d'une importation ou d'une exportation des données. Par

conséquent, s'il est possible de transférer toutes les informations d'un format de données vers un autre par le biais d'une transformation de données, les données sont syntaxiquement et structurellement interopérables entre elles. → Limitation : toutefois, cela ne veut pas encore dire que la signification réelle des données transformées doit également être identique.

4. **L'interopérabilité sémantique** concerne la compatibilité du contenu des données, afin de pouvoir également les transférer correctement d'un système vers un autre en termes de signification. Des normes sémantiques avec des ontologies étendues et des métadonnées uniformes peuvent résoudre cette problématique.

4.4.1.3 Résultats scolaires, un exemple tiré du contexte éducatif

Dans les écoles suisses, les notes sont habituellement assignées de 1 à 6, 1 étant la plus mauvaise et 6 la meilleure note. Dans certaines écoles, on attribue également des demi-notes, dans d'autres des quarts de notes. D'autre part, en Allemagne, on octroie également des notes entre 1 et 6, avec toutefois 1 comme meilleure et 6 comme plus mauvaise note. Aux États-Unis, on attribue habituellement les lettres A à F, dans d'autres institutions d'éducation, les notes varient entre 1 et 10 ou entre 1 et 100. Par ailleurs, outre les notes en tant que telles, des informations d'accompagnement sont souvent assignées en plus des évaluations, mentionnant par exemple qui a attribué la note, quel rôle cette personne joue, comment la note est-elle composée (par ex. à partir de notes partielles), pour quelles raisons cette note a-t-elle été attribuée (justification), pour quelle preuve de performance a-t-elle été rendue, où a-t-elle été archivée, à quel moment ou durant quelle période la note a-t-elle été appliquée, dans quelle école a-t-elle été décernée, etc.

D'un point de vue technique, l'interopérabilité des systèmes (1) permet de transférer les résultats scolaires d'un système à un autre, tout en utilisant probablement une syntaxe différente (chiffres versus lettres). Cette problématique est résolue par l'interopérabilité syntaxique (2), en rendant par ex. les différentes échelles comparables grâce à une transformation des valeurs. Il se peut toutefois que les informations importantes d'accompagnement soient perdues lors d'un transfert de notes, car les différents systèmes ne peuvent pas enregistrer les mêmes données. À l'aide de l'interopérabilité structurelle (3), il est garanti que la totalité des informations soit par ex. transférée par le biais d'un inventaire des formats de données.

Toutefois, il n'est pas encore possible de garantir que les contenus des différents champs de données sont réellement comparables, puisqu'ils utilisent probablement des types de description différents ou la signification des données peut être comprise différemment. Seule l'interopérabilité sémantique (4) assure une compatibilité du contenu des données. Les ontologies communes, les réseaux de relations logiques entre les objets et leurs relations constituent la base de l'interopérabilité sémantique. Ceux-ci peuvent être représentés dans des métadonnées uniformes, de sorte que, par exemple, les informations d'accompagnement des notes soient appliquées de façon cohérente. Toutefois, il est par conséquent difficile et laborieux pour les organisations impliquées de s'accorder sur tous les aspects pertinents et les différentes possibilités d'interprétation et de points de vue et de les rédiger d'une manière claire.

4.4.2 Interopérabilité dans d'autres secteurs

L'interopérabilité ne représente pas seulement un défi dans le secteur de la formation. Depuis de nombreuses années, les activités de recherche et les efforts visant à améliorer l'interopérabilité sont en cours dans d'autres secteurs également.

Par exemple, l'interopérabilité sémantique joue un rôle important dans le secteur de la santé depuis plus de dix ans (Bicer et al. 2005 ; Mead 2006 ; Garde et al. 2007). Il est tout à fait compréhensible que la signification exacte des diagnostics joue un rôle vital dans ce secteur. On a donc mis au point les normes dites « Health Level 7 » (HL7) dans le but de créer une compréhension uniforme de tous les objets et processus pertinents dans le domaine médical et de clarifier ainsi leur signification (sémantique).²⁰⁸ En Suisse, eHealth Suisse a publié plusieurs documents à cet égard qui démontrent l'importance élevée de l'interopérabilité sémantique dans le secteur de la santé et qui donnent en même temps des critères d'évaluation concrets et des recommandations concernant la sémantique et les métadonnées.²⁰⁹

208 <http://www.hl7.org/>

209 <https://www.e-health-suisse.ch/fr/technique-semantique/interopabilite-semantique/normes-semantiques.html>

D'un point de vue historique, l'interopérabilité sémantique a été abordée en premier lieu dans le cadre des systèmes d'information géographique (SIG) (Bishr 1998 ; Bishr et al. 1999 ; Harvey et al. 1999 ; Ouksel and Sheth 1999). De même que dans le cas des données dans le domaine de la formation et de la santé, l'importance d'informations uniformes joue également un rôle important dans le cas des géodonnées. Il est donc possible d'importer et de traiter rapidement de nombreuses données syntaxiquement par le biais des différentes applications SIG, mais ce n'est que leur interopérabilité sémantique qui garantit un contenu réellement comparable.

Dans d'autres branches également, comme par exemple dans le secteur de l'énergie, le traitement automatisé de quantités importantes de données joue un rôle toujours plus important (Uslar 2005 ; GridWise 2009 ; Howell et al. 2017). La mise en réseau croissante des appareils via les réseaux électriques intelligents (« *Smart Grids* ») ainsi que la libéralisation du marché de l'électricité augmentent l'importance de l'interopérabilité de manière déterminante.

4.4.3 Avantages et défis de l'interopérabilité

Dans le secteur de la formation, l'interopérabilité constitue une condition technique préalable afin de pouvoir tirer profit du potentiel des données disponibles. Plus le niveau d'interopérabilité est élevé (cf. ci-dessus), plus les données peuvent être facilement échangées et comparées entre les systèmes. Les systèmes interopérables dans le secteur de la formation créent les conditions préalables pour un échange intégral de données, ce qui représente un avantage concret de manière différente (Wayman et al. 2004 ; Collins et al. 2007 ; U.S. Department of Education 2010 ; Ise 2014) :

1. **L'interopérabilité réduit la dépendance envers les concepteurs de logiciels.** Si les produits logiciels et les services en ligne provenant de différents fournisseurs peuvent importer et exporter leurs données dans un format commun, il est techniquement facile de passer d'une solution logicielle mise en œuvre à une autre.
2. **L'interopérabilité améliore l'échange des données.** Lorsque les enfants et les adolescents changent d'école, il est possible, dans le cas de systèmes administratifs scolaires interopérables, de reprendre par exemple les données éducatives du système informatique de l'école actuelle et de les transférer de façon fiable vers le système de la nouvelle école. À cet égard, les enseignants et parents ne doivent saisir plus qu'une fois les données sur les élèves (principe « *Once Only* »), ce qui encourage la qualité et la cohérence des données.
3. **L'interopérabilité permet la connexion de données pertinentes.** Ce faisant, les enseignants, responsables d'école et parents peuvent obtenir par exemple un aperçu complet et récent des prestations des élèves. Par ailleurs, la connexion des données permet la comparaison de classes et d'établissements scolaires afin d'identifier les meilleures pratiques (« *Best Practices* ») ou de faire appel à des données récentes et exhaustives comme source adéquate d'informations dans le cas de questions concrètes et de processus décisionnels.
4. **L'interopérabilité facilite les analyses statistiques.** Ainsi, par le biais de données anonymisées et agrégées, les décideurs, chercheurs, journalistes, ONG et la société civile peuvent par ex. obtenir de nouveaux résultats en ce qui concerne les performances, les certifications, l'égalité des chances, les filières de formation, etc., ce qui conduit à une plus grande transparence et à une égalité des chances accrue dans le système éducatif.
5. **L'interopérabilité encourage les innovations dans le secteur de la formation.** Par le biais de normes de données uniformes, les pouvoirs publics et les fabricants privés de logiciels peuvent développer de nouveaux produits et services en ce qui concerne l'apprentissage en ligne (« *e-Learning* »), le portefeuille électronique (« *e-Portfolio* »), la « *Block-chain* », le web des données (« *Linked Data* »), etc., ce qui génère des innovations et améliore finalement la performance du système éducatif.

Les défis typiques de l'interopérabilité dans le secteur de la formation concernent trois domaines (Ise 2014) :

1. **Organisation** : il est difficile de parvenir à une compréhension commune et à une entente claire sur la syntaxe et la sémantique des données de formation.
2. **Architecture** : dans le cas des écoles et d'autres établissements de formation, il n'existe pas d'architecture uniforme d'entreprise (outils logiciels [« *Software Tools* »], banques de données, etc.).
3. **Technique** : l'infrastructure technique n'est pas axée sur l'interopérabilité des données et des processus.

En outre, en raison du système éducatif fédéral, il n'existe guère de normes et de prescriptions spécifiques au contexte suisse précisant en ce qui concerne les formats de données pour les données de formation dans les écoles. Aussi bien les cantons que les communes sont autonomes dans le choix de leurs outils logiciels et des normes de données qui y sont liées. L'offre actuellement utilisée en Suisse de solutions d'administration scolaire et de services en ligne, qui stockent toutefois leurs données principalement dans leurs propres banques et formats de données propriétaires, est donc très diversifiée (cf. ci-dessous).

4.4.4 Normes internationales d'interopérabilité dans le secteur de la formation

4.4.4.1 Cadre d'interopérabilité des écoles / des systèmes (« *Systems / School Interoperability Framework SIF* »)

Le cadre d'interopérabilité des écoles, également nommé cadre d'interopérabilité des systèmes (« *Systems / School Interoperability Framework* », SIF), est une norme ouverte de données pour l'échange de données scolaires du jardin d'enfants à l'école secondaire et placée sous la responsabilité de l'« *Access 4 Learning (A4L) Community* ». Initialement, le SIF a été mis en œuvre aux États-Unis, au Canada, au Royaume-Uni et en Australie. Actuellement, il se généralise toutefois toujours plus en Inde et dans d'autres pays (Jakimoski 2016). Grâce à des données de formation SIF, les écoles et autres établissements de formation peuvent échanger horizontalement entre les systèmes informatiques et verticalement entre les institutions (Collins et al. 2007).

D'un point de vue technique, la norme SIF communique depuis 2013 en tant que service Internet via le protocole dit « *Representational State Transfer* » (REST). Le modèle de données est décrit dans le format « *Extensible Markup Language* » (XML), les données elles-mêmes sont consultables à l'aide de requêtes « *Hypertext Transfer Protocol* » (HTTP) sous le format « *JavaScript Object Notation* » (JSON). La spécification de tous les champs de données et éléments de contenu comporte plusieurs centaines de pages, étant disponible pour l'Australie, les États-Unis et le Royaume-Uni²¹⁰ en une version chacun.

4.4.4.2 « *Common Education Data Standards* » (CEDS)

Les « *Common Education Data Standards* » (CEDS) sont une initiative lancée en 2009 aux États-Unis, afin de créer une compréhension commune des données dans le domaine de la formation. Un modèle unifié de données ainsi qu'une collection de métadonnées doivent améliorer la compréhension, la comparaison et l'échange de données de formation entre les États et les districts. À cet égard, les CEDS ont publié un guide complet²¹¹, dans lequel tous les objets pertinents sont exactement spécifiés. Au total, la norme consiste actuellement en plus de 800 tableaux et plus de 5000 champs de banques de données. Par ailleurs, de nombreux outils logiciels pour le traitement des données CEDS ainsi que des manuels supplémentaires et des applications tierces sont disponibles pour la norme.

4.4.4.3 « *IMS Global Learning Consortium* »

L'« *IMS Global Learning Consortium* » a été fondé en 1997 aux États-Unis et a, depuis lors, publié de nombreuses normes de données dans le secteur de la formation. Ces normes sont rendues publiques et sont accessibles gratuitement. Le format XML « *IMS Content Packaging* » décrit par exemple les structures de stockage d'objets d'apprentissage en ligne ou l'« *IMS Global Competencies and Academic Standards Exchange* » (CASE) définit les compétences possibles des élèves. Outre le modèle des données, CASE comporte par ailleurs une documentation REST API exhaustive concernant la requête de données au format

210 <https://www.a4l.org/page/SIFSpecifications>

211 <https://ceds.ed.gov/dataModel.aspx>

JSON²¹². L'« IMS Global Learning Consortium » certifie également des produits logiciels qui implémentent les normes IMS. Actuellement, 426 solutions logicielles sont mentionnées dans le répertoire des produits²¹³.

4.4.4.4 « Ed-Fi Alliance » et « Project Unicorn »

La « *Ed-Fi Alliance* », financée par la Fondation Michael & Susan Dell, a pour but d'améliorer la saisie et l'utilisation de données de formation par le biais d'une norme ouverte de données et d'outils logiciels. Lors de la création du « *Project Unicorn* », elle implique également l'organisation InnovateEDU, qui souhaite également accroître l'interopérabilité des données dans le secteur de la formation²¹⁴.

4.4.5 État actuel en Suisse

Basé sur un échantillon de solutions logicielles dans le secteur de la formation²¹⁵, on a constaté qu'aucune norme uniforme de données n'est actuellement mise en œuvre dans le domaine de la formation suisse. Les quelque 20 concepteurs de solutions d'administration scolaire suisses utilisent par exemple largement leurs propres normes et formats de données, de sorte qu'un échange de données n'est possible qu'à l'aide de conversions laborieuses des données. Aucune norme de données ne s'est établie, ce qui est également le cas dans d'autres domaines logiciels proches de l'éducation tels que les systèmes de gestion de l'apprentissage (« *Learning Management Systems* », LMS), l'interopérabilité n'étant pas non plus garantie dans ce domaine.

Dans le domaine des ressources électroniques d'enseignement et d'apprentissage, educa.ch a publié en 2018 une adaptation suisse du standard international « *Learning Object Metadata Standard* » (LOM), dit standard LOM-CH²¹⁶. Il est donc possible d'échanger les contenus numériques qui doivent être enseignés ou leurs métadonnées entre différents systèmes.

Par ailleurs, dans le contexte suisse, le projet FIDES (Fédération de services d'identités pour l'espace suisse de formation)²¹⁷ peut probablement fournir une contribution à l'interopérabilité dans l'instruction publique. Le projet a pour objectif de relier entre elles les identités numériques d'une manière uniforme. L'accès aux services en ligne pertinents est donc simplifié pour les acteurs individuels de la formation.

4.4.6 Recommandations d'action

Basé sur les résultats indiqués ci-dessus, on en déduit les recommandations d'action suivantes pour le contexte suisse :

1. **Une enquête effectuée auprès des fournisseurs suisses actuels de solutions d'administration scolaire et d'autres prestataires de services dans le domaine de la formation** vise à clarifier les fonctions détaillées des applications et les types de données qu'elles garantissent, dans quels formats et systèmes de stockage. À cet égard, il convient également d'interroger les fournisseurs s'ils connaissent déjà les normes des données et les cadres d'interopérabilité (SIF, CEDS, etc.) et, le cas échéant, s'ils les ont implémentés et quels stockages futurs des données ils recommandent ou ont prévu d'implémenter. Ces indications permettent de faire un tour d'horizon sur l'offre de prestations du paysage actuel des logiciels éducatifs suisses, qui peut servir de base à d'autres décisions concernant une future norme d'interopérabilité des données.
2. À l'aide d'une **enquête complémentaire auprès des cantons et des communes, il convient de clarifier la diffusion des solutions logicielles actuellement utilisées dans le domaine de la formation.** Par ailleurs, dans le cadre de cette enquête, il convient de clarifier si et, le cas échéant, quels sont les besoins supplémentaires concernant les fonctionnalités et l'échange des données et quels formats et normes de données sont connus ou seraient souhaités.

212 <https://www.imsglobal.org/activity/case#PublicDocuments>

213 <https://www.imsglobal.org/cc/statuschart.cfm>

214 <https://www.prounicorn.org/>

215 Relevés d'educa.ch au cours des années 2015-2018.

216 <https://www.educa.ch/fr/acces-ligne/lom-ch>

217 <https://www.educa.ch/fr/dossiers/eid/une-federation-de-services-d-identites-fides>

3. Sur la base des résultats de l'enquête et des considérations stratégiques relatives à l'utilisation de données de formation, il convient d'examiner **l'adaptation d'une norme d'interopérabilité existante telle que SIF ou CEDS au contexte suisse**. Ce faisant, il est possible de tirer profit de l'expérience acquise sur plusieurs années à l'étranger et, parallèlement, de prendre en compte les besoins du système éducatif suisse. Si la mise en place d'une telle norme se révèle appropriée, il convient d'élaborer les spécifications correspondantes au cours d'une procédure participative et de les partager publiquement et gratuitement. Il faut examiner une collaboration avec l'association eCH qui pourrait fonctionner comme comité de standardisation : elle a développé plus de 130 normes d'interopérabilité dans le domaine de la cyberadministration et les actualise régulièrement²¹⁸.
4. Il faut **implémenter cette nouvelle norme suisse de données de formation comme logiciel « Open Source »** après la phase de spécification, de sorte que les bibliothèques de programmes correspondantes facilitent l'intégration de la norme aux concepteurs de logiciels et accroissent la diffusion de la norme.
5. Sur la base de la nouvelle norme des données de formation, il sera possible de **mettre en place une ou plusieurs plateformes cantonales**, par le biais desquelles les écoles et autres établissements de formation pourront échanger dans un format uniforme.
6. Par ailleurs, la norme des données de formation permet de garantir l'interopérabilité des systèmes logiciels en tant que **nouveau critère dans les appels d'offres publics** de solutions d'administration scolaire.
7. À plus long terme, une norme de données uniforme permettra des innovations dans le contexte de la formation. Afin d'accélérer le processus, il est recommandé de procéder au **lancement des projets pilotes avec des données de formation**, par exemple pour promouvoir la mise en œuvre technique de « *Linked Data Endpoints* » (interfaces ouvertes) ou d'applications basées sur la « *Block-chain* » avec des données de formation.

4.4.7 Références

- Bicer, V., Laleci, G.B., Dogac, A., Kabak, Y., 2005. Artemis Message Exchange Framework : Semantic Interoperability of Exchanged Messages in the Healthcare Domain. SIGMOD Rec 34, 71–76. <https://doi.org/10.1145/1084805.1084819>
- Bishr, A.Y., Pundt, H., Kuhn, W., Radwan, M., 1999. Probing the Concept of Information Communities-A First Step Toward Semantic Interoperability, in : Goodchild, M., Egenhofer, M., Fegeas, R., Kottman, C. (Eds.), Interoperating Geographic Information Systems, The Springer International Series in Engineering and Computer Science. Springer US, Boston, MA, pp. 55–69. https://doi.org/10.1007/978-1-4615-5189-8_5
- Bishr, Y., 1998. Overcoming the semantic and other barriers to GIS interoperability. Int. J. Geogr. Inf. Sci. 12, 299–314. <https://doi.org/10.1080/136588198241806>
- Chen, D., Doumeingts, G., Vernadat, F., 2008. Architectures for enterprise integration and interoperability : Past, present and future. Comput. Ind., Enterprise Integration and Interoperability in Manufacturing Systems 59, 647–659. <https://doi.org/10.1016/j.compind.2007.12.016>
- Collins, L., Fruth, L., Sessa, M., Laird, E., 2007. The Right Data to the Right People at the Right Time : How Interoperability Helps America's Students Succeed. Data Qual. Campaign 12.
- Garde, S., Knaup, P., Hovenga, E.J.S., Heard, S., 2007. Towards Semantic Interoperability for Electronic Health Records : Domain Knowledge Governance for openEHR Archetypes. Methods Inf. Med. 46, 332–343. <https://doi.org/10.1160/ME5001>
- GridWise, 2009. Reliability Benefits of Interoperability. GridWise Architecture Council.

²¹⁸ <https://www.ech.ch/de/standards/overviewlist>

- Harvey, F, Kuhn, W., Pundt, H., Bishr, Y., Riedemann, C., 1999. Semantic interoperability : A central issue for sharing geographic information. *Ann. Reg. Sci.* 33, 213–232. <https://doi.org/10.1007/s001680050102>
- Howell, S., Rezgui, Y., Hippolyte, J.-L., Jayan, B., Li, H., 2017. Towards the next generation of smart grids : Semantic and holonic multi-agent management of distributed energy resources. *Renew. Sustain. Energy Rev.* 77, 193–214. <https://doi.org/10.1016/j.rser.2017.03.107>
- IEEE, 1991. IEEE Standard Computer Dictionary : A Compilation of IEEE Standard Computer Glossaries : 610. Inst of Elect & Electronic, New York, NY, USA.
- Ise, O.A., 2014. Towards a Unified University Information System : Bridging the Gap of Data Interoperability. *Am. J. Softw. Eng.* 2, 26–32. <http://pubs.sciepub.com/ajse/2/2/3/index.html>
- Jakimoski, K., 2016. Challenges of Interoperability and Integration in Education Information Systems. *Int. J. Database Theory Appl.* 9, 33–46. <https://pdfs.semanticscholar.org/19cb/b868f2c483f5638dc6b45a-64814c95133aa1.pdf>
- Mead, C.N., 2006. Semantic Interoperability : Now Possible but Still Difficult, Do We Really Need a Better Mousetrap? *J. Healthc. Inf. Manag.*
- Nagarajan, M., Verma, K., Sheth, A.P., Miller, J., Lathem, J., 2006. Semantic Interoperability of Web Services – Challenges and Experiences, in : 2006 IEEE International Conference on Web Services (ICWS'06). Presented at the 2006 IEEE International Conference on Web Services (ICWS'06), pp. 373–382. <https://doi.org/10.1109/ICWS.2006.116>
- Nagarajan, M., Verma, K., Sheth, A.P., Miller, J., Lathem, J., 2006. Semantic Interoperability of Web Services – Challenges and Experiences, in : 2006 IEEE International Conference on Web Services (ICWS'06). Presented at the 2006 IEEE International Conference on Web Services (ICWS'06), pp. 373–382. <https://doi.org/10.1109/ICWS.2006.116>
- Ouksel, A.M., Sheth, A., 1999. Semantic Interoperability in Global Information Systems. *SIGMOD Rec* 28, 5–12. <https://doi.org/10.1145/309844.309849>
- Ouksel, A.M., Sheth, A., 1999. Semantic Interoperability in Global Information Systems. *SIGMOD Rec* 28, 5–12. <https://doi.org/10.1145/309844.309849>
- Sheth, A.P., 1999. Changing Focus on Interoperability in Information Systems : From System, Syntax, Structure to Semantics, in : *Interoperating Geographic Information Systems*. Springer US, Boston, MA, pp. 5–29. https://doi.org/10.1007/978-1-4615-5189-8_2
- Sheth, A.P., 1999. Changing Focus on Interoperability in Information Systems : From System, Syntax, Structure to Semantics, in : *Interoperating Geographic Information Systems*. Springer US, Boston, MA, pp. 5–29. https://doi.org/10.1007/978-1-4615-5189-8_2
- Tiemann, M., 2006. An objective definition of open standards. *Comput. Stand. Interfaces* 28, 495–507. <https://doi.org/10.1016/j.csi.2004.12.003>
- Tiemann, M., 2006. An objective definition of open standards. *Comput. Stand. Interfaces* 28, 495–507. <https://doi.org/10.1016/j.csi.2004.12.003>
- U.S. Department of Education, 2010. Digital Systems Interoperability [WWW Document]. URL <https://www.ed.gov/open/plan/digital-systems-interoperability> (accessed 12.25.18).
- U.S. Department of Education, 2010. Digital Systems Interoperability [WWW Document]. URL <https://www.ed.gov/open/plan/digital-systems-interoperability> (accessed 12.25.18).
- Uslar, M., 2005. Semantic Interoperability Within the Power Systems Domain, in : *Proceedings of the First International Workshop on Interoperability of Heterogeneous Information Systems, IHIS '05*. ACM, New York, NY, USA, pp. 39–46. <https://doi.org/10.1145/1096967.1096976>

Uslar, M., 2005. Semantic Interoperability Within the Power Systems Domain, in : Proceedings of the First International Workshop on Interoperability of Heterogeneous Information Systems, IHIS '05. ACM, New York, NY, USA, pp. 39–46. <https://doi.org/10.1145/1096967.1096976>

Wayman, J.C., Stringfield, S., Yakimowski, M., 2004. Software enabling school improvement through analysis of student data. The Johns Hopkins University.

Wayman, J.C., Stringfield, S., Yakimowski, M., 2004. Software enabling school improvement through analysis of student data. The Johns Hopkins University.

5 Utilisation des données dans le domaine de la formation en comparaison internationale

5.1 Méthode

Afin d'obtenir des valeurs de comparaison sur les objectifs, les réglementations et les formes d'organisation de l'utilisation des données dans le système éducatif, nous avons entrepris de premières recherches exploratoires dans l'espace européen (Danemark, Estonie, France, Irlande, Norvège et Suède) ainsi qu'aux États-Unis. En outre, un sondage²¹⁹ a été envoyé aux réseaux internationaux suivants : European Schoolnet²²⁰, SwissCore²²¹ et groupe de travail DELTA²²². Les recherches réalisées par educa.ch se sont limitées à l'étude du niveau national des pays respectifs concernés puisque le présent rapport met l'accent sur ce niveau, et que des recherches plus détaillées seraient sorties du cadre du présent rapport.

5.2 Efforts déployés dans des pays européens

Les retours des sondages effectués parmi les membres des réseaux internationaux d'educa.ch n'ont pas répondu aux attentes. Les réponses qui n'ont été que brièvement rédigées provenant des réseaux précités ainsi que les recherches réalisées dans différents pays, à savoir l'Estonie, la France, l'Irlande, la Norvège et la Suède, ne permettent de formuler qu'une esquisse superficielle et très générale de la situation en matière d'utilisation des données dans le domaine de la formation dans les pays européens.

219 Voir Annexe III.

220 European Schoolnet est une organisation à but non lucratif qui a son siège à Bruxelles. Elle a pour mission de mettre en réseau des représentantes et représentants de ministères européens de l'éducation et de diriger des projets dans le domaine de la numérisation de l'éducation ainsi que plusieurs groupes de travail sur ce thème.

221 SwissCore est le bureau suisse d'information et de liaison pour les politiques et programmes européens dans le domaine de la recherche, de l'innovation et de la formation.

222 Voir sous : <http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3407&NewSearch=1&NewSearch=1> Le groupe de travail « Digital Education : Learning, Teaching and Assessment » (WG DELTA) a pour but d'aider les États membres de l'UE à s'attaquer aux défis décisifs de leurs systèmes d'éducation en ce qui concerne la formation numérique. Il s'agit de soutenir l'évolution politique au niveau national et européen par des échanges mutuels (*Mutual Learning*) et l'identification de pratiques qui ont déjà fait leurs preuves dans ce domaine.

