Position Paper

Digital services for science – where is the journey heading?

Digital Information Priority Initiative
of the Alliance of Science Organisations in Germany
Digital Tools and Services Working Group

Citation reference
Konrad, Uwe; Förstner, Konrad; Reetz, Johannes; Wannemacher, Klaus; Kett, Jürgen; Mannseicher, Florian (2020): Position Paper – Digital Services for Science.
Published by the Research Software Working Group as part of the Digital Information Priority Initiative of the Alliance of Science Organisations in Germany.
With the collaboration of Bornschein, Mathias; Brembs, Björn; Dandekar, Thomas; Dogan, Zeki Mustafa; Eisengräber-Pabst, Dirk; Erben-Russ, Michael; Feulner, Georg; Franke, Michael; Fritzsch, Bernadette; Fuhrmann, Jürgen; Goedicke, Michael; Janosch, Stephan

German version
For the German version of this publication, see
http://doi.org/10.5281/zenodo.4301924

English translation (this publication)
http://doi.org/10.5281/zenodo.4301947

Contents


maxresde
maxresdefault.jpg

Introduction and motivation

Due to advances in information technology, ever-increasing cross-domain networking in science and research and the need to share services and resources, distributed digital services are being increasingly used by stakeholders in research and science. This development is reinforced by the requirements imposed by funding agencies (both state actors and project funding bodies) to optimise research processes, secure results on a sustainable basis and make them fully available for use. The necessary digital transformation is accelerating this process, and research must endeavour to play a proactive role in it, taking on the challenges while maintaining digital self-determination.

In addition to infrastructural services, digital services also include scientific information services that are often created within the research communities themselves or are used by external providers. Digital services in research practice can be divided into the categories of infrastructure, collaboration and science, with scientific communities mostly requiring coordinated services platforms in all categories.

Infrastructural IT services, including authentication and authorisation services, scientific computing services, and services for research data storage, have long been the subject of diverse institutional, national and international activities and funding programmes.

The focus of this paper are scientific information services, which can include tools for collaborative work, the preparation and analysis of data and also scientific publishing services, as well as research software development services.

This position paper explores the following key questions:

- Which types of services are in demand among research actors and which of these really are subject to widespread use?
- What are the advantages and disadvantages of current use practices, and what opportunities are there to establish the necessary services in addition to or as an alternative to commercial offerings such as those provided by Google, Amazon, Microsoft, Elsevier or Digital Science, on a national and international level?
- How can scientific institutions and subject-specific communities be enabled to provide thematic services under their own responsibility, especially from the point of view of (financial) resource efficiency? In particular, this involves services that were originally developed from (often financially supported) scientific projects with a specialist context, including the software, descriptions and content to be further maintained.

The aims of the position paper are:

- Raise awareness of the requirements and necessities of developing, providing and operating digital scientific services, especially those that have emerged directly from research communities
- Examine the issue of integration in the national and international context
- Propose potential solutions and recommendations for action so as to initiate further discussion of the topic within the scientific community.

This position paper is addressed to

- Researchers and institutions using digital scientific services
- Developers and operators of digital scientific services, including institutional and commercial providers, as well as research communities and researchers offering such services
- Institutions, research organisations and funding bodies that strategically support digital science services for their communities.

Definitions

- Digital scientific service: Information technology service that provides environments, tools and solution components for the scientific work of researchers and research groups.
- Digital generic IT service: Infrastructure service for the purpose of identity and authorisation management and for the transfer, storage, processing, sharing, archiving and retrieval of data and information (e.g. high-performance computing, data management, sync&share)
- Cloud service[1]: IT paradigm that describes ubiquitous access to shared pools of configurable resources and IT services that are provided dynamically, usually via the internet. The main features of cloud services are that IT resources are offered on demand in an elastic (scalable) form and services are requested on a self-service basis. Due to their flexible use and scalable architecture, cloud services are increasingly replacing traditional IT services for many applications.
- NFDI: National Research Data Infrastructure[2] in Germany
- EOSC: European Open Science Cloud[3]

1 Cloud as an enabler for the European Commission Digital Strategy, Document Version 1.01 dated 16/05/2019, https://ec.europa.eu/info/sites/info/files/ec_cloud_strategy.pdf

2 https://www.bmbf.de/de/nationale-forschungsdateninfrastruktur-8299.html

3 https://www.eosc-portal.eu/glossary

Landscape and Setting

Background

The starting point for these considerations is the current scientific practice at university and non-university research institutions: this is characterised by a multitude of services that are often not coordinated with each other and are barely optimised in terms of long-term availability, security or data protection. Only in a few instances is it possible to speak of a sustainable service portfolio.

In practice, infrastructural digital services are offered to all users via data centres and IT departments, but often still as conventional, workstation-based installations that are usually limited to a group of users at the respective institution. However, researcher mobility and increasing cross-organisational, national and international collaboration requires services that can be shared interactively by small, dynamic groups of researchers as well as large communities. This requires technologies, procedures and policies aimed at shifting services to the internet and realising them in the form of cloud services, for example, requiring coordinated procedures for the authentication and authorisation of users. Many such cloud services are available from commercial providers, most of whom are based in the US: these are increasingly used by researchers and often fail to address aspects of data protection, sustainability and long-term cost.