Il s'est avéré que le thème de l'utilisation des données est surtout mis en relation avec celui de la protection des données et avec la question de savoir qui reçoit l'accès à ces données. Ainsi, on se contente souvent d'enregistrer dans quel cadre on utilise les sources de données mises à disposition, comment on les collecte et à quelle fin elles peuvent être traitées. L'utilisation des données est considérée en priorité comme un outil pour le monitoring de la formation, et elle est utilisée dans ce but.

Dès lors – dans la mesure où nous avons pu nous pencher sur le sujet dans le cadre des recherches effectuées pour le présent rapport – on ne rencontre dans l'espace européen de la formation aucun effort comparable à ceux qui sont déployés dans le présent rapport afin d'analyser l'utilisation des données dans le secteur éducatif.

5.3 Exemples pratiques (Europe)

Bien que nous n'ayons pu trouver aucune orientation stratégique de rang hiérarchique supérieur touchant à l'utilisation des données, dans le cadre de notre recherche, grâce à notre participation à une réunion d'experts du réseau CIDREE sur le thème « Using Data for School Improvement »²²³, nous avons néanmoins découvert des applications intéressantes en matière d'utilisation des données qui mettent en évidence le potentiel que recèle une exploitation systématique des données de l'éducation.

Les exemples pratiques énumérés ci-après sont axés sur la mise en place, resp. l'extension d'un entrepôt national de données (*National Data Warehouse*) (voir chapitre 4.1.5.2, p. 62). Un tel entrepôt national de données permet en effet d'étendre le monitoring national de la formation, de mettre à disposition du grand public des informations sur le système éducatif, et offre la possibilité aux écoles d'améliorer l'utilisation comparative de données sur les processus et les prestations dans le cadre de projets de développement (par ex. dans le cadre d'un programme qui exige le développement de l'école et de l'enseignement).

Danemark – Data Warehouse

L'entrepôt de données (ou *Data Warehouse*) a été mis en place en 2014 par le Ministère danois de l'éducation. Initialement, des informations statistiques sur les écoles primaires ont été collectées dans cet entrepôt de données aux fins de développement de la qualité des écoles, et elles ont été mises à disposition des écoles et des communes. Début 2015, des données générales du niveau ISCED 3 sont venues compléter le Data Warehouse, et enfin, vers la fin 2015, des données provenant de la formation professionnelle lui ont été ajoutées.

Le Data Warehouse offre aux institutions, aux régions, aux communes et au grand public un accès à une série de rapports et graphiques prédéfinis contenant des informations statistiques.

Les rapports sont dynamiques et peuvent être modifiés dans les zones de données existantes à l'aide des variables mises à disposition. En fonction de vos intérêts, vous pouvez être affichés soit au niveau national, soit au niveau régional (communes), soit au niveau de l'institution. En outre, au niveau de l'institution, on peut procéder au suivi de l'évolution au fil du temps de certaines catégories de données, et on peut comparer certaines institutions entre elles.

Le Data Warehouse est notamment utilisé pour soutenir le développement de la qualité dans le domaine de l'éducation. Les informations provenant du niveau de la scolarité secondaire se rapportent aux objectifs de qualité à atteindre pour la formation professionnelle (Undervisningsministeriet 2018). Des indicateurs ont été définis à cet effet. Les données qui en font partie sont continuellement mises à disposition des écoles par le biais du *Data Warehouse*.

Il s'agit donc ici d'un outil d'analyse dynamique. Simultanément, le Data Warehouse offre un accès facile aux données de l'éducation issues de différentes sources (Ministère de l'éducation, écoles, exploitants de solutions d'administration scolaire) ; ce recours au Data Warehouse s'explique par l'exigence des milieux politiques qui entendent soutenir les écoles dans le cadre du développement continu de la qualité des éta-

223 Voir sous : <http://www.cidree.org/cidree-expert-meeting-luxembourg-march-14-15-2019>. Cette réunion d'experts était organisée par le service luxembourgeois SCRIPT (Service de Coordination de la Recherche et de l'Innovation pédagogiques et technologiques du Ministère de l'éducation luxembourgeois) et s'est tenue les 14 et 15 mars 2019 à Luxembourg. educa.ch a participé à cette réunion en qualité de représentante du CSRE.

blissements scolaires.

À l'avenir, cet entrepôt de données sera également étendu aux données provenant du domaine préscolaire. Cela aidera les parents à comparer entre elles les offres de crèches et d'encadrement des enfants dans leur région. Pour ce faire, il faudra réunir et préparer dans le Data Warehouse des données provenant de structures d'accueil collectif de jour pour enfants selon certains indicateurs déterminés.

Estonie – Estonian Education Information System (EHIS)

Le Système d'Information Estonien sur l'Éducation (Estonian Education Information System [EHIS]) (voir chapitre 4.1.5.2, p. 62) fonctionne de manière similaire. La banque de données a déjà été mise en place en 2005 et, comme de coutume en Estonie, elle est raccordée à toutes les autres banques de données nationales.

Au sein de l'EHIS, on collecte aussi bien des informations administratives sur la fonction d'une personne donnée (apprenants, enseignants, institutions) et sur son histoire (par ex. salaire, formations continues ...), que des informations sur les prestations atteintes (diplômes, résultats de tests standardisés, notamment PISA). L'EHIS reçoit ces informations tant directement des institutions éducatives qu'indirectement, également de la part d'autres banques de données, comme par ex. de l'« Exam Information System ».

L'accès direct à la banque de données (*Live Data*) est restreint pour le grand public, et elle est utilisée la plupart du temps par des décideurs et des décideuses. Afin de mettre ces données à disposition du grand public, le système EducationEye a été mis sur pied. C'est dans l'EducationEye que les données provenant de l'EHIS peuvent être visualisées et mises à disposition du grand public et des décideurs. On peut faire afficher les informations jusqu'au niveau des écoles individuelles, et on peut comparer entre eux les institutions éducatives et les régions selon certains indicateurs déterminés. Au niveau des différents individus, ces informations peuvent être consultées par les apprenants concernés (et leurs parents), ainsi que par leurs enseignants et par la direction de l'école concernée (mais pas par le grand public). Les données sont transférées quasi manuellement depuis l'EHIS dans l'EducationEye et ne sont pas synchronisées en live.

L'EHIS est basé sur les exigences d'une stratégie nationale de formation et repose sur une base légale créée spécifiquement à cet effet.

Luxembourg – monitoring national de la formation

Au Luxembourg, les informations administratives sur les établissements scolaires sont stockées dans une série de banques de données. On y trouve des informations sur les apprenants, les enseignants, les diplômes, les reconnaissances de diplômes, les plans horaires, les formations continues des enseignants, etc.

Les « Épreuves standardisées » pour cinq niveaux scolaires différents, que tous les élèves doivent passer tous les deux ans, constituent une autre source de données importante. Elles fournissent des renseignements sur les compétences de chaque enfant en mathématiques, allemand et français. Les résultats de ces tests sont mis à disposition du grand public et de l'administration des écoles au niveau national, et de la direction de l'école au niveau de l'école elle-même, ainsi que des enseignants, des parents et des apprenants au niveau des élèves individuels. Chaque niveau reçoit un accès spécifique par le biais d'un tableau de bord individualisé.

Toutes ces données sont intégrées peu à peu et étendues progressivement dans une banque de données nationale centralisée. Il existe par exemple des efforts qui visent à compléter cette banque de données par des informations supplémentaires relatives aux enseignants.

À partir de cette banque de données nationale, on peut non seulement obtenir des informations pour le monitoring national et pour la planification du système éducatif, mais encore des informations sur des apprenants individuels et leur évolution au fil du temps.

Le système se base sur l'introduction d'un plan de développement scolaire obligatoire valable pour chaque école depuis l'entrée en vigueur de la loi du 6 février 2009 portant organisation de l'enseignement fonda-

mental.²²⁴ Le plan de développement scolaire est obligatoire pour chaque école. C'est l'école elle-même qui détermine son contenu et qui l'élabore ; toutefois, selon l'analyse de la situation, elle est tenue de s'orienter sur les données enregistrées par ses soins (par ex. sondages réalisés auprès des enseignants ou auprès des parents) et sur les données mises à sa disposition (par ex. résultats de tests standardisés).

Le Luxembourg dispose globalement d'un système assez avancé de collecte et d'analyse de données destinées à l'enseignement et au développement plus étendu de ses écoles.

5.4 Efforts déployés aux États-Unis

Aux États-Unis également, des efforts similaires, notamment au niveau national, ont été identifiés. Dans ce pays, on constate que l'on s'est attaqué à la question de l'utilisation des données en tenant compte de la formation de manière spécifique. Toutefois, tout comme en Europe, on s'est chaque fois limité à certains contextes déterminés de l'utilisation des données (monitorage de la formation, protection des données, Open Data, etc.). Il n'existe ni orientation stratégique de rang hiérarchique supérieur en matière d'utilisation des données, ni cadre harmonisé sur cette base allant dans le sens du présent rapport.

Toutefois, dans ces différents contextes, les efforts sont plus avancés aux États-Unis, comme le montrent les exemples suivants :

National Center for Education Statistics – fixation de standards de données

Au niveau du Federal Department of Education, on s'efforce de faire progresser fortement l'interopérabilité dans le secteur éducatif. Dans ce contexte, des standards de données (*Common Education Data Standards*) ont été formulés par le National Center for Education Statistics. Ces standards permettent aux différents groupes cibles d'améliorer la portabilité des données dans leurs domaines respectifs et de tirer parti au maximum de ces formats de données standardisés (CEDS 2017). Par comparaison avec la situation qui prévaut en Suisse, en déployant ce genre d'efforts, les États-Unis ont une longueur d'avance.

National Education Technology Plan – utilisation des données et technologie de la formation

Aux États-Unis, l'Office of Educational Technology (OET) élabore des stratégies nationales pour l'utilisation des technologies de l'éducation, et ce, tous niveaux de scolarité confondus. Ce faisant, les thèmes prioritaires sont notamment la fracture numérique (*Digital Divide*) et l'accessibilité (*Accessibility*).

Avec la publication du *National Education Technology Plan Update*, qui est actualisé chaque année (voir Office of Educational Technology, 2017), les réflexions d'ordre stratégique de l'OET sont largement diffusées. On trouve non seulement dans ce plan des questions sur les achats et l'infrastructure, mais encore des études sur la manière dont les données peuvent être utilisées pour la personnalisation de l'apprentissage et pour rendre visibles les progrès d'apprentissage des élèves.

Avec l'utilisation des données dans le système éducatif, on mise ainsi explicitement sur un niveau stratégique applicable à l'échelon national, mais on se limite, dans cet exemple, aux relations avec les technologies de la formation.

5.5 Conclusion

Au final, on peut partir du principe que, du moins en Europe, dans une grande partie des pays membres de l'UE, les efforts déployés en direction d'un mode d'approche global spécifique aux besoins de l'éducation et une régulation de l'utilisation des données telle que thématifiée dans le présent rapport ne sont pas encore très avancés.

Les observations générales suivantes peuvent être formulées à ce sujet :

- L'utilisation des données dans le domaine éducatif a lieu surtout dans le cadre du monitoring national sur la formation ou en lien avec les efforts déployés en faveur de l'Open Data. Les dispositions juridiques

²²⁴ La loi est consultable en français sous : <http://legilux.public.lu/eli/etat/leg/loi/2009/02/06/n3/fo>

déterminantes pour les possibilités et les limites de l'utilisation des données résident dans les lois sur la protection des données en vigueur dans ces pays (resp. dans le RGPD).

- Dans la plupart de ces pays, les aspects spécifiques au système éducatif de l'utilisation des données ne sont pas réglementés explicitement (sauf dans le cadre du monitoring de la formation).

Toutefois, pour étayer encore davantage ces énoncés, une étude approfondie de la situation qui prévaut dans les différents pays observés serait nécessaire.

Sur la base des questions en retour que nous avons reçues de la part des réseaux internationaux précités, nous avons pu constater qu'hormis dans certains contextes précis que sont le monitoring de la formation, la protection des données ou l'Open Data, les besoins de clarification sont encore prédominants. Les questions posées n'ont pas été comprises comme relevant d'un mode d'approche global, ce qui a rendu difficiles la discussion ainsi que l'accès aux sources d'informations.

Il s'agit ici de sensibiliser tous les acteurs, à l'échelle internationale, à la nécessité de disposer d'une politique d'utilisation des données qui tienne compte de tous les contextes d'utilisation et qui, ce faisant, est en mesure de maintenir un équilibre entre la protection, la sécurité et la création de valeur.

6 Potentiels et défis de l'utilisation des données dans l'éducation

6.1 Introduction

Le débat public actuel en cours sur les défis et les potentiels de la transformation numérique touche aussi le secteur de la formation.²²⁵ Outre des questions d'utilisation de terminaux numériques au sein de l'école et dans l'enseignement, on débat également de manière accrue sur les conséquences du recours à des processus automatisés et à des systèmes de contrôle analytique des données. Les mots-clés de ce débat sont les concepts de « Big Data », « Adaptive Learning », « Learning Analytics » ou « intelligence artificielle ». Ils décrivent les différentes possibilités qui résultent de l'interaction entre une puissance de calcul machine toujours plus performante, les progrès vertigineux accomplis dans le domaine de l'apprentissage algorithmique, et les volumes de données toujours plus étendus et plus détaillés qui en découlent. Cette interaction permet à un grand nombre d'acteurs individuels (dans des conditions optimales également à tous les acteurs) d'observer en temps réel un système, de collecter des informations complètes sur eux, leur comportement et leurs relations, et de reconnaître des modèles dans leur comportement et leurs interactions. On peut en déduire des connaissances sur **des rapports descriptifs et des rapports de causalité** (par ex. quelles sont les caractéristiques qui vont de pair avec un résultat d'apprentissage plus élevé), ainsi que des **énoncés prédictifs** pour de futurs événements (par ex. il est possible de déceler précocement si un élève risque d'être assigné à un autre niveau scolaire ou à une autre classe) et des **propositions prescriptives** pour la planification ou pour des interventions (par ex. quels cours un étudiant devrait suivre pour pouvoir terminer ses études avec le meilleur résultat possible).

Pour le système éducatif, les possibilités d'utilisation offertes par ces technologies sont multiples. Elles vont de la structuration individualisée et centrée sur l'élève de l'apprentissage et de l'enseignement, en passant par une réduction de la charge de travail pour les activités administratives ou répétitives pour les enseignants, jusqu'à un monitoring et un pilotage simplifié et plus rapide du système éducatif (et de ses composants requis pour l'administration et la politique de l'éducation) guidé par les faits (Baer & Norris 2017 ; Kolderie & McDonald 2009 ; New 2016). À l'heure actuelle, ce sont surtout les gains d'efficacité considérables dus à l'utilisation de ces technologies à chaque niveau du système éducatif qui dominent le débat public et scientifique. Ces gains sont réalisés grâce à l'accélération et à l'individualisation des processus, grâce aux décisions empiriquement fondées prises sur la base de ces derniers, et grâce à l'accès aux nouvelles

²²⁵ Ainsi, par ex., le système éducatif se voit confronté de manière accrue à l'exigence consistant à satisfaire à la demande croissante en personnel spécialisé dans le domaine MINT, à dispenser à toutes les apprenants des compétences en informatique et en analyse des données, et à garantir simultanément l'acquisition d'une culture générale dûment fondée à toutes les couches de la population (digitalswitzerland 2017 ; Minsch & Wehrli 2018). Et ce, alors que les responsables sont constamment soumis à une obligation croissante de rendre des comptes, et que les ressources mises à leur disposition sont limitées – lorsqu'elles ne sont pas carrément en recul.

connaissances que permettent de manière générale ces technologies. Une étude de McKinsey de l'année 2013 fait valoir qu'avec le recours systématique aux données de l'éducation à l'échelle mondiale, on pourrait réaliser des gains d'efficacité estimés entre 900 milliards et 1,2 billions de dollars (Manyika et al. 2013). Selon cette étude, près d'un tiers de cette valeur résulte de l'amélioration consécutive de l'enseignement et du niveau plus élevé des compétences ainsi obtenu par les élèves. Le reste de ces gains provient d'un meilleur appariement entre les apprenants et les programmes de formation, resp. les métiers appris, tout comme de l'allocation et de l'utilisation plus efficiente des ressources dans le système éducatif. Toutefois, avec la collecte complète de données granulaires dans le système éducatif, il en résulte également un ensemble de données qui recèle un potentiel d'innovation considérable pour la recherche et l'économie (New 2016 ; Cope & Kalantzis 2015). À l'heure actuelle, leurs effets ne sont encore guère mesurables. Un rapport actuel de l'organisation britannique à but non lucratif Jisc montre, par exemple, qu'à partir des données générées provenant de systèmes d'apprentissage adaptatifs, de telles données pourraient aussi être utilisées afin d'identifier précocement des maladies psychiques ou des problèmes de santé dus au stress chez les étudiants, et qu'elles pourraient offrir de l'aide dans les domaines requis (Hall 2018). Selon des experts, les données de l'éducation pourraient être exploitées avec profit pour la planification et la structuration d'autres tâches publiques. Par exemple, en intégrant dans le système des informations sur la répartition géographique des apprenants et des établissements scolaires, et sur la répartition dans le temps du début des cours afin de planifier les capacités disponibles des moyens de transports publics. À l'inverse, on pourrait imaginer que des données provenant d'autres domaines sociétaux puissent être intégrées dans les modalités d'organisation de l'école et de l'enseignement. Ainsi, par ex., on pourrait tirer parti de la combinaison de systèmes d'apprentissage avec des registres d'inscription numérisés afin d'adapter la définition des arrondissements scolaires aux effets potentiels de la composition des classes sur les résultats de l'apprentissage. Ou on pourrait intégrer ainsi dans l'aménagement des plans horaires des informations sur le comportement de mobilité d'autres groupes de personnes afin de réduire au maximum les heures d'absence des apprenants et des enseignants imputables aux trajets.

La réalisation de ces potentiels présuppose cependant la saisie et la mémorisation étendue et standardisée des données, tout comme l'accès dûment réglementé à ces dernières. Ce projet n'est toutefois pas réalisable sans des investissements considérables dans l'infrastructure technique à tous les niveaux du système éducatif. En outre, à l'heure actuelle, les compétences et les méthodes qui permettraient d'analyser avec efficacité les volumes de données qui en résultent font défaut, ce qui empêche d'en déduire des connaissances effectivement applicables. De plus, l'enregistrement et l'utilisation des données de l'éducation se rapportant à des personnes soulèvent une série de questions juridiques et éthiques complexes s'agissant de la transparence, du consentement, des possibilités de faire opposition, et des possibilités de contrôle qui ne sont pas définitivement clarifiées jusqu'à présent (Mitros et al. 2017 ; Krüger & Lischka 2018).

6.2 Potentiels

Faire appel de manière intensive aux données granulaires sur tous les participants au système éducatif permet de développer continuellement et d'optimiser deux tendances qui ont considérablement gagné en importance au cours de ces dernières décennies dans le système éducatif : la tendance à l'individualisation et la tendance au recours aux preuves empiriques. D'une part, l'exploitation intensive de ces données permet une **individualisation** beaucoup plus large des relations avec les différents acteurs, par ex. en réaction à des problèmes et défis spécifiques que ce qui était possible auparavant dans de vastes domaines, notamment dans le déroulement de l'enseignement. Ainsi, par ex., des concepts d'enseignement individualisé et centré sur l'élève ont certes été au cœur des débats pédagogiques depuis des décennies (Bohl, Bönsch, Trautmann, & Wischer 2012). En revanche, il arrive souvent que leur réalisation concrète échoue en raison des exigences élevées posées aux apprenants et aux enseignants (Vock & Gronostaj 2017). D'autre part, ce recours intensif aux données permet un développement continu et systématique de **mécanismes de prise de décisions fondés sur des preuves empiriques**. Il est impressionnant de constater l'importance croissante des preuves empiriques en tant que point de départ des décisions administratives et politiques au vu des nombreuses réformes de l'éducation basées sur des données enregistrées (*Standard-based Reforms*) de ces 30 dernières années (Volante 2018 ; 2012).

La tendance à l'**individualisation** décrit une adaptation des méthodes de pilotage, du mode de répartition des ressources, et des mécanismes de contrôle à la situation et aux conditions préalables propres à un acteur individuel donné. Le but poursuivi est de permettre à cet acteur de se développer au mieux afin d'atteindre un objectif bien défini. À cet égard, les développements les plus avancés concernent le domaine de l'apprentissage personnalisé, c.-à-d. l'adaptation de contenus et de méthodes d'apprentissage aux intérêts, aux besoins et aux capacités de performance de chaque apprenant (Office of Educational Technology, 2017). L'apprentissage personnalisé donne à chaque apprenant « la chance de développer pleinement son potentiel sensorimoteur, intellectuel, émotionnel et social et de lui apporter un soutien via des mesures [individuelles] appropriées » (Eckert 2014). Des approches qui sont basées sur des analyses de données, à savoir sur des « systèmes d'apprentissage adaptatif », apportent une solution qui permet d'éviter la charge élevée en ressources qui est à l'origine de l'applicabilité limitée de cette méthode (Pugliese 2016).²²⁶ Des systèmes automatisés sont en mesure d'estimer le niveau de performance actuel d'un apprenant à partir de ses données historiques, de lui donner des devoirs adaptés à ce niveau de performance, d'évaluer les réponses de l'apprenant à ces devoirs, de lui donner un feedback immédiat et, sur la base de son comportement en matière de réponses, d'adapter la complexité des devoirs à son niveau de prestation (Educause 2017 ; Siemens 2013). Ces systèmes automatisés « imitent » le comportement d'enseignants humains en essayant d'amener progressivement les apprenants à maîtriser de nouvelles capacités, et en les plaçant dans une situation où ils acquièrent les contenus d'apprentissage à un rythme optimal pour eux. Une étude d'ensemble actuelle du Joint Research Centre (JRC) de l'Union européenne (Ferguson et al. 2016) fournit un aperçu complet de l'état de la technique, de l'implémentation et de l'application de ces systèmes d'apprentissage dans le domaine de la formation générale et professionnelle. Cette étude montre qu'on estime généralement que le potentiel de cette technologie est élevé, et qu'elle est utilisée dans certains cas à grand échelle. Toutefois, des évaluations complètes font encore défaut. Cependant, dans des études individuelles portant sur un petit échantillon de personnes testées, on a pu montrer de manière répétée que des systèmes de ce genre étaient comparables à des tuteurs humains du point de vue de leur efficacité (VanLehn 2011).

Simultanément, les informations générées par des systèmes d'apprentissage permettent d'évaluer rapidement et de manière complète le niveau des connaissances et les progrès de chaque apprenant dans une classe. Ils permettent ainsi aux enseignants de disposer en tout temps d'une vue d'ensemble actualisée du niveau de prestation de leur classe et de son évolution dans le temps. En outre, sur la base de ces informations, on peut identifier les apprenants qui sont distancés par rapport à la majorité de la classe, et ce, sans devoir attendre les résultats des évaluations sommatives des prestations. Ils permettent donc d'élaborer beaucoup plus tôt que ce qui est possible aujourd'hui des interventions spécifiques pour ce type d'élèves. À l'inverse, on peut aussi identifier des contenus d'apprentissage qui posent problème à une majorité des apprenants. Les enseignants peuvent donc réagir immédiatement à d'éventuels problèmes. Ils peuvent par ex. adapter les contenus ou les techniques d'enseignement au moment même où ils sont dispensés, ils peuvent immédiatement répéter la matière, ou peuvent procéder à un remaniement du matériel de cours dont l'efficacité est contrôlable.

Par le biais de ces systèmes d'apprentissage, des informations sur les prestations des apprenants individuels peuvent être aussi rendues accessibles à d'autres acteurs impliqués, par ex. les parents ou l'administration de l'école et, last but not least, également aux apprenants eux-mêmes, ce qui permet d'informer ainsi très rapidement et de manière plus complète tous les groupes cibles. En règle générale, un feedback personnalisé est proposé sous la forme de tableaux de bord (*Dashboards*) qui préparent visuellement les informations et les mettent à disposition sous une forme numérisée (la plupart du temps via un site web) (Verbert, Duval, Klerkx, Govaerts, & Santos 2013).

Puisque les données d'utilisation fournissent des informations en tout temps sur le niveau de connaissances de chaque apprenant, elles permettent en outre une plus forte pondération des évaluations formatives (*Formative Assessments*) dans le cadre de l'évaluation des compétences. En lieu et place d'une note de bulletin scolaire se présentant actuellement sous une forme globale (par ex. pour les mathématiques), on pourrait notamment attribuer des notes pour des composantes individuelles de ces compétences (par

²²⁶ Dans un contexte d'enseignement classique, le soutien individuel à un élève dispensé par l'enseignant nécessite que tous les autres élèves doivent travailler de manière autonome afin d'atteindre l'objectif poursuivi. Le soutien individualisé ne peut donc être accordé que ponctuellement à certains élèves seulement.

ex. pour maîtriser les dérivées de fonctions simples). Parce que les systèmes d'apprentissage adaptatif répartissent les contenus des cours en composantes individuelles des compétences (Liu & Koedinger 2017), ces notes peuvent alors être attribuées sous la forme de « Micro-Credentials » ou de « Badges » (Office of Educational Technology 2017). De cette façon, des énoncés peuvent être formulés sur le pourcentage de composantes qu'un élève aura acquis au cours d'une période déterminée. Si ces composantes sont standardisées, des *Micro-Credentials* permettent en outre de saisir facilement de manière détaillée le niveau d'apprentissage de l'apprenant et de procéder à une comparaison des niveaux d'apprentissage des apprenants sans devoir pour autant leur faire passer des tests explicitement standardisés à cet effet. Non seulement cela simplifie le monitoring du système éducatif, mais cela réduit en outre le temps à y consacrer et le stress lié à ces tests pour les élèves – en dernière analyse, il s'agit en effet d'un temps improductif qui ne sert pas à faire progresser l'apprentissage.²²⁷ De plus, la standardisation et la portabilité des « badges » permettent d'évaluer avec précision les capacités des apprenants en cas de changement d'école et facilite ainsi leur intégration dans une nouvelle communauté de classes. Enfin, la combinaison entre des données provenant de traces d'apprentissage et des données issues de traces de comportement avec d'autres informations externes offre la possibilité de développer des modèles prédictifs pour certains événements (par ex. concernant l'interruption de l'école ou le résultat de certaines activités post-scolaires sélectionnées) qui permettent de procéder à une intervention précoce ou de dispenser des conseils couronnés de succès (Baer & Norris 2017).

L'individualisation basée sur des données du déroulement de l'enseignement promet toute une série d'avantages, également au niveau de l'école et de l'administration de l'école (New 2016 ; Baer & Norris 2017). D'une part, l'individualisation peut contribuer à décharger les enseignants de certains travaux consistant à assumer une série de tâches répétitives (par ex. corrections de devoirs d'exercices) et administratives (par ex. la communication de notes et d'évaluations de prestations à l'attention de la direction et de l'administration de l'école). Elle donne ainsi aux enseignants le temps et les ressources nécessaires qui pourront être mis à profit de manière productive pour la structuration de l'enseignement ou pour l'encadrement ciblé de certains apprenants.

Pour le système scolaire dans son ensemble, le recours aux données d'apprentissage individuelles promet, d'une part, une amélioration de la motivation d'apprentissage et des prestations d'apprentissage (Tomasik, Berger, & Moser 2018) et, de ce fait, de l'efficacité et de l'efficience du système. D'autre part, les informations sur le comportement individuel d'apprentissage et l'acquisition des compétences offrent de nouvelles possibilités d'améliorer l'appariement entre les apprenants et les offres de formation, ainsi qu'entre les personnes qui ont terminé leurs études et les emplois disponibles (Manyika et al. 2013). Elles permettent aux apprenants, aux parents et aux écoles de mieux évaluer où se situent des lacunes, et selon quelle ampleur, entre les compétences d'un apprenant et les exigences du cursus scolaire ou de la formation qu'il vise. Parallèlement, sur la base de ces données, ces informations permettent de comparer des énoncés prédictifs sur les perspectives de succès d'un apprenant au-delà des différentes options de formation et, dès lors, de leur dispenser de meilleurs conseils en matière de choix de profession et d'orientation de carrière. Enfin, à partir de l'agrégation de données d'apprentissage individuelles, il est possible de se procurer plus rapidement et de manière plus sûre de nombreux indicateurs du monitoring de la formation axés sur les élèves (par ex. le pourcentage des apprenants qui passent de la scolarité obligatoire à la formation postobligatoire) (Cope & Kalantzis 2015 ; Mitros et al. 2017 ; Wolter et al. 2018).²²⁸

Toutefois, les véritables avantages de l'enregistrement et de l'évaluation des données d'apprentissage individuelles pour la gestion et l'administration scolaire résident dans le fait qu'ils permettent de prendre une multitude de décisions sur la base d'évaluations basées sur des preuves empiriques. **Les processus de décision fondés sur des preuves empiriques** désignent l'évaluation de propositions et concepts de solutions concurrents effectuée sur la base de méthodes expérimentales et scientifiques. Outre des études de contrôle randomisées, ces dernières incluent aussi notamment des « méta-analyses », à savoir l'éva-

227 Une critique formulée de manière répétée à l'encontre du système des tests standardisés réside dans le fait qu'en raison de l'importance qui leur est attribuée, ils renforcent encore la pression déjà élevée qui s'exerce sur les élèves. Ils contribuent ainsi à la multiplication croissante des maladies dues au stress chez les apprenants (Abeles & Rubenstein 2016 ; Simpson 2016). Cette critique s'est exprimée de manière particulièrement frappante dans une lettre ouverte de chercheurs et praticiens en éducation adressée au directeur de l'OCDE pour l'éducation (Meyer & Zahedi 2014).

228 De manière plus sûre notamment lorsque, grâce à l'utilisation des données d'apprentissage, on possède des informations sur la totalité des élèves au lieu de ne disposer, comme jusqu'à présent – par exemple pour mesurer les compétences – que d'un échantillon d'élèves.

luation systématique et la synthèse des résultats de nombreux études (expérimentales) individuelles portant sur le même thème (Borenstein, Hedges, Higgins, & Rothstein 2009). Les résultats de ces analyses permettent donc d'établir un classement (*Ranking*) de concepts de solutions possibles dans la perspective de leur probabilité ou de leur degré d'efficacité (à savoir de leur effectivité) ainsi que de l'efficacité par type de ressource utilisée (à savoir de leur efficence). Ils mettent ainsi à disposition une base de décision intersubjective facilement compréhensible qui tire sa justification d'observations menées scientifiquement dans le monde réel. Les principes de tels processus de prise de décision sont déjà bien ancrés aujourd'hui dans de larges domaines du système éducatif. Ils constituent le fondement argumentatif pour l'implémentation de contrôles standardisés des compétences au niveau national et international tels que les enquêtes PISA ou les enquêtes COFO (Volante 2012 ; 2018).

Les idées et les résultats de la recherche en éducation basée sur des preuves empiriques ont été communiquées à un large public également par le biais des discussions menées autour des travaux de John Hattie (2008 ; 2012) et du groupe de travail autour de Robert Marzano (Dean, Hubbell, Pitler, & Stone 2012) dans le domaine du développement de l'enseignement. Ces enquêtes regroupent les résultats d'une multitude d'études individuelles et de méta-analyses. L'objectif est de pouvoir formuler des énoncés empiriquement fondés sur l'efficacité de différentes méthodes d'apprentissage pour le succès de l'apprentissage, et de donner ainsi aux enseignants des modes d'emploi concrets d'action afin d'optimiser le déroulement de l'enseignement.

Selon ces experts, le potentiel d'application des méthodes expérimentales peut être nettement accru par une saisie et une collecte systématiques de données granulaires sur l'apprentissage et l'enseignement. D'une part, parce que cela élargirait considérablement la base des données sur lesquelles reposent de telles enquêtes. Les enseignants et les directions des écoles pourraient, par exemple, en cas de problèmes concrets, rechercher des situations comparables dans des banques de données et pourraient reprendre chaque fois la meilleure solution. De surcroît, une telle collecte de données faciliterait l'implémentation et la mise en œuvre de petites études de contrôle au niveau des classes ou des écoles. Des questions concrètes, par ex. sur l'efficacité d'un nouveau moyen d'enseignement, pourraient être analysées systématiquement en très peu de temps. Aux États-Unis, l'Office of Educational Technology a déjà développé un tel outil (<https://edtechrce.org/>). Il est déjà diversement utilisé en de nombreux endroits par les écoles et l'administration des écoles (voir cas d'application 6, p. 121). Compte tenu de l'étendue de la collecte d'informations effectuée sur le comportement d'apprentissage et sur les résultats d'apprentissage de chaque apprenant, il a même été possible de déduire des conclusions de ce genre au niveau individuel. À partir de données sur l'historique d'apprentissage, des énoncés spécifiques à des personnes pourraient être formulés pour chaque apprenant en précisant quelle méthode d'apprentissage pourrait aider à résoudre quel type de problème avec le taux de probabilité le plus élevé (Kolowich 2013).

Également au niveau de l'école ou de l'administration scolaire, une collecte croissante et plus étendue de données qui sont générées à chaque niveau du système éducatif contribue à améliorer la prise de décisions en simplifiant et en accélérant notamment le monitoring (Mitros et al. 2017). Au lieu de procéder, par ex., comme aujourd'hui, à des enquêtes standardisées à des intervalles pluriannuels (et avec un retard considérable lors de la publication des résultats) pour pouvoir saisir le niveau de compétences des apprenants, une telle information pourrait être obtenue à n'importe quel moment en consultant une banque de données. Une telle observation en temps réel permettrait à l'administration de l'éducation d'identifier très rapidement des problèmes et, par exemple, de réagir par un changement de répartition des ressources au sein du système. Dans le cas le plus simple, on pourrait par ex. étudier les effets de certaines applications d'apprentissage concurrentes sur les prestations d'apprentissage. De cette façon, on pourrait par exemple évaluer si une utilisation de cette application dans l'ensemble du système scolaire serait appropriée, ou si des variantes plus coûteuses amélioreraient le niveau des prestations des élèves de manière suffisamment élevée pour justifier son choix (New 2016). De surcroît, des collectes de données étendues conviennent à l'analyse des effets de certaines caractéristiques administratives et de certaines décisions. Par exemple pour la planification des cours, l'élaboration de l'offre en matière d'activités extrascolaires, ou encore l'étude des effets sur les prestations des élèves de l'ampleur des travaux de coopération et de collaboration au sein du corps enseignant en tant qu'organe collégial (Gräsel 2015) ainsi que pour l'adaptation correspondante des lignes directrices (New 2016).

L'affinement constant et l'accélération du cycle d'enregistrement, d'analyse et de feed-back des données qui est rendu possible par les outils numériques permettent d'adapter, dans une ampleur encore jamais connue jusqu'ici, les méthodes de pilotage, les ressources et les contrôles à chaque acteur individuel. Ils offrent une base scientifiquement fondée pour des prises de décisions à chacun des niveaux du système éducatif. C'est pourquoi Williamson (2016) parle d'un changement de paradigme tiré par la technologie dans la politique de l'éducation : on abandonne ainsi un pilotage plutôt lent basé sur des changements de conditions cadres au profit d'un contrôle et d'une mise au courant de l'acteur individuel en temps réel.

Une base de preuves empiriques toujours croissante pour des prises de décisions dans le système éducatif peut aussi déployer des effets démocratisants. Et ce, lorsqu'un accès facile aux données et aux résultats est garanti, et que les écoles et les enseignants qui ne disposent pas de compétences et de ressources pour évaluer leurs propres données peuvent faire appel aux expériences et aux bonnes pratiques (*Best Practices*) d'autres acteurs (New 2016). De manière générale, on attribue un potentiel considérable à un accès ouvert aux données pertinentes de l'éducation dans le cadre de l'exploitation de nouvelles possibilités de participation et de discussion (voir chapitre 4.3).

6.3 Cas d'application

À l'heure actuelle, il existe un écart considérable entre le potentiel attribué à l'utilisation des données pour le développement de l'enseignement et du système éducatif et l'utilisation effective à grande échelle de ces concepts dans la vie quotidienne de l'école et de l'administration scolaire. Certes, le nombre d'études scientifiques publiées sur ce thème ces dernières années a fortement augmenté. Mais elles se limitent presque exclusivement aux dépenses effectuées dans le domaine de la formation postobligatoire des États-Unis, de la Grande-Bretagne et de l'Australie. En outre, compte tenu de la nouveauté de nombreuses technologies, ces enquêtes ne fournissent souvent aucun renseignement sur les conséquences à long terme de leur utilisation. Une étude d'ensemble actuelle sur l'utilisation et le recours au « Learning Analytics » parvient donc à la conclusion selon laquelle

« there is no overwhelming evidence that learning analytics have fostered more effective and efficient learning processes and organisations. However, there is convincing evidence [...] that companies and organisations believe learning analytics will do this in the future, and are prepared to invest time and resources [...] in them. » (Ferguson et al. 2016, p. 25).

Néanmoins, on rencontre quelques exemples d'applications internationaux et nationaux qui peuvent servir d'inspiration à une utilisation à long terme couronnée de succès des données de l'éducation afin de développer l'enseignement, l'école et le système éducatif. Parmi ces exemples, nous vous en présentons six dans les lignes qui suivent :

Cas d'application 1 : développement de l'enseignement – implémentation d'un système d'apprentissage adaptatif pour des cours d'appui en mathématiques

Situation initiale

Avec plus de 80 000 étudiantes et étudiants inscrits, l'Arizona State University est la plus grande université publique des États-Unis. L'un des défis à relever découlant de cette grande taille réside dans les écarts énormes qui existent entre les nouveaux étudiants en raison de leur formation antérieure et de leur niveau de connaissances au moment de leur immatriculation. En particulier dans le domaine des mathématiques, près de 30 % des étudiants de chaque volée annuelle ont des connaissances lacunaires à insuffisantes dans ce domaine. Afin de compenser ces écarts, des cours d'appui en mathématiques sont proposés. Dans ce contexte, on a constaté un taux de corrélation marqué entre la survenance d'interruptions dans ces cours et la probabilité que les étudiants interrompent prématurément leurs études. Les étudiants qui terminaient ces cours d'appui avec une mauvaise note avaient un taux de probabilité 50 % plus élevé d'interrompre leurs études sans décrocher de diplôme. L'amélioration de l'enseignement des mathématiques et des compétences en la matière était donc un objectif important de l'administration de cette haute école.

Méthode

En août 2011, en collaboration avec Knewton, aussi bien des cours en présentiel que des cours en ligne ont été convertis et transférés sur un nouveau système d'apprentissage adaptatif basé sur des données. Pour ce faire, tous les contenus d'apprentissage ont été subdivisés en « composantes de compétences », et les cours ont été structurés selon le principe de la « Flipped Classroom ». Cela signifie que les apprenants visualisent des vidéos d'apprentissage (à titre de devoirs à domicile) sur différents domaines thématiques avant le moment où le cours sera dispensé. Pendant le cours, les étudiants traiteront les devoirs d'exercices assistés par ordinateur qui ont été conçus sur mesure pour répondre à leur niveau de compétences spécifique. Leurs progrès sont suivis et enregistrés, et de nouvelles tâches et domaines thématiques sont automatiquement adaptés à ce niveau. Un feedback individualisé est donné à propos de chaque tâche ou devoir. Les enseignants n'enseignent plus *ex cathedra*, mais répondent au contraire de manière ciblée aux questions des étudiants. En outre, un « cockpit » numérique permet de suivre l'évolution de chaque étudiante et de chaque étudiant et d'apporter ainsi un soutien ciblé à celles et ceux qui ont le plus de problèmes.

Résultat

Dans les deux semestres après l'introduction de ce nouveau système, le nombre des étudiants qui interrompaient le cours d'appui de mathématiques a reculé de 13 % à 6 %, et le pourcentage de ceux qui terminaient ce cours avec succès a progressé de 66 % à 75 %. De surcroît, plus de la moitié de tous les participants ont atteint les objectifs de cours visés quatre semaines avant la fin du semestre. En conséquence, ce système a été étendu à d'autres cours en économie politique, physique, chimie et statistique.

Situation en Suisse

En Suisse également, il existe une série de systèmes d'apprentissage adaptatif destinés à l'enseignement. À l'heure actuelle, ils sont utilisés à tous les niveaux du système éducatif. Par exemple, le Center for Active Learning du Département de biologie de l'EPFZ à Zurich offre plusieurs cours qui sont composés de cours en présentiel et de cours en ligne (disponibles en partie en mode « Flipped Classroom »). En règle générale, l'évolution dans le temps du niveau de compétences est mesurée de manière numérique par des tests de type « Multiple-Choice Questions ». Les résultats sont mis immédiatement à disposition des enseignants. Selon le site web du département, cette forme d'enseignement a rencontré un grand écho auprès des enseignants et des étudiants. Au niveau de la scolarité obligatoire, depuis l'année scolaire 2017/2018, la plateforme d'apprentissage adaptatif [Mindsteps](#) développée par l'Institut pour l'évaluation de la formation de l'Université de Zurich est utilisée partout dans les cantons de l'espace suisse de formation du Nord-Ouest. À l'heure actuelle, cette plateforme couvre les compétences pour quatre matières scolaires (allemand, anglais, français et mathématiques) allant de la 3^e classe à la 9^e classe (Tomasik, Berger, & Moser 2018). Jusqu'à présent, on n'a pas procédé à une évaluation de la plateforme. Des systèmes adaptatifs similaires sont mis aussi à disposition par la maison d'édition de matériel pédagogique « Lehrmittelverlag St. Gallen » pour le niveau primaire ([Lernlupe](#)) et pour le niveau secondaire ([Lernpass plus](#)) dans les branches allemand, anglais, français et mathématiques. Ils permettent de mettre à disposition des apprenants des devoirs qui sont adaptés à leurs prestations dans le cadre de mesures standardisées des compétences. Ils rendent ainsi possible la combinaison de l'encouragement individuel avec une évaluation du niveau d'apprentissage et des progrès d'apprentissage des classes et des volées annuelles. Depuis l'année scolaire 2018/2019, ils sont mis à disposition de toutes les écoles dans le canton de St-Gall, étant précisé que les écoles des cantons de Zurich et d'Appenzell Rhodes-Extérieures peuvent aussi recourir à ces systèmes ou à certains de leurs composants (par ex. aux mesures standardisées des compétences). Pendant une phase pilote, ils ont été utilisés dans douze écoles primaires et dans huit écoles secondaires. Jusqu'à présent, une évaluation des effets de ces systèmes sur le taux de satisfaction des apprenants et des enseignants n'a pas non plus été réalisée dans ce contexte.

Cas d'application 2 : développement de l'école – introduction d'un système de recommandations automatique pour la sélection ciblée de cours pendant les études

Situation initiale

En règle générale, le début des études universitaires est synonyme d'importants défis à relever pour les nouveaux étudiants. Ils doivent trouver leurs marques dans un nouveau système qui exige nettement plus d'autonomie et de motivation de leur part que le système scolaire. Au vu de la multitude d'options et de possibilités de choix qui satisfont toutes aux exigences du cursus des études, choisir concrètement les séminaires et cours individuels adéquats constitue un obstacle de taille. Très souvent, il n'est pas clair, par ex., de savoir quelle combinaison correspond particulièrement bien à ses propres inclinations ou conduit de manière particulièrement efficace au diplôme de fin d'études. L'ennui, la sollicitation excessive ou le manque d'intérêt peuvent virer à la frustration et, pour finir, peuvent entraîner l'interruption des études. C'était l'un des problèmes rencontrés par l'Austin Peay State University (APSU) à Clarksville (Tennessee).

Méthode

Inspirée par le système de recommandations réalisé par des entreprises comme Netflix ou Amazon, l'université a développé un système de recommandations des cours baptisé « Degree Compass ». Il a pour but d'apparier aux mieux les étudiants et les cours. « Au mieux » signifie en l'occurrence qu'ils correspondent aussi bien aux points forts de l'étudiant qu'aux exigences du cycle de formation. À cet effet, un algorithme prédictif combine des informations historiques sur le comportement en matière de choix et des informations sur les notes de tous les étudiants précédents avec le comportement actuel en matière de choix et avec les notes actuelles de l'étudiant concerné. Les cours qui ont été suivis avec succès par des étudiants dont le profil historique en matière de haute école était similaire, par le passé, à celui de l'étudiant concerné ont plutôt tendance à faire l'objet d'une recommandation. Les exigences posées aux études sont prises en compte. Avec cette méthode, le système donne ses recommandations les plus fortes aux cours qui sont nécessaires pour terminer une formation, dont l'importance est décisive pour le cursus et dont l'algorithme prédit qu'ils seront suivis avec succès.

Résultat

Le facteur de succès le plus important propre au système de recommandations réside dans sa capacité de répartir le temps d'études des étudiants dans des cours où ils ont le maximum de succès. Des sondages montrent que tant les enseignants que les étudiants se sont en principe félicités de l'existence de ces informations supplémentaires, et qu'après une courte période d'initiation, ils ont pu interagir avec compétence avec l'interface utilisateurs programmée.

L'algorithme fournit une prédiction comparativement précise de la note finale qui sera obtenue à l'issue d'un cours choisi. Une comparaison rétroactive des prédictions du modèle avec des notes réelles a fait état d'une concordance de 90 %. De surcroît, lors de la comparaison des prédictions avec les notes réelles, il s'est avéré que les étudiants qui ont choisi les cours recommandés par le système ont obtenu des notes qui étaient en moyenne supérieures de 0,46 points à celles des étudiants qui n'avaient pas suivi la recommandation.

Situation en Suisse

Selon nos informations, un système de recommandations comparable basé sur les prestations passées n'existe pas en Suisse. Toutefois, une série d'entreprises et d'institutions mettent à disposition des plateformes de tests afin de réaliser des tests standardisés, à savoir des tests d'intérêts, des tests de personnalité et des tests de prestations. Ces tests servent majoritairement à évaluer si les apprenants et les métiers pour lesquels ils se forment sont concordants, et permettent de leur dispenser des conseils d'orientation professionnelle empiriquement et psychologiquement fondés. Nous renvoyons ici, par exemple, à la plateforme de tests en ligne [OTP](#) destinée aux conseillers en orientation du Centre suisse de services Formation professionnelle / orientation professionnelle, universitaire et de carrière ([CSFO](#)) ou aux offres de l'entreprise [gateway.one](#). En règle générale, les outils proposés sont dûment fondés sur le plan théorique et ont été empiriquement évalués. Leur utilisation a gagné en importance au cours de ces dernières années. Ainsi, en 2017, plus de 30 000 tests ont été réalisés sur OTP (CSFO 2018) et près de 50 000 tests proposés ont été effectués par le biais de [gateway.one](#). Dans le cadre de son initiative « [Campus Analytics](#) », le Centre pour l'éducation à l'ère digitale (CEDE) de l'EPFL prévoit notamment l'implémentation d'un outil de planification de cours et de carrière guidé par les données.

Cas d'application 3 : développement du système – coordination des plans d'études entre les écoles et les hautes écoles

Situation initiale

Le programme « Graduate New York City (GNYC) » est un projet commun développé par la City University of New York (CUNY) et par le New York City Department of Education (DoE). Il repose sur le fait reconnu qu'une grande part des étudiants de la CUNY (74 % d'entre eux) proviennent d'écoles qui sont gérées par le DoE. Simultanément, près de 60 % des anciens élèves du DoE qui visent une formation tertiaire choisissent d'étudier à la CUNY. Ces deux établissements gèrent ainsi à peu près les mêmes effectifs d'apprenants. Mais malgré ces recoupements considérables, les échanges entre ces deux institutions sont restés longtemps limités. Du point de vue des écoles, ce qui était frappant, c'est qu'un nombre comparativement restreint d'apprenants parvenaient à être admis dans la haute école. De l'avis de la haute école, les compétences de la plupart des élèves qui avaient fini leur scolarité étaient lacunaires, ce qui se traduisait par un faible de taux réussite au diplôme de fin d'études universitaires.

Méthode

Le travail du GNYC, fondé en 2010 grâce au financement de la Bill & Melinda Gates Foundation, se fonde sur l'hypothèse selon laquelle un changement significatif des taux de réussite locaux au diplôme d'études secondaires et au diplôme de fin d'études universitaires nécessite une étroite coordination entre les établissements éducatifs de la ville.

Un élément central de la collaboration entre les deux institutions consistait à développer un processus automatisé ayant pour but d'harmoniser et d'échanger des données sur les apprenants et à créer une banque de données commune afin de mémoriser ces informations. Cette banque de données permet aux écoles de suivre l'évolution universitaire ultérieure de leurs anciens élèves et d'identifier les problèmes qui surviennent le plus fréquemment. En outre, ce système a permis aux différentes écoles de comparer entre elles les compétences des élèves ayant terminé leur scolarité et de définir ainsi des exemples de meilleures pratiques (*Best Practices*).

Les écarts considérables qui ont été mis en évidence ont donné lieu, en 2012 et 2013, à une plus forte standardisation des plans d'études (Curriculum Alignment Project) où des professeures et professeurs de la CUNY ont rencontré des enseignants et spécialistes du DoE afin de développer des standards communs pour l'enseignement de l'anglais et des mathématiques.

Résultat

Les résultats d'une étude de 2014 montrent que pendant la période qui s'est écoulée depuis le début du projet, on constate clairement la présence d'une série de changements positifs dans le système scolaire (Graduate NYC 2014) :

- Le taux de réussite au diplôme des écoles secondaires (*High Schools*) est passé de 47 % en 2005 à 65 % en 2012.
- Entre 2002 et 2012, le nombre d'étudiants de la CUNY du niveau bachelor qui avaient obtenu un diplôme d'une école gérée par le DoE a progressé de 61 % depuis 2002 (par comparaison avec une progression générale de 32 % de l'effectif des étudiants).
- Pendant la même période, le nombre des étudiants en bachelor de la CUNY ayant étudié par le passé dans une école gérée par le DoE qui ont interrompu leurs cours ou leurs études a reculé de 14,8 % à 11,5 % (interruption des études), resp. de 24,2 % à 19,4 % (cours interrompu ou cours non réussi).

Situation en Suisse

En Suisse également, depuis le milieu des années 1990, différents projets ont été lancés qui avaient pour but de faciliter le passage du gymnase vers la haute école, et de garantir la capacité générale de faire des études universitaires pour les titulaires de la maturité fédérale. Ainsi, par exemple, du côté de la CDIP, une série de recommandations correspondantes ont été élaborées (CDIP 2016). Elles contiennent la définition

des compétences techniques de base²²⁹, une harmonisation entre les examens de maturité cantonaux, ainsi que l'implémentation d'un échange institutionnalisé entre les écoles moyennes et les hautes écoles. Avec le projet [HSGYM](#), un travail similaire bien établi a été réalisé depuis 2006 également dans le canton de Zurich (HSGYM 2014). Ici aussi, la collaboration entre les enseignants et les professeurs a été institutionnalisée par le biais de groupes techniques et de pools d'experts communs. En outre, on offre la possibilité aux enseignants de suivre un semestre de recherche libre dans l'une des hautes écoles zurichoises (Université de Zurich ou EPFZ). Toutefois, un échange systématique des données sur les apprenants entre les écoles moyennes et les hautes écoles n'a pas lieu dans le cadre de ces projets. Une évaluation des résultats n'a pas été réalisée non plus jusqu'à présent, à notre connaissance.

Cas d'application 4 : développement du système – implémentation de standards aux fins d'exploitation de sources de données existantes

Situation initiale

Le nombre de technologies et de programmes différents qui sont utilisés pour saisir différentes données constitue un problème toujours plus important propre à la collecte (électronique) et à l'utilisation des données dans le secteur éducatif. Ainsi, très souvent, des notes scolaires, des absences, des commentaires, etc. – même au sein d'une seule et même école – sont saisis et mémorisés dans des systèmes différents. En outre, la plupart du temps, différentes écoles utilisent d'autres systèmes pour saisir la même information. Dès lors, il en résulte un grand nombre de solutions et systèmes locaux avec lesquels on peut codifier, mémoriser et consulter des données. Cela accroît la charge de travail administratif et le risque d'erreurs propre au système, et rend simultanément difficile le regroupement de données émanant de sources différentes. Cela concerne par ex. les enseignants et la direction de l'école au sein d'une même école (par ex. lorsqu'il s'agit d'ajuster entre elles les notes et les absences), mais cela touche aussi les administrations scolaires locales et régionales (lorsqu'il s'agit de procéder à des comparaisons de plusieurs écoles entre elles).

Méthode

En réaction à cette problématique, plusieurs États fédéraux américains ont lancé des standards destinés au système éducatif. Ces standards constituent un ensemble de règles de portée globale qui définissent comment il faut saisir des entrées de données dans des systèmes déterminés. Ils permettent ainsi de relier ensemble des données provenant de sources qui n'étaient pas (ou seulement très difficilement) reliables auparavant. Un de ces ensembles de règles de portée globale est mis à disposition, avec le soutien financier de la Michael & Susan Dell Foundation, de l'ONG Ed-Fi qui se charge de leur gestion. Cet ensemble de règles de portée globale a été notamment repris et adapté par les États fédéraux américains de l'Arizona, du Michigan et du Wisconsin.

L'objectif clé poursuivi dans chaque cas consistait à garantir l'accès aux données portant sur l'ensemble des sources, et ce, pour tous les acteurs impliqués du système éducatif (notamment les enseignants et l'administration scolaire). Il s'agissait de réduire ainsi la charge de travail administratif (pour l'administration scolaire et les enseignants), de faciliter l'adaptation de l'enseignement aux besoins des apprenants (enseignants), et de simplifier la collecte d'information pour le monitoring de la formation (administration scolaire et enseignants).

Résultat

Dans toutes les études de cas mises à disposition par l'ONG Ed-Fi, on mentionne la facilité, la rapidité et le caractère global de la mise à disposition des informations (Ed-Fi Alliance 2018).

Le Département de l'éducation de l'Arizona (état : 2015) s'attend à des économies de 57 millions de dollars grâce aux gains d'efficacité liés à l'introduction de ce système (Ed-Fi Alliance 2015). En outre, le temps nécessaire pour calculer les statistiques sur la formation à l'échelle de l'ensemble de l'État fédéral de l'Arizona a été réduit de cinq jours à une heure (Arizona Department of Education 2018).

²²⁹ Les compétences techniques de base désignent les connaissances et le savoir-faire dans la première langue parlée et en mathématiques qui constituent des prérequis pour de nombreux cursus universitaires et dont la maîtrise est par conséquent nécessaire pour être admis avec succès dans de nombreux cycles d'études universitaires (CDIP 2016). Ces compétences sont établies sur la base d'enquêtes et d'interviews réalisées avec des étudiants universitaires de différentes orientations spécialisées et sur la base d'une analyse du matériel de cours de la première année d'études (Eberle, Brüggengbrock, Rüede, Weber, & Albrecht 2015).

Situation en Suisse

Un standard de portée globale applicable aux données de l'éducation en Suisse n'existe pas jusqu'à présent (voir chapitre 4.4). Toutefois, avec l'association eCH, qui a pour mandat d'élaborer des standards nationaux d'*e-government* pour la Suisse (Conseil fédéral et CdC 2015), les conditions organisationnelles préalables requises ont été instaurées pour leur élaboration. Même s'il n'existe pas encore jusqu'à présent de groupe technique consacré à l'étude du domaine thématique « Formation », quelques standards existent déjà, par exemple pour l'échange d'identifications de personnes (Steimer & Stingelin 2014) qui pourraient être utilisés dans un futur standard relatif aux données de l'éducation. En outre, dans le cadre de ses enquêtes, l'OFS définit des standards de données (par exemple dans le [Manuel technique Statistique des diplômés des formations institutionnalisées](#)) qui peuvent servir de base à l'élaboration d'un tel standard.

Cas d'application 5 : développement du système – collecte de données aux fins d'amélioration du financement des hautes écoles (système d'information suisse sur les hautes écoles)

Situation initiale

Le système des hautes écoles en Suisse est complexe. Outre les hautes écoles fédérales soutenues par la Confédération, il existe des hautes écoles privées ainsi que des universités, des hautes écoles spécialisées et des hautes écoles pédagogiques dont le financement est, pour l'essentiel, du ressort des cantons (Schmidt 2008). Les cantons qui ne possèdent pas de hautes écoles participent au financement de certaines institutions cantonales dans le cadre de conventions intercantionales. Le soutien est accordé sous la forme d'un « montant forfaitaire par étudiant » pour les étudiants provenant de cantons qui ne possèdent pas de hautes écoles, étant précisé que c'est le domicile légal au moment de l'admission dans la haute école qui est considéré comme le critère d'appartenance au canton (accord intercantonal universitaire [AIU] 1997). Dès lors, un système d'attribution objectif des étudiants aux cantons qui est compréhensible pour tous les groupes cibles joue un rôle central dans la clé de répartition des coûts.

Méthode

Au début des années 1970, le Système d'information universitaire suisse (SIUS) a donc été conçu à cet effet en collaboration avec les universités, les cantons et la Confédération. Il saisit, selon un rythme annuel, des informations sur toutes les étudiantes et tous les étudiants qui suivent leurs études dans des hautes écoles publiques de Suisse. Des informations sont regroupées dans une banque de données de l'Office fédéral de la statistique. Outre le lieu de domicile au moment de l'immatriculation, une série de caractéristiques et informations sociodémographiques de la situation d'études sont également saisies. En outre, les étudiants peuvent être identifiés clairement par un numéro matricule et – depuis 2012/2013 – par le nouveau numéro AVS « AHVN13 » (OFS 2018c).

Résultat

Alors que le SIUS continue d'être utilisé également pour son but initial, à savoir pour enregistrer les effectifs des étudiants selon leur canton d'origine, à partir du jeu de données rassemblées depuis 1975, une multitude d'autres possibilités d'application en ont résulté. Elles permettent ainsi d'exécuter des analyses statistiques en différé (par ex. changement de haute école et changement de branche, taux de réussite, durée des études). En outre, la banque de données sert de base pour le monitoring de la formation ainsi que pour des enquêtes scientifiques sur le comportement pendant les études suivies dans les hautes écoles suisses, notamment pour les indicateurs relatifs aux hautes écoles et pour les prévisions sur la formation réalisées par l'OFS.

Cas d'application 6 : développement de l'enseignement et de l'école – outil d'évaluation rapide et systématique du matériel d'enseignement (numérique) et des programmes scolaires

Situation initiale

La quantité croissante des offres d'enseignement et d'apprentissage numériques accroît encore l'offre déjà très étendue de matériel pédagogique, et ce, dans une mesure toujours plus rapide. L'augmentation constante des possibilités de choix qui va de pair avec cette tendance permet certes un appariement toujours meilleur entre les besoins des enseignants et ceux des apprenants, mais elle place aussi les décideurs travaillant dans l'école et l'administration scolaire devant le défi important de devoir faire des choix judicieux et conduisant à l'objectif poursuivi. Et ce, également parce que, très souvent, il n'est pas clair de savoir si oui ou non, et si oui, selon quelle ampleur, du nouveau matériel pédagogique potentiellement plus onéreux apporte une valeur ajoutée effective à l'enseignement. En outre, compte tenu du manque de ressources et de connaissances disponibles, il est souvent impossible à de nombreux responsables de procéder eux-mêmes à des évaluations systématiques et scientifiquement correctes de ce matériel pédagogique.

Méthode

La plateforme Internet [Rapid Cycle Evaluation Coach](#) (RCE), développée en 2016 par l'entreprise [Policy Research](#) sur mandat de l'[Office of Educational Technology](#) (OET), a pour mission de combler ces lacunes. Elle offre un mode d'emploi facile à suivre et un accompagnement pas à pas lors de l'exécution d'expériences scientifiques et d'analyses de données, et rend ainsi les directeurs et les directrices d'écoles et les enseignants capables d'évaluer de manière autonome, rapidement et à moindres coûts du matériel pédagogique (numérique). Appelées « Rapid Cycle Evaluations », ces évaluations permettent d'identifier, sur la base d'échantillons de taille relativement petite, des écarts systématiques dans les prestations entre les apprenants avec et sans accès à une technologie ou à un programme déterminé, et de formuler ainsi des énoncés sur leur efficacité et leur efficience. En outre, le projet met à disposition des protocoles sous licence ouverte (des modes d'emploi sur des modalités de procédures méthodologiques) afin d'exécuter de telles évaluations à destination des praticiens, des développeurs et des chercheurs qui peuvent les utiliser au-delà du cadre de cette plateforme. Le but consiste, d'une part, à améliorer les compétences et possibilités d'analyse des enseignants et des responsables des écoles et, d'autre part, à mettre à disposition de meilleures mesures de l'utilité de ce matériel pédagogique afin de prendre des décisions en matière d'achats.

Résultat

Une série d'[études de cas](#) montrent qu'il est approprié de recourir à cet outil pour répondre à un large éventail de questions, et que les connaissances ainsi acquises peuvent entraîner des conséquences considérables pour l'organisation de l'enseignement et de l'école. Ainsi, par ex., la plateforme RCE a été utilisée dans les Uplift Education Charter Schools au Texas afin d'évaluer un programme de tutorat pour des apprenants ayant des capacités de lecture « faibles » à « très faibles ». À cet effet, ces apprenants ont été retirés de l'enseignement hebdomadaire habituel, chaque fois pour une période de deux heures scolaires, et ont suivi des cours dans une situation réelle et concrète (*1:1-Situation*) dispensés par des volontaires spécifiquement formés. Le résultat de l'analyse a montré que cette intervention n'a amélioré que marginalement – si tant elle qu'elle les ait améliorées – les capacités de lecture des personnes concernées. Par conséquent, le programme a été suspendu, et de nouveaux enseignants ont été engagés avec l'argent ainsi libéré (Place & Manley 2017).

Situation en Suisse

En Suisse également, il existe des instruments qui facilitent l'exploitation des données afin de développer les écoles et l'enseignement. En voici un exemple : l'offre « Evaluationscenter » de la plateforme Internet [IQES online](#), qui permet aux responsables des écoles et aux enseignants de créer et d'exécuter facilement des sondages et de les évaluer (de manière automatisée). À l'heure actuelle, cette offre est utilisée par quelque 1200 écoles en Suisse alémanique (IQES online 2018). Une évaluation externe dans le cadre d'un projet pilote quadriannuel réalisée dans cinq écoles du canton de Berne a montré

que la plateforme a été qualifiée de « très utile » et de « très positive » par les directions des écoles et par les enseignants, et qu'elle a été utilisée pour un large éventail de tâches allant de sondages réalisés auprès des parents à l'établissement de rapports destinés à l'inspectorat des écoles (Siegel & Wyler 2012). Un exemple tiré du domaine de l'évaluation du matériel d'enseignement est l'instrument [levanto](#), de la [Centrale intercantonale des moyens d'enseignement](#) (CIME). Il met en priorité à disposition des commissions cantonales du matériel scolaire un moyen de systématiser l'évaluation du matériel scolaire sur la base de critères spécifiques aux différentes branches et de critères interbranches. Il permet d'établir une synthèse visuelle des évaluations de plusieurs personnes considérées selon ces critères, et il aide ainsi à structurer les processus d'évaluation de manière plus transparente. Toutefois, les évaluations n'incluent aucune information externe (par ex. des informations sur les prestations des élèves).

6.4 Défis à relever

Les possibilités qui résultent de la numérisation croissante sont considérables pour le pilotage du système éducatif en général, et pour la conduite de son déroulement en particulier. Tirer parti au maximum de ces potentiels est un objectif stratégique de la politique de l'éducation suisse (CDIP 2018b). Pour atteindre cet objectif, il faut remplir une série de conditions préalables qui ne permettent pas seulement la collecte, la mémorisation et l'analyse des données de l'éducation d'un point de vue technique, mais qui garantissent que l'on dispose aussi de compétences et de ressources suffisantes, et que l'utilisation de ces données bénéficie de l'acceptation des groupes cibles concernés. Ces deux aspects sont décisifs pour assurer le bon fonctionnement d'un système de prise de décision assisté par des données. Même un système arrivé techniquement à maturité peut échouer en raison d'un manque de connaissances, d'une absence de motivation ou d'attentes excessives. Un document d'ensemble portant sur quatre volumineuses études de cas relatives à l'exploitation des données au niveau de la scolarité secondaire souligne, par ex., en tant que résultat clé, le fait qu'à elle seule, l'existence de systèmes techniques « does not guarantee effective decision making. Having data does not necessarily mean that they will be used to drive decisions or lead to improvements. » (Marsh, Pane, & Hamilton 2006, p. 10). Parallèlement, il est évident qu'une infrastructure technique qui permet la collecte, la mémorisation et l'évaluation de données de l'éducation constitue une condition préalable de base à remplir pour l'utilisation plus étendue de ces données. De plus, des solutions techniques peuvent contribuer à développer des compétences dans le domaine de l'utilisation des données ou peuvent accroître l'acceptation pour des mécanismes de prise de décision basés sur des données en rendant possibles un contrôle et une vérification de ces mécanismes. Deux rapports actuels sur les processus algorithmiques de prise de décision proposent notamment de recourir à de telles solutions automatiques (Krüger & Lischka 2018 ; SRV 2018).²³⁰

Conformément à ce qui précède, il existe des défis à relever à plusieurs niveaux pour l'utilisation des données dans le secteur éducatif. Sclater (2015) subdivise en trois grands domaines thématiques les défis qui prennent naissance dans le cadre de l'utilisation des méthodes du Learning Analytics : il s'agit des défis logistiques, des défis éthiques et des défis juridiques. Ces défis correspondent aux dimensions de l'écosystème des données présenté au chapitre 3. Même si l'utilisation des données dans l'ensemble du système éducatif comprend des aspects supplémentaires, l'approche figurant dans les lignes qui suivent suit donc cette taxonomie. Outre la tâche consistant à mettre à disposition une infrastructure technique à la fois sûre et fiable pour la collecte, la mémorisation et l'évaluation des données (défis de nature technique), ces défis incluent surtout des questions sur les conséquences qui découlent de l'utilisation des données pour la participation d'un acteur individuel du système éducatif à la société (défis éthiques et culturels), ainsi que sur l'adaptation des conditions cadres juridiques et la mise à disposition des ressources financières requises (défis juridiques). Les défis éthiques, juridiques et techniques ne peuvent pas toujours faire l'objet de distinctions aussi tranchées. Par exemple, l'implémentation d'un processus de consentement des personnes concernées relatif à la collecte de données clairement compréhensible et pertinent est une réponse à une

²³⁰ Le Conseil allemand des experts pour la recherche sur les consommateurs propose, par ex., en tant que moyen de garantir la qualité de prédictions et évaluations algorithmiques, de recourir à une « interface de transparence » (Gigerenzer, Müller, & Wagner 2018). Cette interface est basée sur le principe du « Black Box Tinkering » (Perel & Elkin-Koren 2017). Ce faisant, on fait varier systématiquement des entrées de données, et on évalue ensuite les sorties de l'algorithme. Ainsi, on peut en tirer des déductions sur le mode de fonctionnement d'un algorithme même si l'on n'en connaît pas le code source. En outre, on peut identifier assez facilement d'éventuelles distorsions ou des erreurs systématiques de l'algorithme (Gesellschaft für Informatik 2018).

exigence éthique selon laquelle les groupes cibles concernés doivent pouvoir comprendre à quelles fins, pour quelle destination et avec quels moyens des collectes et des évaluations de données sont réalisées. En revanche, ancrer sans ambiguïté ce processus dans des dispositions légales est un défi de nature juridique. Et la mise en œuvre correspondante réalisée dans le cadre d'un tel processus de collecte de données est souvent liée à des possibilités techniques.

6.4.1 Défis de nature technique

Lorsqu'on fait appel à des données numériques à grande échelle dans le domaine de la formation, il est indispensable de disposer d'une **infrastructure** technique qui rend possible une telle utilisation. Cela signifie qu'il faut prévoir une dotation complète des établissements éducatifs en terminaux numériques ainsi que la mise à disposition de capacités (de serveurs) suffisantes nécessaires à l'évaluation rapide des données ainsi collectées pour pouvoir développer ou acquérir les programmes correspondants. En outre, il faut pouvoir disposer d'un nombre de spécialistes suffisamment bien formés pour mettre en place un système d'analyse des données et en assurer la maintenance et le développement. Dès lors, il est vraisemblable que les investissements initiaux et les coûts de maintenance courants destinés à un tel système seront considérables. Toutefois, à l'heure actuelle, nous ne disposons d'aucune estimation sur le montant des coûts auquel il faut effectivement s'attendre. Leuthold et Lubich (chapitre 4.2) renvoient également à la contradiction potentiellement problématique entre la stagnation des budgets de l'administration publique de l'éducation et les besoins financiers accrus nécessaires à l'acquisition et à la maintenance d'une infrastructure IT sécurisée. Mais compte tenu des expériences faites dans le domaine de la formation tertiaire et des petites et moyennes entreprises (PME), une série de mesures de réduction des coûts peuvent être proposées.

De plus, une autre tâche qui revêt une grande importance, selon les contributions des experts sollicités, réside dans la définition de **standards** de validité globale concernant la collecte des données, l'architecture des données et l'échange des données. En règle générale, il existe des solutions locales isolées qui ne sont pas incompatibles pour la saisie et la mémorisation des données. Ces solutions isolées rendent considérablement plus difficiles l'échange et l'harmonisation des données et accroissent le risque de dépendance vis-à-vis de certains fabricants de logiciels (voir chapitre 4.4). De plus, des standards de validité globale permettent de définir des lignes directrices uniformes sur la qualité des données et rendent **possibles** l'implémentation d'un management de la qualité des données correspondant. De cette façon, on pourrait réduire au maximum des distorsions systématiques de données dues à des méthodes de mesure différentes ou à des mécanismes de transfert différents. Cela pourrait aussi offrir une meilleure vue d'ensemble sur les groupes de données collectés et enregistrés dans le domaine de la formation, et cela faciliterait aussi leur échange entre les acteurs pertinents. À l'heure actuelle, il n'est souvent pas clair de savoir si un besoin supplémentaire en données – comme cela a été annoncé, par exemple, dans le rapport actuel sur l'éducation en Suisse (Wolter et al. 2018) – s'explique par le fait que des données ne sont pas enregistrées, ou simplement par le fait que ces données ne sont pas accessibles. Un enregistrement complet de l'état des ensembles de données dans le domaine de l'éducation et un meilleur ajustement entre l'offre et la demande en matière de données sont également considérés par Thouvenin et al. (chapitre 4.1) et par Kellerhals (chapitre 4.3) comme des défis importants à relever. Comme le soulignent toutefois Leuthold et Lubich (chapitre 4.2), des limites sont assignées à la standardisation du côté de l'infrastructure et de la gestion de la sécurité de l'information en raison des différences qui existent dans la situation en matière de menaces et le besoin de protection.

En ce qui concerne la qualité, des défis résultent également des méthodes d'analyse et des algorithmes d'analyse utilisés. À ce jour, il n'existe en effet aucun **standard de qualité** généralement applicable et imposable **pour les analyses de données** et autres outils algorithmiques. Jusqu'à présent, les fabricants ne sont pas tenus de garantir ou d'apporter la preuve du caractère correct et de la qualité de leurs méthodes, et les utilisateurs ne peuvent pas vérifier cette qualité, ou seulement au prix d'une charge de travail disproportionnée. Cela ne rend pas seulement plus difficile l'évaluation de l'efficacité de certains outils spécifiques, mais cela empêche encore de découvrir des erreurs et distorsions systématiques dans l'analyse des données (voir plus loin). Alors qu'on discute actuellement d'une série de possibilités visant à valider à l'ex-

terne la qualité des décisions algorithmiques²³¹, il n'existe à ce jour que des lignes directrices non contraignantes rédigées en termes très généraux sur la manière dont les résultats de ces évaluations doivent être évalués (Krüger & Lischka 2018 ; Calo 2017). Il n'existe pas de valeurs limites généralement valables et techniquement définissables qui permettraient, par ex., de classer un système dans la catégorie « suffisamment valide ». C'est donc la tâche des milieux politiques (le cas échéant en collaboration avec les responsables de la recherche) de définir et d'imposer de telles valeurs limites.

De plus, la collecte globale de grandes quantités d'informations personnelles sensibles, par exemple les notes et évaluations scolaires, les résultats d'enquêtes de médecine scolaire ou des profils de personnalité, nécessite une approche globale en matière de garantie de la **protection et de l'intégrité des données**. Ceci est d'autant plus important que, dans de nombreux cas, des données provenant du secteur éducatif concernent des personnes légalement mineures dont la participation aux activités scolaires et, partant, à une collecte de données basée à l'école, est prescrite par la loi à titre obligatoire – du moins pour certaines parties de ces activités. La garantie de la sécurité de l'information pour les données issues du système éducatif est donc aussi un point d'importance décisive de la stratégie de numérisation de la CDIP (2018b). Les enfants et les adolescents ont en outre tendance à être moins critiques en ce qui concerne les risques actuels et futurs propres à une telle collecte de données (Livingstone, Stoilova, & Nandagiri 2018 ; Livingstone 2018). Lorsque ces données deviennent la propriété de l'apprenant, par exemple à l'issue de la période scolaire obligatoire, pouvoir garantir la protection et l'intégrité de ces données à ce moment-là est aussi un défi à relever. Puisque ces données contiennent des informations étendues sur le comportement d'apprentissage, les préférences et l'intelligence d'une personne donnée, elles revêtent un grand intérêt pour d'autres groupes cibles, p. ex des employeurs potentiels, des banques ou des compagnies d'assurance. Même une transmission facultative et volontaire de ces données qui serait effectuée par des élèves individuels pourrait entraîner des distorsions considérables sur le marché du travail, ou en cas d'octroi de crédits bancaires. Cela pourrait inciter, de leur côté, d'autres acteurs impliqués à mettre à disposition également leurs données (Kolowich 2013).²³² La question de savoir comment on peut réglementer une utilisation ultérieure de ces données par des tiers qui ne sont pas des acteurs directement liés au système éducatif n'a absolument pas été clarifiée jusqu'à présent – également et précisément lorsque ces données sont transmises à titre volontaire et facultatif.

En raison de l'évolution vertigineuse de la technologie, notamment dans le domaine de l'apprentissage machine, il en résulte une série de défis supplémentaires à relever pour la sécurité de l'information d'un système d'utilisation des données dans le secteur éducatif. Ces technologies augmentent en particulier la surface d'attaque et l'arsenal possible pour des cyberattaques, mais aussi les possibilités de défense du système contre de telles attaques.²³³ Par exemple, bon nombre des modèles algorithmiques qui sont utilisés pour des classifications, des prédictions ou des propositions peuvent être manipulés de manière ciblée par la mise à disposition systématique de fausses informations – il s'agit d'une technique connue sous le nom d'« Adversarial Machine Learning » (Kurakin, Goodfellow, & Bengio 2017). Un exemple facile à comprendre de manipulations de ce genre dans la vie quotidienne est le mode de fonctionnement des « Social Bots » qui faussent le contenu des médias sociaux. Ils essaient d'influencer la logique de décision propre aux réseaux sociaux en reprenant de manière automatisée certains contenus, en les copiant et en les rediffusant pour donner l'impression selon laquelle un thème donné suscite un intérêt général plus élevé que ce n'est effectivement le cas.

Une autre conséquence de l'évolution extrêmement dynamique du domaine de l'apprentissage machine, des structures d'intelligence artificielle basées sur ce dernier et de ses applications, réside dans le fait que les applications techniques existantes ont souvent un cycle de vie comparativement court, et qu'elles sont rapidement remplacées par de nouveaux développements. Une compréhension technique complète fait

231 Outre la possibilité déjà décrite du « Black Box Tinkering », une série de techniques supplémentaires a été proposée (Krüger & Lischka 2018). En font partie la vérification sur la base de l'analyse du code source (Zweig 2016), des jeux de données de tests standardisés (SRV 2018), des techniques dérivées du « Black Box Tinkering » (Perel & Elkin-Koren 2017), ou des explications d'application sous la forme de ce qu'on appelle des « Counterfactuals » (Wachter, Mittelstadt, & Russell 2018). Ces derniers décrivent le plus petit changement requis d'un attribut donné (par ex. l'attribut relatif au revenu) qui serait nécessaire pour atteindre un résultat d'analyse souhaité (par ex. pour qu'une évaluation soit considérée comme « digne de crédit »).

232 Par exemple parce que le refus de transmettre ces données est considéré comme le signe d'un risque de sécurité accru.

233 Dans un concours lancé en 2016 par l'agence de défense américaine Defense Advanced Research Projects Agency, on a montré que des systèmes apprenants automatisés sont en mesure de trouver et d'éliminer de manière autonome des erreurs de logiciels (Defense Advanced Research Projects Agency [DARPA], 2016).

souvent défaut pour pouvoir analyser les avantages et les inconvénients des différentes méthodes d'analyse pour différentes problématiques, et les risques liés au recours à ces méthodes (Kop, Fournier, & Durand 2017). Et ce, en particulier parce que les algorithmes sur lesquels sont basées de nombreuses applications sont des algorithmes propriétaires et que, de ce fait, ils ne sont pas accessibles au grand public intéressé. Pour tenir le rythme de cette évolution, et afin de pouvoir évaluer de manière valide leurs risques (et possibilités) techniques indépendamment des prestataires respectifs, il sera indispensable de mettre en place une expertise complète du côté des autorités régulatrices et un **encouragement durable à la recherche scientifique** dans ce domaine.

6.4.2 Défis éthiques et culturels

Un système intégré et sécurisé de collecte et d'évaluation de données standardisées dans le secteur éducatif n'a pas seulement pour mission centrale de rendre possible l'utilisation des données sous l'angle de la technique de gestion de l'information, mais de renforcer également la confiance dans cette utilisation et d'accroître ainsi son acceptation auprès des groupes cibles concernés. Créer cette confiance est une condition nécessaire à remplir pour que les connaissances tirées d'éventuelles analyses de données soient perçues comme utiles et équitables, et que les recommandations d'action qui en résultent soient aussi effectivement mises en œuvre. Pour ce faire, nous avons non seulement besoin de solutions techniques qui fonctionnent bien, mais encore d'autres solutions qui ont déjà fait leurs preuves pour résoudre une série de défis culturels et éthiques qui découlent de l'utilisation des données et de sa perception.

En font partie, d'une part, l'encouragement à une large compréhension des possibilités et des chances qui résultent de l'utilisation systématique des données de l'éducation. Et ce, d'autant plus qu'actuellement, la discussion publique tourne en priorité autour des risques liés à l'utilisation des données et sur les aspects de sécurité. Une analyse de contenu d'articles sur le thème du « Big Data » parus dans trois quotidiens suisses importants (NZZ, Tages-Anzeiger, Blick) entre 2001 et 2016 montre par exemple que les comptes rendus sur ce thème sont surtout critiques, et sont axés sur des problèmes (Huonder 2017). Dès lors, les Suissesses et les Suisses sont inquiets et redoutent que l'utilisation des données puisse enfreindre la protection de leur sphère privée. Selon un sondage de « World Internet Project – Switzerland » de 2017, 51 % des personnes interrogées font part de cette inquiétude au sujet des entreprises, 46 % d'entre elles au sujet des personnes privées, et près de 40 % d'entre elles au sujet des gouvernements, étant précisé que cette inquiétude croît avec l'intensité d'utilisation de l'Internet (Latzer, Büchi, Festic, & Just 2017). Il est hors de doute que l'accent mis sur les aspects de l'utilisation des données pertinents pour la sécurité est justifié.²³⁴ Néanmoins, si l'on s'oriente uniquement sur le débat public, cela rend plus difficile une analyse et une évaluation détaillées et complètes de ces technologies qui tient compte de manière appropriée tant des opportunités que des risques.

La nouveauté et la complexité relative des technologies utilisées constitue une raison importante de l'attitude réticente quant à l'évolution du potentiel de l'utilisation des données au sein du grand public et auprès des acteurs du système éducatif. Certes, une proportion croissante de la population dispose de bonnes compétences dans l'utilisation des médias numériques (OFS 2018b), mais en règle générale, ces compétences concernent la capacité de se procurer des informations sur Internet ou d'utiliser un programme de traitement de texte. En revanche, **les compétences et l'acquisition de compétences dans le domaine de l'analyse des données** sont encore d'un niveau plutôt faible (Ravaux et al. 2018). Le renforcement de ces capacités au sein des groupes cibles pertinents du système éducatif est donc une autre condition préalable élémentaire à remplir pour pouvoir faire appel avec succès à l'utilisation des données dans ce domaine. Mais cela ne sert pas seulement à gagner la confiance, car au-delà de ce qui précède, il est important d'avoir une compréhension des possibilités, des limites et des risques des analyses de données afin de parvenir à un consentement éclairé et bien informé concernant l'utilisation des données. Il s'agit de traduire les connaissances acquises en actions effectives et de faire face aux risques inhérents au traitement des informations tiré par des données. Ces risques sont connus, par ex., dans la psychologie des organisations, où l'on recourt parfois à l'expression « Glass-Cockpit-Syndrom » pour les caractériser (Angell & Samonas 2009). Ce syndrome décrit le phénomène selon lequel l'utilisateur final, en raison d'un afflux excessif d'informations et de leur visualisation, devient « aveugle », et il tend donc à laisser entièrement au système le

²³⁴ Dans cette même étude, par exemple, le pourcentage des personnes interrogées qui déclarent ne jamais lire les dispositions sur la protection des données des services utilisés est plus élevé (29 %) que celui des sondés qui indiquent lire souvent de telles dispositions (20 %).

soin de prendre des décisions, ignorant ainsi des informations supplémentaires (parfois importantes). Dans le domaine scolaire, ce serait par ex. le cas si des prédictions ou des propositions déduites d'un système d'apprentissage adaptatif étaient reprises sans que l'on ne contrôle si elles sont adaptées ou non à la situation concrète d'un apprenant. Comme le fait remarquer Zeide (2017), procéder à un tel contrôle est important puisque des systèmes automatisés sont certes souvent « adaptatifs », mais qu'en dernière analyse, ils se fondent sur des probabilités et des valeurs moyennes. Dès lors, ils ne sont guère en mesure de réagir à des aspects et circonstances spécifiques uniques en leur genre, notamment lorsque ces derniers – par ex. en raison de données d'entraînement incomplètes ou non spécifiques – ne peuvent pas être modélisés de manière appropriée.

Un défi important pour l'utilisation des données dans l'éducation est donc d'établir une **culture d'organisation et de gestion** qui permette un traitement éclairé des données. En fait partie la compréhension pour les possibilités de l'utilisation des données, et le fait de garantir des marges de manœuvre suffisantes pour qu'elles soient utilisées de manière innovante par les acteurs individuels. Parallèlement, les limites de ces applications doivent être à la fois connues et respectées. Des analyses de données, et les algorithmes basés sur ces dernières, devraient être considérés comme constituant un système de mise à disposition d'informations dans un but de soutien au processus de prise de décisions. Cela inclut également une prise de conscience générale portant sur les moments où il faut faire ces analyses de données, sur la manière dont ces analyses de données doivent être réalisées pour pouvoir répondre à certaines questions spécifiques, et sur la façon dont les résultats de ces analyses peuvent être mis en œuvre dans des situations concrètes. Dans la littérature spécialisée sur les systèmes d'apprentissage adaptatifs, on souligne aussi régulièrement à quel point il est primordial de disposer d'une culture d'organisation responsable pour pouvoir utiliser matériellement ces données de manière correcte. Un document de prise de position du groupe de travail ECAR Analytics du réseau Educause souligne, par exemple, que « before deploying predictive learning analytics solutions, an institution should ensure that its organizational culture understands and values data-informed decision-making processes » (Alhadad et al. 2015, p. 2) [passages soulignés par les auteurs]. Ces derniers sont en train de développer un modèle facile à utiliser d'analyse des facteurs de succès pour le processus d'apprentissage. Sur cette base, Baer et Norris (2017) développent un outil d'évaluation afin d'apprécier les progrès lors de l'implémentation de systèmes d'apprentissage adaptatifs au niveau tertiaire. Il comprend 30 points individuels, subdivisés en cinq catégories, qui permettent de dresser un état des lieux sous la forme d'une comparaison entre l'état actuel et l'état visé. Parmi ces cinq catégories, quatre d'entre elles incluent des aspects non technologiques des défis organisationnels à relever (Leadership, Culture/Behaviour, Politiques/Processus/Pratiques, Skills/Talent Development). Dans ce contexte, la direction a pour mission importante, par ex., de trouver un équilibre défendable entre la liberté de l'enseignement et de la recherche (et donc aussi la liberté dans le choix du matériel d'enseignement) d'une part, et la sécurité de l'information optimisable en limitant ces libertés, d'autre part (voir chapitre 4.2).

Mais afin de parvenir à améliorer la compréhension pour les possibilités et limites des analyses de données globales, il ne faut pas seulement informer et dispenser des cours de formation continue aux personnes qui sont directement confrontées aux résultats correspondants (enseignants, administration scolaire, parents, etc.). Il existe encore actuellement des lacunes considérables en matière de connaissances sur les conséquences et les effets des analyses de données sur l'amélioration des prestations des élèves et l'amélioration du pilotage du système éducatif. Alors que, par ex., les effets des plateformes d'apprentissage adaptatif ont pu être présentés de manière répétée dans de petites études individuelles, aujourd'hui comme hier, des évaluations globales d'implémentations réalisées à grande échelle font défaut (voir chapitre précédent). Jusqu'à présent, il n'est pas clair non plus de savoir comment les effets se différencient selon qu'il s'agit de tel ou tel groupe d'apprenants. Pouvoir répondre à la question de savoir à qui, sous quelle forme et dans quelles conditions des analyses de données sont utiles et, à l'inverse, dans quels domaines de telles analyses sont plutôt nuisibles est un grand défi à relever (Kop, Fournier, & Durand 2017). En outre, jusqu'à présent, des effets secondaires potentiels des systèmes de décisions basés sur des données dans le secteur éducatif n'ont pas été systématiquement étudiés, quand bien même de tels points critiques sont soulevés de manière répétée dans le débat scientifique et dans le débat public sur cette thématique (Kop, Fournier, & Durand 2017 ; Fenwick 2015). En font partie, par ex., la focalisation sur des compétences (facilement) quantifiables, mais qui ne sont pas forcément pertinentes, les effets négatifs du sentiment d'être sous (auto-) observation permanente sur la motivation des mauvais apprenants ou sur les membres de minorités so-

cialement stigmatisées, ou encore l'encouragement non intentionnel à un excès de gardiennage et à un empiètement exagéré des parents dus à la mise à disposition de données d'apprentissage actualisées en permanence (Kemenetz 2016). Il est dès lors indispensable de poursuivre **l'encouragement à la recherche scientifique** dans ce domaine.

6.4.3 Défis juridiques

Des prises de décisions algorithmiques basées sur des données peuvent avoir une influence considérable sur la prospérité et la participation des personnes à la société. Des modèles prédictifs sur les résultats futurs des études universitaires peuvent par ex. être inclus dans les décisions d'admission des apprenants à certains cursus d'études, ce qui contribue à déterminer leur avenir académique et économique. Alors que globalement, cela peut incontestablement avoir des effets économiques positifs pour les apprenants, les établissements et la société dans son ensemble, des décisions basées sur des données peuvent porter préjudice à certains individus ou à certains groupes (Prinsloo & Slade 2017). Et ce, d'autant plus que toute forme d'analyse des données repose sur l'hypothèse selon laquelle des modèles qui ont été établis dans des données historiquement collectées auparavant seront également valables pour l'avenir. Ainsi, par ex., pour prédire les futurs résultats des études universitaires, les algorithmes se basent sur des informations portant sur la relation entre certaines caractéristiques et les résultats d'études d'anciennes étudiantes et d'anciens étudiants. Ces modèles prédictifs peuvent donc avoir tendance à autorenforcer et à approfondir des inégalités existantes (Fenwick 2015). Un recours à des technologies de ce genre nécessite donc un cadre sûr qui permet aux personnes concernées de faire valoir leurs droits, et pas seulement sous l'angle de la sécurité de l'information. En fait partie également l'instauration de la transparence sur le recours aux collectes de données et aux analyses de données ainsi que le fait de garantir une possibilité institutionnalisée de consulter les données et de faire opposition s'agissant de leurs résultats. C'est également important parce que la participation au système scolaire – jusqu'à la fin de l'école primaire – est obligatoire. Les apprenants et les parents n'ont donc que des possibilités limitées de refuser de participer à des enregistrements de données dans la vie quotidienne de l'école et de l'enseignement.

Parallèlement, le risque existe qu'une régulation excessive fasse obstacle à l'utilisation innovante et productive des données et réduise ainsi à néant une partie du potentiel de cette utilisation pour le secteur éducatif. Cela concerne en particulier une régulation qui serait unilatéralement axée sur la question de la sécurité de l'information et qui, de ce fait, rendrait considérablement plus difficile l'accès aux données. Une régulation de ce genre contribuerait à isoler et à « couper du monde » les collectes de données publiques. Elle renforcerait ainsi le déséquilibre qui existe déjà actuellement en faveur des grands prestataires privés et nuirait ainsi à la concurrence et à l'innovation dans l'utilisation des données de l'éducation. Le développement et l'amélioration constante des systèmes de prise de décision algorithmiques modernes reposent sur l'accès à de grandes quantités de données qui sont utilisées pour l'« entraînement » des algorithmes. Lorsqu'un système acquiert davantage d'utilisateurs, et donc davantage de données, il a donc des chances plus élevées de poursuivre plus rapidement son développement. De petites différences initiales entre des systèmes concurrents peuvent donc très rapidement s'accroître. Pour le domaine de l'évolution des systèmes algorithmiques en général, cette tendance peut déjà être clairement constatée. Un très petit nombre de grandes entreprises privées dominant aussi bien le marché que le panorama de la recherche scientifique (Calo 2017). Même la possibilité de réutiliser des analyses de données pour garantir le contrôle et la transparence portant sur d'autres formes d'utilisation et d'évaluation des données serait compromise par une limitation stricte de l'accès aux données (par ex. parce que les données nécessaires provenant d'ensembles de données publiques ne peuvent pas être mises à disposition). Ainsi, si l'on a pu déduire que certains des systèmes automatisés qui sont utilisés aux États-Unis pour évaluer le risque de récidive de délinquants condamnés commettent des erreurs systématiques aux dépens de délinquants afro-américains, c'est parce qu'on s'est basé sur la comparaison systématique de données prédictives avec des événements effectivement survenus (Krüger & Lischka 2018). Ce sont des droits d'accès à ces informations ancrés dans la loi qui ont permis l'accès à ces données (Larson, Mattu, Kirchner, & Angwin 2016). Cet exemple montre bien que l'un des défis clés à relever dans la régulation de l'utilisation des données réside dans la recherche d'un équilibre entre la protection de la sphère privée des individus (par ex. avant une éventuelle réidentification au sein des données relatives aux probabilités de récidive) et la garantie d'un accès public à des données propriétaires d'un autre type (et aux résultats des analyses effectuées avec ces données).

Une utilisation judicieuse des données de l'éducation présuppose donc la **définition d'une gouvernance des données** qui crée un équilibre défendable entre les droits légitimes de l'individu à la sécurité et à la protection des données et les besoins de la société d'utiliser ces données pour améliorer le système éducatif et les résultats de l'apprentissage. Comme l'expliquent Thouvenin et al. (chapitre 4.1), il est nécessaire de créer à cette fin des bases légales qui prévoient explicitement l'échange et l'utilisation des données de l'éducation. À l'heure actuelle, il n'existe guère de pistes de réflexion à cet effet, quand bien même ces données doivent être impérativement disponibles pour le traitement des données personnelles effectué par des organes étatiques. En outre, les auteurs attirent l'attention sur le fait que la création de telles bases légales constitue un défi décisif à relever pour pouvoir concrétiser ultérieurement les objectifs d'une utilisation des données dans le secteur éducatif. Car ce n'est qu'ainsi qu'on pourra tenir compte des différentes obligations imparties, notamment en ce qui concerne la protection des données, le fait de lier des données à une destination précise, et l'utilisation minimale des données en cas de recours plus intensif aux données.

7 Une politique cohérente d'utilisation des données pour l'espace suisse de formation

7.1 Les données dans l'éducation – les données pour l'éducation

En raison de l'utilisation toujours croissante des terminaux et applications numériques à tous les niveaux, le système éducatif se voit confronté à une augmentation vertigineuse de la quantité des données. Il s'agit désormais d'apprendre à les gérer. Puisque les possibilités d'évaluation des données s'accroissent aussi continuellement, il en résulte de nombreuses possibilités pour des utilisations de données qui génèrent de la valeur ajoutée tant pour les acteurs individuels de l'éducation que pour les écoles et pour le système éducatif dans son ensemble (voir chapitre 6.2 Potentiels). Toutefois, ces potentiels ne peuvent déployer leur effets que si l'on crée un cadre qui soit en mesure de répondre aux défis actuels, resp. que si l'on établit une politique d'utilisation des données pour l'espace suisse de formation qui soit axée en conséquence sur l'avenir.

7.2 Nécessité d'un cadre commun

Utiliser des données de manière sécurisée et éthique et en générant de la valeur ajoutée dans et pour le système éducatif présuppose de mettre en place un cadre correspondant qui soupèse avec soin et qui équilibre entre eux les différents intérêts des acteurs individuels de l'éducation et ceux du système dans son ensemble, et qui recherche des solutions constructives²³⁵ (selon la devise : protéger au mieux les individus – tirer parti au mieux des données²³⁶). Ce cadre comprend différentes composantes (composantes juridiques, spécifiques à la sécurité de l'information, éthiques, etc.) (voir le cadre de l'écosystème des données de l'éducation, chapitre 3.3) qui sont interdépendantes et doivent donc être harmonisées entre elles. En fonction de l'ampleur des besoins de régulation, il s'agira de trouver chaque fois le niveau adéquat (avons-nous besoin d'une base légale, ou une solution à bas seuil est-elle suffisante ? Des principes éthiques peuvent-ils conduire à l'objectif poursuivi et/ou une sensibilisation des acteurs est-elle appropriée ?, etc.), de faire appel aux acteurs adéquats et, ce faisant, de penser au-delà des structures actuelle-

235 Voir à ce sujet le concept de « politique constructive en matière de données » (Golliez 2018).

236 Tant le système dans sa globalité que le sujet des données doivent pouvoir utiliser et réutiliser des données en leur faveur. Cela signifie qu'il faut pouvoir réaliser si possible toutes les analyses de données qui apportent des avantages au système éducatif et à ses acteurs. Le sujet des données lui-même devrait pouvoir en particulier continuer d'utiliser ses données en sa faveur (approche « MyData »).

ment existantes – qui ont été créées pour un système éducatif fonctionnant selon les règles du monde analogique.

De manière générale, les principes suivants peuvent être fixés pour la mise en place d'un tel cadre :

- Le cadre doit être applicable à l'échelle de toute la Suisse (« il n'existe qu'un seul espace de données [de l'éducation] ») :
Dans une société numérisée, des solutions isolées n'ont aucun sens ; elles empêchent l'efficacité, rendent les innovations impossibles et créent des inégalités (par ex. pour les écoles de différents cantons ou, pour les prestataires, en raison de l'existence de directives différentes sur la protection des données). La Confédération, les cantons et les communes travaillent de concert afin de garantir cette perméabilité – également au-delà du seul secteur éducatif.
- Le cadre sera établi par le biais d'une approche globale et intégrative :
Il ne s'agit pas de trouver des solutions individuelles, mais au contraire de règlementer l'utilisation des données dans sa globalité. Il faut intégrer au même titre tous les éléments cadres dans les pistes de solutions.
- Le cadre devra être agile et flexible :
Afin de pouvoir réagir en conséquence aux futurs défis, on choisira des solutions qui permettent de réévaluer continuellement la situation ainsi que l'intégration d'adaptations à un rythme élevé pouvant aller jusqu'à des remaniements fondamentaux.
- Le cadre sera développé en commun (« politique constructive en matière d'utilisation des données ») :
La mise en œuvre ne peut pas être réalisée par les instances et institutions concernées du système éducatif à elles seules, mais elle nécessite au contraire une collaboration étroite entre tous les groupes cibles. Leurs intérêts respectifs seront soupesés et seront repris dans une solution consensuelle. Et ce, non seulement en raison des aspects financiers, mais surtout également parce que des compétences techniques sont requises à cet effet.

7.3 Caractère urgent

Un autre principe qui est à la base de la mise en place d'un cadre commun pour l'utilisation des données dans le système éducatif réside dans son caractère urgent. Compte tenu de son importance, un chapitre propre lui est ici consacré.

Il est approprié d'agir d'urgence pour les raisons suivantes :

- **Aujourd'hui, nous ne pouvons tirer entièrement parti des potentiels de l'utilisation des données dans le système éducatif.** Outre les lacunes dans l'équipement de base technique et des compétences plutôt limitées jusqu'à présent, le fait que de nombreux aspects juridiques et procéduraux de la protection des données et de la disponibilité des données ne sont pas encore clarifiés explique les incertitudes de la situation actuelle.
- **Risque d'accentuation du fédéralisme des données.**²³⁷ Il faut aussi agir d'urgence parce que les cantons, les communes et les écoles sont en train de rechercher actuellement leurs propres solutions, ce qui renforcera encore l'hétérogénéité / les solutions isolées, créera des inégalités et ce qui rendra considérablement plus difficiles des pistes de solutions ultérieures de portée globale.
- **Extension croissante et poursuite continue du développement de la situation en matière de menaces.** Des menaces prennent naissance pour les établissements scolaires et pour les établissements universitaires de recherche et d'enseignement, d'une part, en raison de changements sociétaux, organisationnels, économiques et techniques survenant hors de leur sphère d'influence directe et, d'autre part, en raison de défauts dans la structuration, l'exploitation et l'adaptation continue des dispositifs de protection correspondants (voir chapitre 4.2.2, p. 68). Le système éducatif doit avoir pour mission centrale de garantir une protection appropriée. Et ce, d'autant plus que, de ce fait, on ne garantit pas

²³⁷ Voir Zehnder (1989).

seulement la sécurité du droit lorsqu'il est fait usage des données, mais qu'on garantit également la capacité d'agir des instances étatiques de pilotage s'agissant de l'utilisation des données.

- **Exploitation unilatérale du potentiel.** Aujourd'hui déjà, des prestataires privés (et aussi étatiques) de matériel pédagogique numérique et d'autres services numériques sont en train de récolter des données détaillées sur le comportement d'apprentissage et sur les résultats obtenus par les élèves et, la plupart du temps, ils les utilisent à titre exclusif. En vertu de conventions d'utilisation actuellement en vigueur, les institutions qui génèrent des données ne peuvent absolument pas y accéder, et les sujets de données ne peuvent y accéder que de manière limitée (par ex. par le biais du droit d'accès dont ils disposent en vertu du droit de la protection des données). De plus, il faut présumer que ces institutions et ces personnes ne sont pas au clair sur l'étendue et la nature de la collecte de données effectuée par les prestataires.

7.4 Champs d'action

Lors d'un workshop²³⁸ qui a eu lieu avec les auteurs des articles techniques (voir chapitre 4) en collaboration avec d'autres expertes et experts²³⁹ – que nous avons sollicités pour procéder à une triangulation relative au contenu – ainsi qu'avec les mandants, les différentes perspectives spécialisées ont été regroupées en tenant compte des bases élaborées aux chapitres 2, 3, 5 et 6, et des pistes de solution ont été discutées en commun. Ce faisant, au cours d'une première étape, des champs d'action dans lesquels des mesures doivent être prises ont été identifiés afin d'atteindre les objectifs figurant ci-après :

Il s'agit de rendre possible une utilisation des données cohérente à l'échelle de toute la Suisse et qui puisse générer de la valeur ajoutée dans le secteur éducatif. Cette dernière est tenue de suivre des principes éthiques appropriés et doit garantir la sécurité nécessaire.²⁴⁰

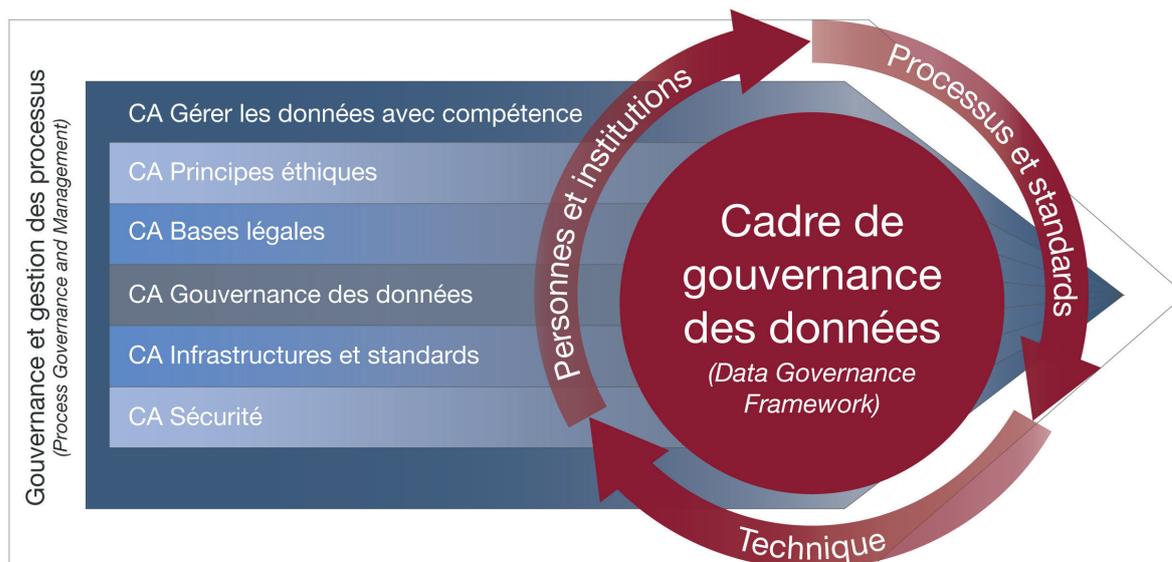
La figure suivante (graphique 2) permet de visualiser les différents champs d'action à gérer. Ils se recoupent typiquement avec les conditions cadres propres à l'écosystème des données de l'éducation (voir chapitre 3.3, p. 23). Globalement, ils ont pour effet que l'utilisation des données peut véritablement avoir lieu et qu'elle se déroule dans un cadre dûment réglementé. Le champ d'action « Gérer les données avec compétence » constitue la base préalable de toute action qui vise à ce que chaque acteur dispose de compétences déterminées dans la gestion des données pour qu'il puisse assumer son rôle en conséquence au sein du système éducatif. Les champs d'action « Principes éthiques » et « Sécurité » représentent aussi simultanément une directive d'objectifs : la future utilisation des données devra se faire de manière sécurisée et devra être guidée par des principes éthiques appropriés. Les champs d'action « Bases légales » et « Infrastructures et standards » offrent deux instruments de régulation²⁴¹, mais à des niveaux différents. Le champ d'action « Gouvernance des données » vise déjà la mise en œuvre des résultats obtenus dans les différents champs d'action (qui est responsable, comment les processus doivent se dérouler, etc.). Enfin, le champ d'action « Gouvernance et gestion des processus » assume la coordination en jouant le rôle d'« interface » entre les différents champs d'action et, de ce fait, constitue le point de départ de l'établissement d'un cadre cohérent dans ce domaine. Car tous les champs d'action doivent être considérés comme des modules interdépendants dont le degré d'importance doit être négocié en interaction réciproque. Cela nécessite une approche globale et transversale qui inclut d'emblée, et au même titre, tous les champs d'action et toutes les conditions cadres.

238 Ce workshop a eu lieu le 22 février 2019 à Berne.

239 Voir la liste des participants en annexe.

240 Cette fixation d'objectif a été intentionnellement formulée en termes vagues. Il s'agit d'instaurer un équilibre entre les droits de l'individu à l'autodétermination informationnelle, à la sécurité de l'information et à l'utilisation minimale des données, et les besoins de la société d'utilisation de ces données afin d'améliorer le système éducatif et les résultats d'apprentissage, et afin de créer d'autres applications possibles (par ex. dans le contexte « Smart City »).

241 Bien entendu, les autres champs d'action servent aussi à réguler l'utilisation des données, mais pas exclusivement.



Graphique 2 : Champs d'action (CA)

CA 1 : Gérer les données avec compétence

L'acceptation de cette gestion des données par les groupes cibles concernés constitue une condition préalable d'importance cruciale à remplir pour qu'elle puisse être introduite avec succès, et pour que l'utilisation des données de l'éducation génère de la valeur ajoutée dans le système éducatif. Cette acceptation est d'autant plus pertinente que le système politique suisse est caractérisé par un degré élevé de participation citoyenne. Une discussion éclairée sur les chances, les limites et les risques de l'utilisation des données dans le secteur éducatif est donc décisive tant pour le processus qui permet de se forger une opinion à ce sujet que pour le processus de prise de décision y afférent.

Le champ d'action « Gérer les données avec compétence » regroupe les pistes de réflexion qui permettent de disposer d'un niveau de compréhension suffisant à tous les niveaux du système éducatif concernant :

- les bases factuelles de l'utilisation des données (où cela se passe-t-il, et qu'est-ce qu'il se passe ?),
- les bases techniques correspondantes (comment cela fonctionne-t-il ?),
- les bases fonctionnelles correspondantes (comment puis-je tirer parti de cela ?) et
- les bases sociétales correspondantes (quel est l'impact sur la société ?).²⁴²

Alors que différents groupes cibles, selon leur fonction respective, doivent disposer d'un niveau de compréhension plus ou moins approfondi de ces aspects propres à l'utilisation des données, les pistes de réflexion figurant dans ce champ visent à garantir une compréhension de base commune sur tous les aspects et sur tous les groupes d'intérêts concernés. Toutefois, en raison de l'axe d'orientation prioritaire du présent rapport, nous mettons ici l'accent sur le niveau systémique.

CA 2 : Bases légales

L'utilisation des données – en particulier des données personnelles – dans le système éducatif est réglementée par une série de bases légales. Ses principaux piliers sont la législation de la Confédération sur la protection des données et les lois correspondantes des cantons. Elles fixent un cadre étroit applicable à la collecte, à la mémorisation et au traitement des données personnelles effectué notamment par les organes de l'administration publique. Le champ d'action « Bases légales » fait la lumière sur les possibilités et les limites offertes par le cadre légal actuel pour une utilisation plus intensive des données dans le secteur éducatif et décrit, le cas échéant, les adaptations nécessaires à apporter afin de faciliter une telle utilisation.

²⁴² Gesellschaft für Informatik (2016) ; les trois perspectives de Dagstuhl-Dreieck ont été étendues à une quatrième, celle des bases factuelles.

CA 3 : Sécurité

Les données provenant du secteur éducatif concernent souvent des informations sensibles sur le comportement, l'évolution et les résultats scolaires de personnes mineures fréquemment identifiables sans aucune ambiguïté. Une protection complète de ces données contre une utilisation non voulue ou une retransmission des données à des tiers est donc une condition préalable déterminante à la poursuite de leur utilisation. En outre, du point de vue de la cybersécurité, il existe plusieurs risques pour les systèmes techniques qui sont nécessaires à la collecte, à la préparation et à l'analyse de ces données (voir chapitre 4.2). Le champ d'action « Sécurité » décrit des approches et moyens possibles de réduire ces risques au maximum pour l'infrastructure technique et pour les ensembles de données détenus à tous les niveaux du système éducatif.

CA 4 : Principes éthiques

Tout particulièrement dans le secteur éducatif, l'utilisation et l'évaluation systématiques des données peuvent avoir des conséquences considérables et à long terme pour la prospérité et la participation des individus à la société. Cette forme d'utilisation des données présuppose donc l'existence d'un consensus éclairé sur les exigences sociétales, sociales et spécifiques à l'éducation auxquelles elle devrait satisfaire : avec quels objectifs, sous quelles conditions et pour quels types de problématiques l'utilisation des données est-elle appropriée et correcte ? Qui peut et qui est autorisé à définir, à contrôler et, le cas échéant, à adapter ce caractère approprié, quand et comment ? Dans le champ d'action « Principes éthiques », il faudrait mettre en évidence des voies et moyens précisant comment un tel consensus peut être atteint au sein de la société. Un profil d'exigences correspondant devrait être esquissé en tant que point de départ de la discussion.

CA 5 : Infrastructures et standards

Afin de pouvoir collecter et évaluer systématiquement des données dans le domaine éducatif, nous avons besoin d'une infrastructure qui rend cela possible. Cela ne concerne pas seulement l'infrastructure physique (par ex. des ordinateurs, des serveurs et un réseau de liaison), mais encore une standardisation de portée globale des processus de saisie, de mémorisation et de traitement des données qui prennent naissance dans le secteur de l'éducation. Le champ d'action « Infrastructures et standards » regroupe les approches qui peuvent favoriser et accélérer une implémentation et une diffusion de tels standards.

CA 6 : Gouvernance des données

La gouvernance des données décrit la collecte de lignes directrices et de règles qui organisent, contrôlent et rendent possible une utilisation standardisée des données de l'éducation. Elles réglementent, au niveau pratique, la disponibilité, l'intégrité, la sécurité des données de l'éducation et leur convivialité pour les utilisateurs. Elles garantissent l'utilisation appropriée et éthique des données ainsi que des analyses effectuées à l'aide de ces dernières. Elles reflètent les résultats des autres champs d'action et formalisent leur mise en œuvre et leur opérationnalisation dans les limites de l'espace de formation suisse. Dans le champ d'action « Gouvernance des données », on discute des éléments qui devront faire impérativement partie d'une telle structure, on identifie des moyens d'établir une telle structure, et on met en évidence des possibilités de l'implémenter.

CA 7 : Gouvernance et gestion des processus

Outre les champs d'action dont le contenu est déterminant, la coordination effective et l'organisation (temporelle) des différentes étapes découlant de ces champs d'action et l'implémentation d'une utilisation des données de portée globale dans le domaine éducatif revêtent une dimension importante. Précisément en raison de la structure fédéraliste des responsabilités et compétences respectives dans le système éducatif, une telle coordination au-delà de tous les champs d'action et de tous les acteurs est une nécessité urgente. Le champ d'action intégrateur « Gouvernance et gestion des processus » décrit donc de possibles principes de base pour une harmonisation optimale des processus dans les autres champs d'action.

7.5 Conclusion

La politique éducative est aussi une politique (d'utilisation) des données

Comment peut-on tirer parti de la richesse des données qui existe dans le secteur de la formation, et sous quelles conditions peut-on la mettre à disposition de tiers ? C'est ce type de questions que doit se poser le secteur éducatif – tout comme tous les autres secteurs de la société – dans un monde numérisé. Ce faisant, il semble important de ne pas considérer les données comme des sous-produits, mais plutôt comme constituant (une partie) des biens d'infrastructure²⁴³. Par conséquent, il s'agit d'investir dans ces biens d'infrastructure, il faut en assurer la maintenance, et il faut en réglementer l'utilisation (voir chapitre 7.2 Nécessité d'un cadre commun) afin que l'éducation puisse profiter de leurs avantages et qu'elle soit soutenue par ces données.

S'attaquer activement au changement de paradigme

Considérer les données comme des biens d'infrastructures nécessaires propres à l'espace de formation numérique nécessite à la fois de repenser les choses au sein du système et de le réinventer de fond en comble : car ce n'est qu'ainsi que ce changement de paradigme pourra réussir à long terme. Pour cela, une approche progressiste commune est nécessaire. L'objectif du présent rapport est de donner le coup d'envoi à cette réflexion. Nous ne recherchons pas des solutions « top-down », mais au contraire des approches constructives qui donneront au système et à ses acteurs la possibilité de participer à sa propre croissance et à son propre processus de maturité afin de pouvoir répondre en fin de compte de manière éclairée à la question de savoir comment nous voulons enseigner et apprendre à l'avenir.

La Confédération et les cantons, en tant qu'organes assumant la responsabilité commune du système éducatif – conformément à leurs responsabilités et compétences respectives – doivent créer les bases nécessaires afin de faire en sorte que l'on puisse s'attaquer de manière coordonnée aux mesures à prendre décrites dans les champs d'action (voir chapitre suivant) et que l'on puisse établir pas à pas une politique d'utilisation des données cohérente correspondante. Le comité de coordination de la numérisation (CC N) est la plateforme appropriée pour assumer cette tâche de coordination.

7.6 Pistes de réflexion en vue de l'élaboration d'une politique d'utilisation des données pour l'espace suisse de formation

Tirer parti de la manière la plus complète possible des potentiels de l'utilisation des données dans le système éducatif est un projet complexe et de grande ampleur qui présuppose chez les responsables, la compréhension de l'écosystème des données de l'éducation et des interactions entre ses éléments (voir chapitre 3). Car ce n'est qu'ainsi que nous pourrions adopter le point de vue global nécessaire à cette thématique qui permettra d'inclure également dans les réflexions la totalité des composants du cadre que nous visons (chapitre 3.3). Ces composants revêtent le même degré d'importance et se déterminent aussi mutuellement.

Dans ce contexte, dans les lignes qui suivent – en nous fondant sur les bases élaborées dans le présent rapport et sur les évaluations des experts – nous mettrons en évidence des approches qui permettront d'élaborer un espace de données de l'éducation cohérent. Il s'agit, d'une part, des potentiels de développement spécifiques aux différents champs d'action qui visent les éléments cadre individuels de l'écosystème. Pour ce faire, l'établissement d'une gouvernance des processus et d'un management correspondant (CA 7) requiert une harmonisation cohérente des différentes mesures (complémentaires). D'autre part, des propositions de mise en œuvre seront présentées qui soulignent la nécessité de l'adoption d'une approche globale et attireront l'attention sur des possibilités d'étapes concrètes.

Tant les mesures spécifiques aux différents champs d'action que les mesures intégratives conduisent aux objectifs visés en vue d'élaborer une politique d'utilisation des données cohérente pour autant que l'on garde la vue d'ensemble sur la totalité du système. Toutefois, idéalement, il s'agit de poursuivre le traite-

243 « Infrastructures are the basic equipment and structures [...] that are needed for a country, region, or organization to function properly » (Frischmann 2012). On considère que les données sont des biens d'infrastructure parce qu'elles ont les trois caractéristiques suivantes : 1. Elles peuvent être utilisées par tous les acteurs dans un contexte de non-rivalité, 2. Elles sont un bien d'investissement (et non un bien de consommation) et 3. Elles peuvent être utilisées pour un nombre illimité d'applications (OCDE 2015). Voir aussi chapitre 4.3.1, p. 84.

ment des différentes pistes de réflexion en les considérant comme un paquet global comportant des parties individuelles auxquelles il sera parfaitement possible et nécessaire de s'attaquer selon des rythmes et des degrés d'intensité différents.

Lors de la mise en œuvre des différentes pistes de réflexion, le travail de mise en réseau à fournir aussi bien au sein du système éducatif qu'au-delà de ce dernier est à la fois synonyme de gains importants à réaliser et de défis à relever. Cette mise en réseau permettra de créer des synergies avec des mesures déjà existantes et d'en tirer parti au mieux. Il s'agit de bénéficier et d'apprendre les uns des autres et de structurer un espace de données commun. Car en dernière analyse, la politique en matière d'utilisation des données de l'éducation doit être considérée comme faisant partie d'une politique générale (transsectorielle) en matière d'utilisation des données. Il faut donc encourager la perméabilité et la capacité de raccordement aux autres secteurs.

Partant, pour toutes les pistes de réflexion figurant ci-après, la règle suivante est applicable : là où cela s'avère possible, il faut recourir aux structures et ressources existantes (étant précisé qu'il faut aussi tenir compte des structures et ressources transversales externes au secteur éducatif), en se conformant à la devise suivante : il s'agit d'utiliser en commun et différemment les structures existantes grâce à la mise en réseau. Ainsi, pour chaque piste de réflexion, il s'agit d'étudier concrètement quels types de structures et de mesures existent déjà, et comment ces dernières pourront être utilisées pour le présent projet.

Cela vaut aussi pour les enregistrements des ensembles de données qu'il faut prévoir dans les différents domaines (saisie systématique des flux de données ; examen des lois cantonales y afférentes concernant le domaine de la formation ; établissement d'une liste de solutions logicielles et de solutions de banques de données qui sont utilisées actuellement).

7.6.1 Pistes de réflexion spécifiques aux différents champs d'action

CA 1 : Gérer les données avec compétence

Encourager la compréhension pour la transformation numérique en général et pour le rôle de l'utilisation des données en particulier, notamment auprès des commissions chargées de prendre des décisions. Cela inclut la sensibilisation à l'importance de l'utilisation des données, la mise en place d'une prise de conscience en matière de données (*Data Awareness*) et d'une initiation aux compétences de base pour la maîtrise des données (*Data Literacy*) dans le système éducatif

CA 2 : Bases légales

Développement continu et coordonné, à l'échelle de toute la Suisse, du droit d'accès aux données en vertu de la législation sur la protection des données afin de garantir la portabilité des données dans l'espace éducatif suisse²⁴⁴

Étudier l'opportunité d'introduire, à l'échelle de toute la Suisse, un droit d'accès aux données de l'éducation qui sont détenues à titre privé.

Étudier l'opportunité d'introduire un droit d'accès privilégié renforcé et standardisé, valable pour toute la Suisse, qui serait accordé aux responsables de la recherche en éducation afin de garantir un accès approprié aux données de l'éducation pour la recherche, l'évaluation et la planification.

²⁴⁴ Si, le cas échéant, dans le cadre de la discussion en cours sur la révision de la loi fédérale sur la protection des données (LPD), le Parlement parvient à la conclusion selon laquelle il faudrait créer un droit suisse de la portabilité des données similaire à celui du Règlement général de l'UE sur la protection des données (RGPD), il faudrait alors contrôler si les besoins du secteur éducatif sont ainsi couverts et/ou si des adaptations deviendront nécessaires à d'autres niveaux (par ex. dans les lois et ordonnances scolaires).

CA 3 : Sécurité

Définition et implémentation de mesures de protection minimales pour la sécurité de l'information dans la totalité des institutions éducatives suisses.

CA 4 : Principes éthiques

Élaborer une première proposition relative à des principes éthiques de base à observer dans le cadre de l'utilisation des données de l'éducation.

CA 5 : Infrastructures et standards

Implémentation d'un standard uniforme, valable pour toute la Suisse, relatif aux données de l'éducation.

Mise en place d'un entrepôt de données (*Data Warehouse*) national pour les données de l'éducation.

Tirer entièrement parti du potentiel des données ouvertes dans le domaine de l'éducation à titre de base préalable en vue d'adopter de nouvelles pistes créatives de résolution de problèmes, pour exploiter de nouvelles connaissances ainsi que pour accroître l'efficacité dans l'administration.

CA 6 : Gouvernance des données

Définition de structures et processus de gouvernance des données appropriées, en accord avec les mesures à prendre dans les autres champs d'action.

CA 7 : Gouvernance et gestion des processus

Garantir la coordination de la totalité des mesures à prendre.

Propositions de mise en œuvre

Centre de compétences pour l'utilisation des données

Mise en place d'un réseau de compétences pour l'utilisation des données dans le système éducatif. Il exploite et met en valeur les compétences techniques pertinentes en matière d'utilisation des données axées sur les différents champs d'action précités et sert de centre d'aiguillage et d'information à tous les acteurs du système éducatif pour répondre à leurs questions dans ce domaine.

Pour une concrétisation rapide, il est recommandé de mettre l'accent, dans un premier temps, sur des problématiques spécifiques à l'éducation. Lors d'une étape ultérieure, il pourrait être tout à fait judicieux d'institutionnaliser ce réseau et de le transformer en un centre de compétences général pour l'utilisation des données au sein duquel on créerait un service chargé de la formation (voir Frey, Rogg, Savolainen, Schmid, & Wandeler 2018).

Encouragement ciblé à des projets pilotes afin d'établir un cadre adéquat

Soutenir des projets pilotes qui feront la lumière sur le fonctionnement de cet « écosystème des données de l'éducation » dans la réalité, concrétiser des objectifs d'utilisation des données, et tester des pistes de solutions dans le champ de tension qui existe entre les intérêts respectifs des différents groupes cibles. Les solutions qui ont fait leurs preuves devront être mises en œuvre immédiatement.

Ce mode de faire rendra simultanément les acteurs capables de maîtriser le système (le caractère urgent et la nécessité d'agir deviendront visibles, et tant les avantages potentiels que les défis à relever deviendront observables et pourront être expérimentés dans la réalité quotidienne du système éducatif). Cela permettra d'analyser dans le détail la nécessité d'agir et d'élaborer des solutions pratiques conformes aux besoins (y compris la définition des rôles et des responsabilités).

8 Glossaire

Accès aux données	L'accès aux données décrit le degré et les modalités de la disponibilité des données.
Algorithme	Le concept d'algorithme désigne un ensemble structuré d'étapes et de règles d'action exemptes d'ambiguïté qui doivent être suivies lors de la résolution d'un problème défini à l'avance. Elles déterminent comment les entrées de données doivent être traitées et transformées en sorties de données. En informatique, ces règles sont formulées dans un langage de programmation (code). (Krüger & Lischka 2018, p. 9)
Anonymisation	L'anonymisation (également : anonymisation absolue) décrit le processus où l'on élimine après coup durablement, et de manière irréversible, la référence à des personnes contenue dans des données. Si une réidentification probabiliste n'est plus possible qu'au prix d'une charge de travail considérable, les données personnelles anonymisées ne relèvent alors plus du domaine de protection propre au droit de la protection des données. (Jensen 2012, p. 14)
Artificial Intelligence (<i>Intelligence artificielle</i>)	L'intelligence artificielle est un domaine partiel de l'informatique qui s'occupe du développement de machines, et notamment de programmes informatiques qui sont en mesure de résoudre certaines tâches dans le monde (McCarthy 2007). Elle peut inclure, mais ce n'est pas obligatoire, des processus d'apprentissage machine ou des processus d'apprentissage statistique. Le programme d'IBM « Deep Blue », qui a battu pour la première fois un champion du monde d'échecs en 1997, était, par ex., basé sur un « algorithme de traversée » qui, lors de chaque round, évaluait des millions de coups possibles, mais ne tirait aucune leçon des résultats du choix finalement retenu. (Campbell, Hoane, & Hsu 2002)
Attribut	Par le terme d'attribut (dans la recherche sociale empirique, on parle aussi de variable), on désigne une caractéristique ou une propriété qui est attribuée à un sujet de données. Dans ce contexte, il faut distinguer entre l'attribut lui-même (par ex. une note scolaire quelconque en mathématiques) et la spécification de cet attribut (par ex. la note scolaire 5). (Pflaumer, Heine, & Hartung 2009, p. 13)
Autodétermination informationnelle	L'autodétermination informationnelle désigne le droit de l'individu à déterminer lui-même s'il entend divulguer ou non et utiliser ses données à caractère personnel. Ce concept provient initialement de la jurisprudence allemande. En tant que droit fondamental non écrit et en tant que principe de rang constitutionnel, il a aussi fait son entrée dans le système judiciaire suisse.

Black Box Tinkering	Le « Black Box Tinkering » (litt. : « bricolage de boîte noire ») est une approche expérimentale qui tente de comprendre le mode de fonctionnement d'un algorithme ou d'un système décisionnel automatisé et qui cherche à évaluer la qualité de ses sorties de données, mais sans connaître l'implémentation technique (le code) de cet algorithme. Pour ce faire, on varie systématiquement les entrées de données, et on les compare avec les sorties de données de l'algorithme basées sur ces entrées. (Perel & Elkin-Koren 2017)
Data Literacy	La « Data Literacy » (<i>littératie des données</i>) décrit la capacité des individus d'extraire des informations à partir de données, et de tenir compte de manière appropriée de ces informations lors de la résolution de problèmes. Cela inclut le fait de découvrir, d'évaluer, de manipuler, de regrouper, d'interpréter et de présenter des données pertinentes pour une problématique déterminée. (Shield 2004 ; Prado & Marzal 2013)
Data Mining	Le « Data Mining » (on parle aussi d'apprentissage statistique) décrit la combinaison de méthodes statistiques et informatiques qui a pour but de découvrir des modèles, des relations et des régularités tirés par des données dans des ensembles de données (il s'agit souvent de données très volumineuses, dispersées et complexes). Le Data Mining sert à extraire des informations jusqu'ici inconnues et potentiellement utiles à partir de données valides (au sens statistique du terme) (Fayyad, Piatetsky-Shapiro, & Smyth 1996). Les méthodes propres au Data Mining constituent la base de l'apprentissage machine, lequel peut à son tour être utilisé dans le cadre du développement de systèmes d'intelligence artificielle.
Data Warehouse (Entrepôt de données)	Un entrepôt de données (<i>Data Warehouse</i>) est le résultat de la mise en œuvre conceptuelle et technique d'une collecte logiquement uniforme et systématique de toutes les données d'une organisation (par ex. du système éducatif) ainsi que de leur stockage et de leur mise à disposition qui sont nécessaires au pilotage et au fonctionnement effectif de ladite organisation. (Gabriel, Gluchowski, & Pastwa 2009)
Décision basée sur des preuves empiriques	Le processus de prise de décision basé sur des preuves empiriques décrit l'évaluation de propositions et concepts de solutions concurrents fondés sur des méthodes expérimentales et scientifiques.
Données	Une suite de signes (ordonnés syntaxiquement) (ou des fonctions continues de signes) qui servent à collecter, mémoriser et représenter des informations, des faits et des observations provenant du monde réel.
Données d'administration	Les données qui prennent naissance ou sont enregistrées dans le cadre du fonctionnement (quotidien) et de l'administration des institutions du système éducatif (écoles, administration scolaire, etc.).
Données d'application	Les données qui sont générées dans le cadre de l'acquisition et de l'utilisation de produits et prestations de services (par exemple pour le pilotage du système éducatif et de ses institutions). Ces données servent actuellement aux acteurs concernés – qui sont issus la plupart du temps de l'économie privée – aux fins de monitoring et d'amélioration de leurs produits, et aux fins d'optimisation de leur offre sous l'angle du comportement des utilisateurs en matière de demande.
Données d'enregistrement	Données qui sont générées dans le cadre de recensements et d'enregistrements institutionnalisés (par ex. enquêtes COFO, PISA, etc.) et « spontanés » (par ex. recherche universitaire ou de hautes écoles) réalisés dans le domaine de la formation.

Données de l'éducation	Données qui sont générées dans le cadre du système éducatif et qui fournissent des énoncés sur ce dernier ou permettent de faire des déductions à son sujet. On les désigne aussi par l'expression « données se rapportant à l'éducation ».
Données de performance ou données de prestations	Décrivent les résultats de processus d'évaluation et d'appréciations de prestations. Elles servent en priorité à évaluer le niveau de compétence d'une personne physique ou d'une organisation.
Données numériques	Données dont la collecte, la mémorisation ou la présentation s'effectuent sous forme électronique et dont le traitement est, de ce fait, rendu possible par des machines.
Données ouvertes	Données ouvertes qui peuvent être utilisées, réutilisées et partagées librement par tout le monde – la seule restriction applicable étant l'obligation de citer le nom de leur auteur. (Open Data Foundation 2019)
Données personnelles	Toutes les informations qui se rapportent à une personne physique ou morale déterminée, ou aussi à une personne seulement déterminable (Datenschutzbeauftragter Kanton Zürich 2019).
Données pouvant être mises en relation avec des personnes	Les données pouvant être mises en relation avec des personnes sont des données qui, certes, ne permettent pas d'identifier sans ambiguïté ou directement une personne déterminée, mais qui permettent de constater l'identité d'une personne à l'aide d'autres informations (→ personne physique déterminable). Il peut s'agir de degrés d'intensité exceptionnels de certains attributs (par ex. si la personne concernée parle une langue rare chez elle) ou de la combinaison de nombreux attributs dont la composition, à son tour, est elle aussi très rare. (Jensen 2012, p. 14)
Données propriétaires	Les données propriétaires (également : données fermées, <i>Closed Data</i>) sont des données auxquelles ne peut faire appel que le sujet de données concerné ou l'entité chargée de les enregistrer. En règle générale, ces données sont protégées par des solutions techniques (par ex. formats de données non ouverts), conventions relatives aux licences, brevets ou droit de propriété intellectuelle. (Open Data Institute 2018)
Données relationnelles	Données qui codifient les croisements entre les sujets de données. Elles permettent de décrire les relations entre des personnes et des organisations.
Données semi-ouvertes	Les données semi-ouvertes (<i>Shared Data</i>) sont des données qui sont mises à disposition d'un cercle d'utilisateurs étendu aux fins d'utilisation, mais dont l'utilisation et la transmission à des tiers sont soumises à des restrictions plus strictes que pour les données ouvertes. (Open Data Institute 2018)
Données sensibles ou données particulièrement dignes de protection	Les données personnelles sensibles sont, d'une part, des informations pour lesquelles, en raison de leur importance, de la nature de leur traitement, ou de la possibilité de croisement avec d'autres informations, il existe un risque particulier d'infraction au droit de la protection de la personnalité (par. 3, al. 4, let. a, de la loi cantonale zurichoise sur l'information et la protection des données [IDG]) et, d'autre part, pour lesquelles des regroupements d'informations permettent une évaluation d'aspects essentiels de la personnalité d'une personne physique (ce qu'on appelle des « profils de personnalité » (par. 3, al. 4, let. b, IDG). (Datenschutzbeauftragter Kanton Zürich 2019)
Données sur le comportement	Les données sur le comportement (ou données comportementales) informent sur des activités ainsi que sur les conditions sous lesquelles ces activités peuvent être observées. On les désigne aussi par l'expression « données d'interaction ». (Meißner & Hartmann 2011)

Données sur les propriétés ou caractéristiques	Les données sur les caractéristiques fournissent des renseignements sur la situation démographique, socio-économique, physique, psychologique et politique d'une personne physique ou d'une organisation.
Educational Data Mining	L'Educational Data Mining est une discipline scientifique qui s'occupe du développement de méthodes de recherche de données provenant des institutions éducatives, et qui utilise ces méthodes pour mieux comprendre les élèves et les conditions cadres dans lesquelles ils apprennent. Il existe de fortes similitudes avec le champ de recherche « Learning Analytics ». (Baker & Inventado 2014)
Gouvernance des données (<i>Data Governance</i>)	La gouvernance des données est l'organisation et la mise en œuvre de lignes directrices, méthodes, structures, rôles et responsabilités qui esquissent et imposent les règles d'utilisation, les droits de décision et les obligations de rendre des comptes applicables à l'administration et au recours à des ensembles de données. (Ladley 2012, p. 11)
Gouvernance et gestion des processus (<i>Process Governance and Management</i>)	La gestion des processus désigne la structuration, le monitoring, le pilotage et la correction de processus, et l'encouragement à l'apprentissage axé sur les processus au sein des organisations dans le but de garantir la capacité de fonctionnement des processus et de promouvoir l'apprentissage axé sur des processus. La gouvernance en matière de processus prescrit les lignes directrices et les conditions cadres pour ces processus. Elle définit les objectifs et les principes de la gestion des processus et fixe des limites de compétences de décision, des formes d'organisation, des rôles et des responsabilités. (Paim & Flexa 2011)
Hygiène élémentaire dans la technique de l'information	Le concept d'hygiène élémentaire dans la sécurité de l'information décrit une combinaison de 10 à 20 mesures de protection fondamentales qui se sont avérées suffisantes pour la sécurité informatique d'une organisation ayant un besoin de protection normal (faible dépendance du fonctionnement quotidien de l'organisation vis-à-vis d'informations sauvegardées sous forme électronique, pas de traitement et de mémorisation systématiques de données d'importance critique pour le compte de tiers, possibilités et ressources financières limitées), mesures qui sont réalisables pour un administrateur IT au bénéfice de connaissances normales en matière de sécurité.
Individualisation	L'individualisation décrit une adaptation des méthodes de pilotage, de la répartition des ressources et des mécanismes de contrôle à la situation et aux conditions préalables propres à un acteur individuel. Le but poursuivi est de permettre à cet acteur le meilleur développement possible en direction d'un objectif bien défini.
Interopérabilité	L'interopérabilité est la possibilité d'échanger des informations à partir de deux ou de plusieurs systèmes et de les utiliser mutuellement.
Le fait de lier des données à une destination précise	Lier des données à une destination précise, c'est décrire le principe général du droit suisse de la protection des données selon lequel seul est autorisé le traitement des données effectué pour la destination précise pour laquelle elles ont été enregistrées. Cette destination peut être modifiée ou étendue par une disposition légale qui prévoit explicitement une autre utilisation ou qui prévoit le consentement de la personne concernée. (Datenschutzbeauftragter Kanton Zürich 2019)

Learning Analytics	La Society for Learning Analytics Research (« SoLAR ») définit le concept de « Learning Analytics » comme étant « le fait de mesurer, de collecter, d'analyser et de faire rapport concernant des données sur les apprenants et leurs interactions aux fins de compréhension et d'optimisation de l'apprentissage et de l'environnement où il se déroule ». Il existe de fortes similitudes avec le domaine de l'« Educational Data Mining ». (Siemens 2013)
Machine Learning (<i>Apprentissage machine</i>)	L'apprentissage machine est un champ de recherche scientifique qui est né de la combinaison entre l'informatique et la statistique. Il s'occupe de la recherche et du développement de programmes informatiques qui apprennent par expérience, c.-à-d. dont la performance lors de la résolution d'une tâche (évaluée sur la base d'un critère prédéfini) s'améliore au fur et à mesure du nombre croissant des tâches exécutées (Mitchell 2006). Une tâche de ce genre peut consister à évaluer correctement des modèles et des régularités dans des ensembles de données (→ Data Mining). Il s'agit par ex. de la détection précoce des élèves dont le passage à une classe de niveau plus élevé est menacé.
Métadonnées	Données structurées qui mettent à disposition des informations sur d'autres données. Elles servent à décrire et à documenter des ensembles de données dans le but de les rendre trouvables et comparables.
Métadonnées systémiques	Données qui prennent naissance lors de la description externe du système éducatif, de ses acteurs, de ses résultats et de ses processus. Les données qui reproduisent les opinions et les jugements de valeur de groupes d'intérêts pertinents (par ex. des médias) sur le système éducatif et ses processus en sont des exemples.
Modèle de données	Un modèle abstrait qui organise les données propres à un domaine d'application déterminé (par ex. les données du département Production, du département Comptabilité ou la totalité des données d'une entreprise) et qui décrit leurs relations mutuelles. (Lackes & Siepermann 2019)
Personne physique déterminable	Personne physique qui peut être identifiée directement ou indirectement par le biais d'un renvoi à des informations qui peuvent être déduites des circonstances ou du contexte. (TERMDAT)
Portabilité des données	On entend par « portabilité des données » (également : « transférabilité des données ») le droit conféré par le RGPD (art. 20) aux personnes physiques concernées de recevoir les données à caractère personnel les concernant qu'elles ont fournies à un responsable de traitement, dans un format structuré, couramment utilisé et lisible par machine ; elles ont le droit de transmettre ces données à un autre responsable de traitement.
Data Awareness	Par l'expression « Data Awareness », on désigne des connaissances sur des données concernant une unité (une personne, une organisation, etc.) et sur leur contenu, ainsi qu'une prise de conscience sur leurs possibilités d'utilisation et sur les risques qui leur sont liés. (Hornick 2018)
Profil de personnalité	On désigne par l'expression de « profil de personnalité » le fait de déduire et d'évaluer certaines caractéristiques d'une personne sur la base de données personnelles traitées de manière automatisée, notamment afin d'analyser ou de prévoir la prestation de travail, la situation économique, la santé, le comportement, les inclinations, le lieu de séjour ou la mobilité (<i>Profiling</i>) (TERMDAT).
Protection des données	Mesures juridiques qui servent à protéger la personnalité et les droits fondamentaux de personnes physiques dont les données sont traitées par des personnes privées ou par des organes de la Confédération. (TERMDAT)

Pseudonymisation	On parle aussi d'anonymisation de fait. Processus qui permet de séparer les données personnelles identifiantes du reste des données personnelles. L'attribution à ces deux domaines de données s'effectue par le biais d'un pseudonyme (par le biais d'un identificateur « non parlant ») qui doit être disponible aussi bien dans les données identifiantes (souvent sous la forme d'un tableau de correspondance) que dans le reste des données. Ainsi, les ayants droits peuvent à nouveau regrouper ces deux parties de données (« dépseudoymisation »/réidentification). (PFPDT 2019)
Registre	Répertoires de personnes, d'entreprises et de bâtiments/logements gérés par l'administration publique qui sont nécessaires à l'exercice des tâches relevant de la souveraineté étatique (par exemple pour le contrôle des habitants). (TERMDAT)
Sensibilité des données	Une catégorisation du besoin de protection des données définie par l'Office fédéral de la statistique (OFS). Elle décrit le potentiel de mise en danger des données pour la protection de la personnalité d'une personne physique ou morale décrite par ces données. On distingue entre les quatre niveaux suivants : (1) données factuelles, (2) données personnelles simples, (3) données personnelles qualifiées et (4) données sensibles. Les restrictions applicables au croisement des données et à leur transmission à des tiers s'accroissent avec l'augmentation du degré de protection. (OFS 2017, p. 13)
Standard de données	Les standards de données sont des conventions acceptées relatives au format, à la présentation, à la définition, à la structuration, à la manipulation, à la caractérisation, au transfert, à l'utilisation et à l'administration de données. Un standard de données décrit le contenu requis et le format dans lequel devraient être présentés et échangés certains types de données. (Baykoucheva 2015)
Standard	Un standard (on dit aussi : une norme) désigne un accord portant sur des règles, des règles, des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné. Outre les données, cela peut aussi concerner des algorithmes, des modèles, des méthodes, des processus, etc. (ISO 2019)
Sujet de données	Le sujet de données est l'unité (par ex. une personne physique) qui est décrite par les informations rassemblées au moyen de données. On le désigne aussi par l'expression « Unité d'observation », « Unité d'enregistrement », « Unité d'examen » (<i>Unit of Observation</i>) ou « Porteur de caractéristique » (art. 4, al. 1 ^{er} GDPR) (le GDPR est la version anglaise du RGPD).
Système d'apprentissage adaptatif	Les systèmes d'apprentissage adaptatifs (<i>Adaptive Learning Systems, Adaptive Learning Platforms, Intelligent Tutoring Systems</i>) sont des plateformes techniques qui sont basées la plupart du temps sur l'apprentissage machine. Elles réagissent aux interactions avec les apprenants en adaptant de manière automatisée des contenus au comportement des apprenants. (EdSurge 2016)
Utilisation minimale des données ou minimisation des données	Le principe d'utilisation minimale des données (synonymes : principe de minimisation des données, principe d'évitement des données, principe de proportionnalité) stipule que seul est autorisé le traitement des données personnelles qui sont nécessaires et appropriées au but de traitement visé. Le traitement de données superflues qui ne sont pas nécessaires à l'atteinte de ce but n'est pas autorisé. (TERMDAT)

9 Bibliographie

La bibliographie contient la littérature prise en compte dans les chapitres 1, 2, 3, 5, 6 et 7. Les articles spécialisés du chapitre 4 contiennent chacun une bibliographie.

- Abeles, V. & G. Rubinstein (2016) : *Beyond Measure: Rescuing an Overscheduled, Overtested, Underestimated Generation*, New York : Simon and Schuster.
- Alhadad, S. et al. (2015) : *The Predictive Learning Analytics Revolution : Leveraging Learning Data for Student Success*, Louisville : CO : ECAR.
- Angell, I. & S. Samonas (2009) : “The risk of computerized bureaucracy”, dans : *Journal of Information System Security* 5 (2009) : 3-25
- Arizona Departement of Education (2018) : *About AzEDS : Efficiency & Savings*. [<http://www.azed.gov/aelas/azeds/> ; 14.08.2019]
- ASSM (2019) : *Éthique de la recherche*. [<https://www.samw.ch/fr/Ethique/Ethique-de-la-recherche-live.html> ; 14.08.2019]
- Babel, J. & L. Gaillard & P. Strübli (2017) : *Bildungsperspektiven : Szenarien 2016 – 2025 für das Bildungssystem*, Neuchâtel : Bundesamt für Statistik.
- Baer, L. L. & D.M. Norris (2017) : “Unleashing the Transformative Power of Learning Analytics”, dans : Lang, C. et al. (eds.) (2017) : 309-318.
- Baker, R., & P. Inventado (2014) : “Educational Data Mining and Learning Analytics”, dans : *Learning Analytics. From Research to Practice* (2014) : 61-75.
- Baykoucheva, S. (2015) : *Managing Scientific Information and Research Data*, Waltham : Chandos Publishing.
- Beglinger, M. (2018) : «Das ist vernichtend» *Die Antworten der Bildungsforscher über die Wirkung der Schulreformen in der Schweiz sind ernüchternd*. [<https://www.nzz.ch/gesellschaft/bildungspolitik-das-ist-vernichtend-ld.1415126?reduced=true> ; 14.08.2019]
- Beynon-Davies, P. (2002) : *Information Systems - An Introduction to Informatics in Organisations*, New York : Palgrave.
- Bohl, T. et al. (2012) : *Binnendifferenzierung : Teil 1 : Didaktische Grundlagen und Forschungsergebnisse zur Binnendifferenzierung im Unterricht* (= Theorie und Praxis der Schulpädagogik. Bd. 17), Immenhausen : Prolog-Verlag.
- Borenstein, M. et al. (2009) : *Introduction to Meta-Analysis*, Chichester : Wiley & Sons.
- Calo, R. (2017) : *Artificial Intelligence Policy : A Primer and Roadmap*. [<http://dx.doi.org/10.2139/ssrn.3015350> ; 14.08.2018]
- CdC (2018) : *Lignes directrices des cantons relatives à l'administration numérique*, Berne : Conférence des gouvernements cantonaux.
- CDIP & SEFRI (2016) : *Mandat du Comité de coordination Numérisation de l'éducation (CC N)*, Berne : CDIP.

- CDIP (2016) : *Recommandations relatives à la garantie à long terme de l'accès sans examen aux hautes écoles avec une maturité gymnasiale*, Berne : Conférence suisse des directeurs cantonaux de l'instruction publique.
- CDIP (2018a) : *Décision de l'Assemblée plénière : Optimisation de l'échange des données dans la formation professionnelle ; programme optima : décision*, Berne : Conférence suisse des directeurs cantonaux de l'instruction publique.
- CDIP (2018b) : *Stratégie Numérique : Stratégie de la CDIP du 21 juin 2018 pour la gestion de la transition numérique dans le domaine de l'éducation*, Berne : Conférence suisse des directeurs cantonaux de l'instruction publique.
- CDIP (sd) : *Législation cantonale*. [<http://www.edk.ch/dyn/31059.php> ; 14.08.2019]
- CEDS (2017) : *Common Education Data Standards (CEDs) Version 7: Data Model Guide*. Common Education Data Standards. [<https://ceds.ed.gov/dataModel.aspx> ; 14.08.2019]
- Centre de compétences Fritic (2017) : *L'équipement et l'assistance informatiques dans les écoles fribourgeoises 2016*. [<https://www.fritic.ch/fr/mitic/resultats-lenquete-annuelle> ; 14.08.2019]
- Résultats de l'enquête menée en 2017
- Chancellerie fédérale (2019) : TERMDAT. [<https://www.termat.bk.admin.ch/Search/Search/language=fr> ; 14.08.2019]
- Connolly, P. & C. Keenan & K. Urbanska (2018) : "The trials of evidence-based practice in education : a systematic review of randomised controlled trials in education research 1980-2016", dans : *Educational Research*, 60 (2018) : 276-291.
- Conseil fédéral & CdC (2015) : *Convention-cadre de droit public concernant la collaboration en matière de cyberadministration en Suisse 2016 – 2019*. [<https://www.admin.ch/opc/fr/federal-gazette/2015/8805.pdf> ; 14.08.2019]
- Conseil fédéral (2017) : *Conséquences de la numérisation sur l'emploi et les conditions de travail : opportunités et risques : Rapport du Conseil fédéral donnant suite aux postulats 15.3854 Reynard du 16 septembre 2015 et 17.3222 Derder du 17 mars 2017*, Berne : Le Conseil fédéral de la Confédération suisse.
- Conseil fédéral (2018a) : *Stratégie en matière de libre accès aux données publiques en Suisse pour les années 2019 à 2023 (Stratégie Open government data, OGD)*. [<https://www.admin.ch/opc/fr/federal-gazette/2019/855.pdf> ; 14.08.2019]
- Conseil fédéral (2018b) : *Communiqué : Le Conseil fédéral approuve les bases de la stratégie suisse de cyberadministration 2020-2023*. [<https://www.admin.ch/gov/fr/accueil/documentation/communiques/communiques-conseil-federal.msg-id-72924.html> ; 14.08.2018]
- Conseil fédéral (Septembre 2018) : *Stratégie Suisse numérique*. [<https://strategy.digitaldialog.swiss/fr/> ; 14.08.2019]
- Cope, B., & Kalantzis, M. (2016) : *Big Data Comes to School : Implications for Learning, Assessment, and Research*. [<https://doi.org/10.1177/2332858416641907> ; 14.08.2019]
- Cope, B., & M. Kalantzis (2015) : "Interpreting Evidence-of-Learning : Educational research in the era of big data", dans : *Open Review of Educational Research*, 2.1 (2015) : 218-239.
- CSFO (2018) : *Rapport annuel 2018*, Berne : Centre suisse de services Formation professionnelle.
- CTIE (2012) : *Recommandations du CTIE destinées aux écoles publiques concernant la protection des données Guide pratique relatif à la transmission des données personnelles des élèves aux fournisseurs de services web*. Berne : educa.ch

- Datenschutzbeauftragter des Kantons Luzern (2009) : Merkblatt «Amtsgeheimnis». [<https://datenschutz.lu.ch/-/media/Datenschutz/Dokumente/Publikationen/dsblumerkblattamtsgeheimnisv10.pdf?la=de-CH> ; 14.08.2019]
- Datenschutzbeauftragter Kanton Zürich (2019) : *Glossar*. [<https://review.datenschutz.ch/datenschutz/global/glossary.php#0b25b30be70f496128a894957bdd60ee> ; 14.08.2019]
- Dean, C. B et al. (2012) : *Classroom Instruction That Works : Research-Based Strategies for Increasing Student Achievement*, Denver : McRel.
- Defense Advanced Research Projects Agency (2016) : «*Mayhem*» Declared Preliminary Winner of Historic Cyber Grand Challenge. [<https://www.darpa.mil/news-events/2016-08-04> ; 14.08.2019]
- Degryse, C. (2016) : *Digitalisation of the Economy and its Impact on Labour Markets. ETUI Research Paper - Working Paper 2016.02*, Brussels : European Trade Union Institute.
- Departement für Erziehung und Kultur Thurgau (2018) : *Bildungsbericht Thurgau 2018*. Frauenfeld : Departement für Erziehung und Kultur Thurgau.
- digitalswitzerland (2017) : *Manifeste digital pour la Suisse*. [<http://digital.swiss/assets/dateien/Manifest-Digital-CH-F.pdf> ; 14.08.2019]
- Döbeli Honegger, B. & M. Hielscher & W. Hartmann (2018) : *Lehrmittel in einer digitalen Welt*, Rapperswil : Interkantonale Lehrmittelzentrale.
- Eberle, F. et al. (2015) : *Basale fachliche Kompetenzen für allgemeine Studierfähigkeit in Mathematik und Erstsprache : Schlussbericht zuhanden der EDK*, Zürich : Institut für Erziehungswissenschaften UZH.
- Eckert, E. (2004) : “Individuelles Fördern”, dans : Was ist guter Unterricht? (2004) : 86–103.
- Ed-Fi Alliance (2015) : *Arizona Department of Education Successfully Launches New Data System*. [<https://www.ed-fi.org/blog/2015/11/arizona-department-of-education-successfully-launches-new-data-system/> ; 14.08.2019]
- Ed-Fi Alliance (2018) : *Success Stories*. [<https://www.ed-fi.org/success-stories/> ; 14.08.2019]
- EdSurge (2016) : *Decoding Adaptive*, London : Pearson.
- Educause (2017) : *7 Things You Should Know About Adaptive Learning*. [<https://library.educause.edu/~media/files/library/2017/1/eli7140.pdf> ; 14.08.2019]
- Ender, S. et al. (2017) : *Bildungsbericht Nordwestschweiz 2017*. Zürich : Institut für Bildungsevaluation.
- Fayyad, U. M. & G. Piatetsky-Shapiro & P. Smyth (1996) : “Knowledge Discovery and Data Mining : Towards a Unifying Framework”, dans : *Second International Conference on Knowledge Discovery and Data Mining* (1996) : 82-88.
- Fenwick, T. (2015) : “Professional responsibility in a future of data analytics”, dans : *Coding/learning : software and digital data in education* (2015) : 68-72.
- Ferguson, R. et al. (2016) : *Research Evidence on the Use of Learning Analytics : Implications for Education Policy*, Brüssel : European Commission.
- FNS (Mars 2018) : *PNR 77 : Transformation numérique*. [<http://www.nfp77.ch/fr> ; 14.08.2019]
- Frey, C. B., & M.A. Osborne (2017) : “The Future of Employment : How Susceptible Are Jobs to Computerisation?”, dans : *Technological Forecasting and Social Change* (2017) : 254–280.
- Frey, F. et al. (2018) : *Digitale Verwaltung Schweiz. Wie gelingt das Heben des Datenschatzes?* [https://www.bcg.com/Images/BCG_Digitale_Verwaltung_CH_E-Government-Strategie_tcm20-194548.pdf ; 14.08.2019]

- Frischmann, B. (2012) : *Infrastructure - The Social Value of Shared Resources*, Oxford/New York : Oxford University Press.
- Gabriel, R. & P. Gluchowski & A. Pastwa (2009) : *Data warehouse & data mining*, Witten-Herdecke : W3L-Verlag.
- Gesellschaft für Informatik (2016) : *Dagstuhl-Erklärung: Bildung in der digitalen vernetzten Welt*, Berlin : Gesellschaft für Informatik e.V.
- Gesellschaft für Informatik (2018) : *Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren. Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen*, Berlin : Sachverständigenrat für Verbraucherfragen.
- Gigerenzer, G., Müller, K.-R., & Wagner, G. G. (22. Juni 2018) : Wie man Licht in die Black Box wirft. *Frankfurter Allgemeine Zeitung*.
- Golliez, A. (2018) : *Swiss Data Space. Für eine konstruktive Datenpolitik in der Schweiz* [https://opendata-consulting.ch/wp-content/uploads/2015/08/Opendata.ch_2018_Swiss_Data_Space_20180703.pdf ; 14.08.2019]
- Graduate NYC (2014) : *Partnering for Educational Success in NYC : NYC DoE & CUNY*, New York : Graduate NYC.
- Gräsel, C. (2015) : Was ist Empirische Bildungsforschung?, dans : *Empirische Bildungsforschung: Strukturen und Methoden* (2015) : 15-30.
- Gregory, A. & P. Heus & Ryssevik, J. (2009) : *Metadata. RatSWD Working Paper Series - Working Paper No. 57.* [https://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_57.pdf ; 14.08.2019]
- Grossenbacher, S. & C. Oggenfuss (2015) : «*Information Bildungsforschung*». *40 Jahre Forschungsdokumentation der SKBF*, Aarau : Schweizerische Koordinationsstelle für Bildungsforschung.
- Häder, M. (2009) : *Der Datenschutz in den Sozialwissenschaften: Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland.* [https://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf ; 14.08.2019]
- Hall, M. (2018) : *Student wellbeing and mental health: the opportunities in learning analytics.* [<https://repository.jisc.ac.uk/6916/1/student-wellbeing-and-mental-health-the-opportunities-in-learning-analytics.pdf> ; 14.08.2019]
- Hattie, J. (2008) : *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*, London : Routledge.
- Hattie, J. (2012) : *Visible Learning for Teachers : Maximizing Impact on Learning*, London : Routledge.
- Hauser, C. et al. (2017) : *Ethische Herausforderungen für Unternehmen im Umgang mit Big Data*, Zürich : Schweizerische Akademie der Technischen Wissenschaften.
- Höhne, J. (2010) : *Verfahren zur Anonymisierung von Einzeldaten.(= Statistik und Wissenschaft, Bd. 16)*, Wiesbaden : Statistisches Bundesamt.
- Hornick, M. (2018) : *Data Science Maturity Model - Data Awareness Dimension (Part 6).* [<https://blogs.oracle.com/r/data-science-maturity-model-data-awareness-dimension-part-6> ; 14.08.2019]
- HSGYM (2014) : *Hochschulreife und Studierfähigkeit - eine Zwischenbilanz*, Zürich : HSGYM.
- Huonder, F. (2017) : *Medieninhaltsanalyse Big Data: Big Data, Datenschutz und Privatsphäre in Schweizer und US-amerikanischen Zeitungen*, Chur : Hochschule für Technik und Wirtschaft.
- IDC (2014) : *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Thing.* [<https://www.emc.com/leadership/digital-universe/2014iview/index.htm> ; 14.08.2019]

- IDC (2018) : *The Digitization of the World: From Edge to Core*. [<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> ; 14.08.2019]
- IQES online (2018) : *Basisinformationen für Schulleitungen und Lehrpersonen*. [<https://www.iqesonline.net/download.cfm?id=3a2a4b24-1517-6203-6054-f4d2634461ac> ; 14.08.2019]
- ISO (2019) : *ISO deliverables: The different types of ISO publications*. [<https://www.iso.org/deliverables-all.html> ; 14.08.2019]
- Jarchow, T. & B. Estermann (2015) : *Big Data: Chancen, Risiken und Handlungsbedarf des Bundes: Ergebnisse einer Studie im Auftrag des Bundesamts für Kommunikation*, Bern : Berner Fachhochschule.
- Jensen, U. (2012) : *Leitlinien zum Management von Forschungsdaten: Sozialwissenschaftliche Umfragedaten*, Köln : GESIS – Leibniz-Institut für Sozialwissenschaften.
- Johnson, L et al. (2015) : *Measuring Learning: An NMC Horizon Project Strategic Brief. Volume 2.2*, Austin : The New Media Consortium.
- Jovanović, J. (2017) : “Learning analytics to unveil learning strategies in a flipped classroom”, dans : *The Internet and Higher Education*, 33 (2017) : 74-85.
- Kemenetz, A. (2016) : *5 Doubts About Data-Driven Schools*. [<https://www.npr.org/sections/ed/2016/06/03/480029234/5-doubts-about-data-driven-schools> ; 14.08.2019]
- Kolderie, T., & T. McDonald (2009) : *How Information Technology Can Enable 21st Century Schools*, Washington D.C. : The Information Technology & Innovation Foundation.
- Köller, O. (2014) : “Entwicklung und Erträge der jüngeren empirischen Bildungsforschung”, dans *Zeitschrift für Pädagogik, Beiheft 60* (2014) : 102-122.
- Kolowich, S. (2013) : *The New Intelligence*. [<https://www.insidehighered.com/news/2013/01/25/arizona-st-and-knewtons-grand-experiment-adaptive-learning> ; 14.08.2019]
- Kop, R. & H. Fournier & G. Durand (2017) : “A Critical Perspective on Learning Analytics and Educational Data Mining”, dans : *Lang, C. et al. (eds.)* (2017) : 319-326.
- Krcmar, H. (2015) : *Informationsmanagement*, Heidelberg : Springer.
- Krüger, J., & K. Lischka (2018) : *Damit Maschinen den Menschen dienen: Lösungsansätze, um algorithmische Prozesse in den Dienst der Gesellschaft zu stellen*, Gütersloh : Bertelsmann Stiftung.
- Kurakin, A. & I. Goodfellow & S. Bengio (2017) : *Adversarial Machine Learning at Scale*. [<https://arxiv.org/pdf/1611.01236.pdf> ; 14.08.2019]
- Lackes, R., & M. Siepermann (2019) : *Datenmodell*. [<https://wirtschaftslexikon.gabler.de/definition/datenmodell-28093> ; 14.08.2019]
- Ladley, J. (2012) : *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*, Waltham : Elsevier.
- Lang, C. et al. (eds.) (2017) : *Handbook of Learning Analytics*. Society of Learning Analytics Research.
- Larson, J. (2016) : *How We Analyzed the COMPAS Recidivism Algorithm*. [<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> ; 14.08.2019]
- Latzer, M. et al. (2017) : *Vertrauen und Sorgen bei der Internetnutzung in der Schweiz 2017: Themenbericht aus dem World Internet Project – Switzerland 2017*, Zürich : Universität Zürich.
- Liu, R., & K.R. Koedinger (2017) : “Going Beyond Better Data Prediction to Create Explanatory Models of Educational Data” dans : *Lang, C. et al. (eds.)* (2017) : 69-76

- Livingstone, S. & M. Stoilova & R. Nandagiri (2018) : *Children's data and privacy online: Growing up in a digital age. An evidence review*, London : London School of Economics and Political Science.
- Livingstone, S. (2018) : “Children : A Special Need for Privacy”, dans : *InterMEDIA*, 46.2 (2018) : 18-23.
- Madaio, M. et al. (2018) : “A climate of support: a process - oriented analysis of the impact of rapport on peer tutoring”, dans : *Proceedings of the 12th International Conference of the Learning Sciences* (2017).
- Manyika, J. et al. (2011) : *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute.
- Manyika, J. et al. (2013) : *Open data: Unlocking innovation and performance with liquid information*, McKinsey Global Institute.
- Marris, S. (2016) : *Blog: Pädagogische Hochschule St. Gallen*. [<http://blogs.phsg.ch/grundlagenmodul/files/2016/08/Vorlesung-Grundlagenmodul-Medienmentor-PHSG-2016-Studentenfassung.pdf> ; 14.08.2019]
- Marsh, J. A. & J.F. Pane & L.S. Hamilton (2006) : *Making Sense of Data-Driven Decision Making in Education*, Santa Monica : RAND Education.
- McCarthy, J. (2007) : *What is AI? / Basic Questions*. [<http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html> ; 14.08.2019]
- Meißner, S. & J. Hartmann (2011) : “Mousetracking - Analyse und Interpretation von Interaktionsdaten”, dans : *Workshop-Proceedings der Tagung Mensch & Computer* (2011) : 133-134
- Mertens, P et al. (2017) : *Grundzüge der Wirtschaftsinformatik*, Berlin : Springer.
- Meyer, H.-D. & K. Zahedi (2014) : *An Open Letter: To Andreas Schleicher*, OECD, Paris. [<https://bildung-wissen.eu/fachbeitraege/basistexte/open-letter-to-andreas-schleicher-oecd-paris.html> ; 14.08.2019]
- Minsch, R. & R. Wehrli (2018) : “Digitalisierung – Herausforderungen und Chancen für die Schule”, dans : *economiesuisse Dossier Politik #02/2018* (2018) : 16.
- Mitchell, T. M. (2006) : *The Discipline of Machine Learning*, Pittsburgh : School of Computer Science, Carnegie Mellon University.
- Mitros, P. et al. (2017) : *Big data for monitoring educational systems*, Brüssel : European Commission.
- Murray Campbell, A. & Joseph Hoane & Feng-hsiung Hsu (2002) : “Deep Blue”, dans : *Artificial Intelligence* 134.1-2 (2002) : 57-83.
- New, J. (2016) : *Building a Data-Driven Education System in the United States*, Washington D.C. : Center for Data Innovation.
- Odermatt, S. et al. (2009) : *Datenschutz in den Volksschulen des Kanton Bern: Leitfaden (Nachschlagewerk)*, Bern : Erziehungsdirektion des Kantons Bern, Amt für Kindergarten, Volksschule und Beratung.
- OECD (2015) : *Data Driven Innovation*. [https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page1 ; 14.08.2019]
- Office of Educational Technology (2017) : *Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update*, Washington D.C. : U.S. Department of Education.
- Office of Educational Technology (2019) : *National Education Technology Plan*. [<https://tech.ed.gov/netp/> ; 14.08.2019]

- OFJ (2006) : *Loi fédérale sur le principe de la transparence dans l'administration* [<https://www.bj.admin.ch/bj/fr/home/staat/gesetzgebung/archiv/oeffentlichkeitsprinzip.html> ; 14.08.2019]
- OFS (2006) : *Wegleitung zum Datenschutz bei der Weitergabe von Einzeldaten an Dritte*. [<https://unstats.un.org/unsd/dnss/docViewer.aspx?docID=2766> ; 16.11.2018]
- OFS (2016a) : *Programme pluriannuel de la statistique fédérale 2016-2019*. [<https://www.bfs.admin.ch/bfs/fr/home/ofs/statistique-publique/systeme-statistique-publique/programme-pluriannuel.assetdetail.350521.html> ; 14.08.2019]
- OFS (2016b) : *Appariement de données pour le compte de tiers*. [<https://www.bfs.admin.ch/bfs/fr/home/services/appariement-donnees/pour-compte-tiers.html> ; 12.08.2019]
- OFS (2017) : *Directives sur l'appariement (version 1.1; 22.02.2017)*. [<https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/publikationen.assetdetail.2760553.html> ; 14.08.2019]
- OFS (2018a) : *Le système de la statistique publique*. [<https://www.bfs.admin.ch/bfs/fr/home/ofs/statistique-publique/systeme-statistique-publique.html> ; 12.08.2019]
- OFS (2018b) : *Compétences numériques*. [<https://www.bfs.admin.ch/bfs/fr/home/statistiques/culture-medias-societe-information-sport/societe-information/indicateurs-generaux/menages-population/competences-numeriques.html> ; 12.08.2019]
- OFS (2018c) : *Fiche signalétique – Enquêtes, sources : Etudiants et examens finals des hautes écoles (SHIS-studex)*. [<https://www.bfs.admin.ch/bfsstatic/dam/assets/5727901/master> ; 14.08.2019]
- Open Data Foundation (2019) : *Was ist Open Data?* [<http://opendatahandbook.org/guide/de/what-is-open-data/>; 14.08.2019]
- Open Data Institute (2018) : *The Data Spectrum*. [<https://theodi.org/about-the-odi/the-data-spectrum/> ; 14.08.2019]
- Pagnossin, E. & F. Armi & A. Matei (2014) : *Vers le monitoring de l'Espace romand de la formation*, Neuchâtel : Institut de recherche et de documentation pédagogique.
- Paim, R., & R. Flexa (2011) : *Process Governance: Definitions and Framework, Part 1*. BPTrends. [<https://www.bptrends.com/publicationfiles/11-01-2011-ART-Process%20Governance-Def%20&%20Framework-Paim-Flexa%20reviewed%20v1.pdf> ; 14.08.2019]
- Perel, M. & N. Elkin-Koren (2017) : “Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement”, dans : *Florida Law Review* 69.1 (2017) : 181-221.
- Pflaumer, P. & B. Heine & J. Hartung (2009) : *Statistik für Wirtschafts- und Sozialwissenschaften: Deskriptive Statistik*, München : Oldenburg Verlag.
- PFPDT (2019) : *Glossaire*. [<https://www.edoeb.admin.ch/edoeb/fr/home/protection-des-donnees/generallites/glossaire.html> ; 14.08.2019]
- Place, K., & M. Manley (2017) : *Case Study Uplift Education: How Can We Use the Coach to Evaluate Non-Tech Initiatives in our Schools?* [https://edtechrce.org/Resources/#case_studies ; 14.08.2019]
- Prado, J. C. & M. Á. Marzal (2013) : “Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents”, dans : *Libri*, 63.2 (2013) : 123-134.
- Prinsloo, P., & S. Slade (2017) : “Ethics and Learning Analytics: Charting the (Un)Charted”, dans : *Lang, C. et al. (eds.)* (2017) : 49-57.
- Pugliese, L. (2016) : *Adaptive Learning Systems: Surviving the Storm*. [<https://er.educause.edu/articles/2016/10/adaptive-learning-systems-surviving-the-storm> ; 14.08.2019]

- Ravaux, J.-C et al. (2018) : *Future-Proofing the Workforce: Accelerating Skills Acquisition to Match the Pace of Change*, Zürich/London/Genf : Adecco Group and the Boston Consulting Group.
- Rehäuser, J., & H. Krcmar (1996) : “Wissensmanagement im Unternehmen”, dans : *Wissensmanagement* (1996) : 1-40.
- Rudin, B. (2014) : *Tätigkeitsbericht des Datenschutzbeauftragten des Kantons Basel-Stadt für das Jahr 2013*. [<http://www.bs.ch/publikationen/dsb/taetigkeitsbericht-datenschutzbeauftragten-jahr-2013.html> ; 14.08.2019]
- Schmidt, J. (2008) : “Das Hochschulsystem der Schweiz: Aufbau, Steuerung und Finanzierung der schweizerischen Hochschulen”, dans : *Beiträge zur Hochschulforschung* 30.2 (2008) : 114-147.
- Sclater, N. (2015) : *A taxonomy of ethical, legal and logistical issues of learning analytics v1.0*. [<https://analytics.jiscinvolve.org/wp/2015/03/03/a-taxonomy-of-ethical-legal-and-logistical-issues-of-learning-analytics-v1-0/> ; 14.08.2019]
- SEFRI (2016) : *Collaboration en matière de formation Confédération – cantons : Bases communes*. [<https://www.sbf.admin.ch/sbf/fr/home/formation/l-espace-suisse-de-formation/collaboration-en-matiere-de-formation-confederation-cantons/bases-communes.html> ; 14.08.2019]
- SEFRI (2017a) : *Plan d'action dans les domaines de la formation, de la recherche et de l'innovation pour les années 2019-2020*. [<https://www.sbf.admin.ch/sbf/fr/home/le-sefri/numerisation.html> ; 14.08.2019]
- SEFRI (2017b) : *Défis de la numérisation pour la formation et la recherche en Suisse*. Berne : Secrétariat d'Etat à la formation, à la recherche et à l'innovation.
- Shield, M. (2004) : “Information Literacy, Statistical Literacy and Data Literacy”, dans : *IASSIST Quarterly*, 28.2/3 (2004) : 6-11.
- Siegel, R. & S. Wyler (2012) : *Bericht zur Schlussevaluation von IQES online im Kanton Bern*, Bern : Institut für Forschung, Entwicklung und Evaluation PHBern.
- Siemens, G. & R. Baker (2012) : “Learning analytics and educational data mining: Towards communication and collaboration”, dans : *2nd International Conference on Learning Analytics and Knowledge* (2012) : 252-254.
- Siemens, G. (2013) : “Learning analytics: The emergence of a discipline”, dans : *American Behavioral Scientist* (2013) : 1380-1400.
- Silber, H. et al. (2019) : *GESIS Survey Guidelines*. [<https://www.gesis.org/gesis-survey-guidelines/home/> ; 14.08.2019]
- Simpson, C. (2016) : *Effects of Standardized Testing on Students' Well-Being*. [https://projects.iq.harvard.edu/files/eap/files/c_simpson_effects_of_testing_on_well_being_5_16.pdf ; 14.08.2019]
- SKBF-CSRE (2018) : *Recherche en éducation*. [<http://www.skbf-csre.ch/fr/recherche-en-education/recherche-en-education/> ; 14.08.2019]
- SRV (2018) : *Verbrauchergerechtes Scoring. Gutachten des Sachverständigenrats für Verbraucherfragen*, Berlin : Sachverständigenrat für Verbraucherfragen.
- Steimen, T. (2018) : *IT-Beschaffungskonferenz 2018 – Bern. LP21 : So unterstützt der Kanton Luzern die Gemeinden mit digitaler Ausrüstung*. [https://www.digitale-nachhaltigkeit.unibe.ch/unibe/portal/fak_wi-so/a_bwl/inst_wi/abt_digital/content/e90971/e562996/e630114/e630383/e720650/e721041/e721042/ITBK2018_AHS1_ThomasSteimen_LP21KantonLuzerndigitaleAusrstung.v2 GER.pdf ; 14.08.2019]
- Steimer, T., & M. Stingelin (2014) : *eCH-0044 Datenstandard Austausch von Personenidentifikationen*. [<https://www.ech.ch/standards/52527> ; 14.08.2019]

- Tippelt, R., & B. Schmidt-Hertha (2018) : *Handbuch Bildungsforschung*, Wiesbaden : Springer Fachmedien.
- Tomasik, M. J. & S. Berger & U. Moser (2018) : “On the development of a computer-based tool for formative student assessment: Epistemological, methodological, and practical issues”, dans : *Frontiers in Psychology* 9 (2018) : Article 2245.
- Undervisningsministeriet (2018) : *Klare mål: Der er opstillet fire overordnede kvalitetsmål for erhvervsuddannelserne*. [<https://uvm.dk/erhvervsuddannelser/skoleudvikling/klare-maal/om-klare-maal> ; 14.08.2019]
- VanLehn, K. (2011) : “The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring”, dans : *Educational Psychologist*, 46.4 (2011) : 197-221.
- Verbert, K. et al. (2013) : “Learning Analytics Dashboard Applications”, dans : *American Behavioral Scientist*, 57.10 (2013) : 1500-1509.
- Vock, M., & A. Gronostaj (2017) : *Umgang mit Heterogenität in Schule und Unterricht*, Berlin : Friedrich-Ebert-Stiftung.
- Volante, L. (2012) : *School Leadership in the Context of Standards-Based Reform: International Perspectives*, Amsterdam : Springer Netherlands.
- Volante, L. (2018) : *The PISA Effect on Global Educational Governance*, New York : Routledge.
- Voss, J. (2013) : *Was sind eigentlich Daten?* [<https://libreas.eu/ausgabe23/02voss/> ; 14.08.2019]
- Wachter, S. & B. Mittelstadt & C. Russell (2018) : “Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR”, dans : *Harvard Journal of Law & Technology*, 31.2 (2018) : 841-887.
- Williamson, B. (2016) : “Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments”, dans : *Journal of Education Policy*, 31.2 (2016) : 123-141.
- Wolter, S. et al. (2018) : *L'éducation en Suisse - rapport 2018*, Aarau : Centre suisse de coordination pour la recherche en éducation.
- Zehnder, C. A. (1989) : “Datenföderalismus”, dans : *Quantitative Methoden in den Wirtschaftswissenschaften* (1989) : 229-235.
- Zeide, E. (2017) : “The Structural Consequences of Big Data-Driven Education”, dans : *Big Data*, 5.2 (2017) : 164-172.

10 Annexes

I. Exemple de hiérarchie des niveaux du droit dans le canton de Soleure

Le tableau suivant illustre les sources du droit pour le domaine scolaire et pour le domaine de la formation du canton de Soleure. Il donne comme exemples des indications générales sur la hiérarchie des niveaux du droit figurant au chapitre 3 (voir Graphique 6) et ne prétend pas à l'exhaustivité.

Bund
<p><i>Im Folgenden wurde darauf verzichtet, alle Erlasse mit Einfluss auf den Schul- und Bildungsbereich aufzuführen. Insbesondere im Bereich der Berufsbildung und der Universitäten besteht auf Bundesebene eine grosse Normendichte. (Für eine vollständige Auflistung: https://www.admin.ch/opc/de/classified-compilation/41.html#41)</i></p>
<p>Bundesverfassung</p> <p>Art. 19 Recht auf ausreichenden und unentgeltlichen Grundschulunterricht</p> <p>Art. 62 Schulwesen SR 101</p>
Kanton
<p><i>Auch im Kanton besteht für den schulischen Bereich eine grosse Normdichte (s. BGS Kanton Solothurn unter 41 Schule https://bgs.so.ch/frontend/texts_of_law). Im Folgenden wurden nur einige Rechtsgrundlagen exemplarisch aufgelistet und thematisch geordnet. Wichtig: Bei den aufgeführten Verordnungen muss zwischen regierungsrätlichen Verordnungen und Verordnungen des Departements unterschieden werden, die aufgrund einer Delegation entstanden.</i></p>
<p>Kantonsverfassung</p> <p>§§ 102-113 Kantonsverfassung betr. Kultur, Unterricht und Bildung BGS 111.1</p>
<p>Volksschule</p> <p>Volksschulgesetz BGS 413.111</p> <p>Vollzugsverordnung zum Volksschulgesetz BGS 413.121.1</p>
<p>Fachmittelschule</p> <p>Gesetz über die Fachmittelschule BGS 414.131</p> <p>Verordnung über die Fachmittelschule des Kantons Solothurn BGS 414.132</p>
<p>Mittelschule</p> <p>Mittelschulgesetz BGS 414.11</p> <p>Mittelschulverordnung BGS 414.113</p> <p>Verordnung über die gymnasialen Maturitätsschulen BGS 414.114</p>
<p>Beurteilen und Übertritt</p> <p>Laufbahnreglement für die Volksschule BGS 413.412</p> <p>Reglement über die kantonalen Schulleistungsprüfungen (Leistungschecks) BGS 413.413</p> <p>Verordnung über die Passerelle Berufsmaturität oder Fachmaturität – universitäre Hochschulen BGS 441.118</p> <p>Promotionsreglement für die Fachmittelschule BGS 414.133</p> <p>Reglement über die gymnasiale Maturitätsprüfung BGS 414.472</p>

Unterricht

Reglement über die Lektionspläne der Volksschule BGS 413.621
Verordnung über die Integration fremdsprachiger Kinder BGS 413.671
Absenzen- und Disziplinarordnung der kantonalen Mittelschulen BGS 414.481
Absenzen- und Disziplinarreglement der Kantonsschule Solothurn BGS 414.482
Absenzen- und Disziplinarreglement der Kantonsschule Olten BGS 414.483

Regionale Abkommen und Koordination

Konkordat über die Schulkoordination BGS 411.211
Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule BGS 411.214.1
Regionales Schulabkommen (RSA 2000) BGS 411.241

Grundlage konfessioneller Unterricht

Verordnung über den Religionsunterricht an den Kantonsschulen BGS 414.651

Lehrmittel

Statut der Interkantonalen Lehrmittelzentrale ilz BGS 411.271
Verordnung über die Lehrmittelkommission BGS 411.273
Geschäftsreglement Lehrmittelkommission BGS 411.274

Datenschutz

Informations- und Datenschutzgesetz BGS 114.1
Informations- und Datenschutzverordnung BGS

Weitere Rechtsquellen des Kantons

Gerade im schulischen Bereich bestehen auf kantonaler Ebene neben den «klassischen» Rechtsgrundlagen weitere Rechtsquellen. Diese treten häufig in Form von Weisungen, Reglementen und Richtlinien auf und werden teilweise nicht publiziert. Hier einige Beispiele für publizierte Rechtsquellen:

Weisung zu den Lehrmittelobligatorien an der Volksschule (https://www.so.ch/fileadmin/internet/dbk/dbk-vsa/Rechtssetzende_Erlasse/weisung_lehrmittelobligatorien.pdf)

Weisung Konfessioneller Religionsunterricht während der obligatorischen Schulzeit (https://www.so.ch/fileadmin/internet/dbk/dbk-vsa/Schulbetrieb_und_Unterricht/Religionsunterricht/weisungen_konfessioneller_ru.pdf)

Weisung Sekundarstufe I – Justierung (https://www.so.ch/fileadmin/internet/dbk/dbk-vsa/Rechtsetzende_Erlasse/Weisung-Sekundarstufe-I_Justierung.pdf)

Weisung zur Unterrichtsplanung (2005) (https://www.so.ch/fileadmin/internet/dbk/dbk-vsa/Schulbetrieb_und_Unterricht/Lehrplan/Treffpunkte/Weisungen_Unterrichtsplanung.pdf)

Weisung Dialekt / Standardsprache (2004) (https://www.so.ch/fileadmin/internet/dbk/dbk-vsa/Schulbetrieb_und_Unterricht/Lehrplan/Treffpunkte/Weisung_Dialekt_Standardsprache.pdf)

Weisung Basisschrift (2015) (https://www.so.ch/fileadmin/internet/dbk/dbk-vsa/Schulbetrieb_und_Unterricht/Lehrplan/Treffpunkte/weisung_basisschrift.pdf)

Teilweise haben auch noch weitere Publikationen Einfluss auf die Arbeit der Schulen. Beispielhaft:

Lehrplan 21 (<https://so.lehrplan.ch/>)

LehrerOffice Zeugnisanleitung (<https://www.lehreroffice.ch/wiki/doku.php?id=lehreroffice:desktop:kantone:so:zeugnisanleitung>)

Gemeinden

Einige Gemeinden haben ebenfalls noch Regelungen im Schulbereich getroffen (hier beispielhaft die Stadt Grenchen):

§ 63 Gemeindeordnung städtisches Reglement 100

Schulordnung städtisches Reglement 310

Reglement Benutzung Schulräume städtisches Reglement 312.10

II. Enquête semi-structurale par courriel

Enquête sur le traitement des données dans le cadre professionnel

Avec la numérisation, de grandes quantités de données sont également générées dans le domaine de l'éducation. Comment, quand, par qui et dans quel but ces données peuvent et doivent être utilisées est une question centrale. Outre la protection et la sécurité de la sphère privée, l'accent est de plus en plus mis sur les possibilités et le potentiel qu'offre l'utilisation de ces données. Par exemple, pour simplifier les processus et les activités administratives, pour accroître la personnalisation de l'enseignement ou pour simplifier les processus d'évaluation du matériel et des techniques pédagogiques.

En collaboration avec différents experts, educa.ch prépare actuellement un rapport à l'intention du Secrétariat d'État à l'éducation, à la recherche et à l'innovation (SEFRI) et de la Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP), qui servira de base à une future politique d'utilisation des données générées par le système éducatif. Dans ce contexte, **les lacunes et les problèmes liés à l'utilisation actuelle des données** sont également identifiés et **les exigences concernant le traitement des données** sont mises en évidence.

En tant qu'acteur important du système éducatif, vous savez comment les données sont traitées dans la pratique. À quoi ça ressemble ? Où voyez-vous un potentiel d'amélioration ? Nous vous remercions de prendre quelques minutes pour répondre aux questions ci-dessous et de contribuer à l'élaboration d'une politique d'utilisation des données qui sera bénéfique pour le système éducatif suisse.

L'utilisation des données est une question complexe et à plusieurs niveaux. La quantité de possibilités techniques et les questions juridiques et éthiques qui leur sont associées rendent difficile une compréhension rapide et complète du sujet. Afin de faciliter la discussion ci-dessous, nous avons structuré les questions selon trois perspectives.

Collecte des données : Quelles données sont collectées, traitées, stockées et utilisées par qui ?

Accès aux données : Quelles données sont mises à la disposition de qui, comment, quand et dans quel but ? Comment les données provenant de différentes sources sont-elles reliées entre elles ?

Analyse des données : Comment les données sont-elles analysées ? Comment ces analyses sont-elles utilisées ?

Veillez considérer les questions ci-dessous comme des lignes directrices générales seulement. Si d'autres informations vous semblent particulièrement importantes, veuillez les ajouter ci-dessous. Si le temps que vous avez à consacrer à ce questionnaire est limité, vous pouvez vous concentrer sur les questions concernant l'utilisation la plus importante que vous faites des données dans le cadre de votre travail quotidien. Toutes les données personnelles seront traitées confidentiellement.

Questions d'introduction :

- Sur une échelle de 1 (extrêmement insatisfait/insatisfait) à 10 (extrêmement satisfait/satisfaite), dans quelle mesure êtes-vous satisfait/satisfaite du traitement des données dans votre travail quotidien ? Veuillez noter un numéro.
- Acquisition de données :
- Accès aux données :
- Possibilités d'analyse :
- Diriez-vous que vous avez une vue d'ensemble des sources de données qui vous concernent et comment elles peuvent être utilisées ? Sur une échelle de 1 (je ne connais pratiquement aucune source de données et les utilisations possibles) à 10 (bien sûr, je connais toutes les sources de données pertinentes qui me concernent et les possibilités de leur utilisation), quelle évaluation faites-vous ?

- Aperçu des sources de données pertinentes :
- Connaissance des utilisations possibles :

A. Collecte des données :

- Dans quels contextes les données sont-elles collectées systématiquement dans votre environnement professionnel ou dans le cadre de vos activités ? De quel type de données s'agit-il, qui les recueille et, lesquelles sont collectées par voie électronique ?
- Collectez-vous des données dans le cadre de votre activité professionnelle qui sont déjà recueillies ailleurs ? Dans l'affirmative, dans quelle mesure est-ce le cas ?
- Quelle proportion de votre temps de travail ou de votre charge de travail est consacrée à la saisie des données ? Avez-vous l'impression que les coûts et les avantages sont proportionnels ?
- Selon vous, comment le processus de collecte ou de génération de données dans votre environnement professionnel devrait-il idéalement être organisé ?

B. Accès aux données :

- Utilisez-vous pour votre travail quotidien des données qui n'ont pas été collectées par vous-même ? Quelles sont vos sources de données les plus importantes et comment l'accès à celles-ci est-il organisé ?
- Mettez-vous les données collectées par vous-même à la disposition d'autres personnes ou institutions ? Dans l'affirmative, cet échange est-il normalisé ou réglementé ? Dans quelles conditions effectue-t-il ?
- Y a-t-il des données dont vous auriez besoin mais auxquelles vous n'avez pas accès ou qui ne peuvent être mises à disposition qu'avec un effort excessif (par exemple, traces d'apprentissage à partir d'une application ou d'un service en ligne) ? Pour quelles raisons n'y avez-vous pas accès (formats incompatibles, pas d'autorisation d'accès...) ? Pouvez-vous donner des exemples concrets ?
- Selon vous, comment l'accès aux données dans votre environnement professionnel pourrait-il être réglementé ou organisé de manière à ce que tous les contributeurs aient accès à toutes les données pertinentes aussi rapidement que possible ? Disposez-vous par exemple de données qui pourraient être pertinentes pour d'autres personnes ou institutions, mais qui ne sont pas actuellement utilisées par elles ?

C. Analyse des données :

- Créez-vous et/ou utilisez-vous des analyses de données dans votre travail quotidien ? Si oui, dans quel but ?
- Voyez-vous un besoin d'amélioration et d'expansion de l'analyse des données dans votre travail quotidien ? Quels sont les principaux obstacles ?

Question finale :

- Est-ce que les réponses données aux points A. Collecte des données, B. Accès aux données et C. Analyse des données justifient les notes données aux questions introductives ? Avez-vous d'autres points qu'il serait important de soulever dans ce contexte ?

III. Politique d'utilisation des données dans le secteur éducatif en Europe – enquêtes

Dans le cadre des recherches effectuées portant sur le chapitre 5, les questions suivantes ont été posées aux représentants des pays membres du réseau European Schoolnet et du groupe de travail DELTA.

Questions à SwissCore, European Schoolnet

- Does your country have mandatory guidelines on the use of data in education (including higher education) ?
- On what national provisions are these guidelines based (if any) ?
- What objectives are pursued in the use of educational data ?
- There are existing directives that deal with the question on what type of data may be processed. Are there also regulations on how these data may be processed ?

Questions au groupe de travail DELTA

Relevance of the topic :

1. Is the use of educational data currently a topic of discussion in your country ? At what level (policy, administration, science, the public) and in what context (data protection, teaching or system development, etc.) is the use of educational data discussed ?
2. Is there a strategy regarding the use of educational data or is such a strategy planned ?

Current regulation, data governance :

1. How is the use of educational data regulated in your country ?
2. At what level and by which measures/instruments are your educational data regulated (laws, guidelines, recommendations) ? Where are the responsibilities ?
3. Are there differences at the individual school levels? For what purposes may educational data be used ?
4. Which educational data do you distinguish (performance data, appraisal data, location data, etc.) ?

IV. Liste des participantes et participants au workshop du 22 février 2019

Früh	Alfred	Center for Information Technology, Society, and Law (ITSL)	Auteur article spécialisé
Kellerhals	Andreas	anciennement : Archives fédérales suisses (AFS), Département fédéral de l'intérieur (DFI)	Auteur article spécialisé
Klausing	Andreas	CDIP	Mandant
Leuthold	Martin	SWITCH, Security & Network	Auteur article spécialisés
Lubich	Hannes	Institut für Mobile und Verteilte Systeme, Hochschule für Technik FHNW	Auteur article spécialisé
Montereale	Barbara	SEFRI	Mandante
Pilloud	Marc	Nextmeme	Expert
Stürmer	Matthias	Forschungsstelle Digitale Nachhaltigkeit, Institut für Wirtschaftsinformatik	Auteure article spécialisé
Tamò-Larrieux	Aurelia	Center for Information Technology, Society, and Law (ITSL)	Auteur article spécialisés
Uttinger ²⁴⁵	Ursula	Datenschutz-Forum Schweiz	Experte
Buchser-Heer	Nelly	educa.ch	
Volland	Benjamin	educa.ch	
Wimmer	Karl	educa.ch	

²⁴⁵ N'a pas pu participer au workshop, mais a donné une évaluation écrite.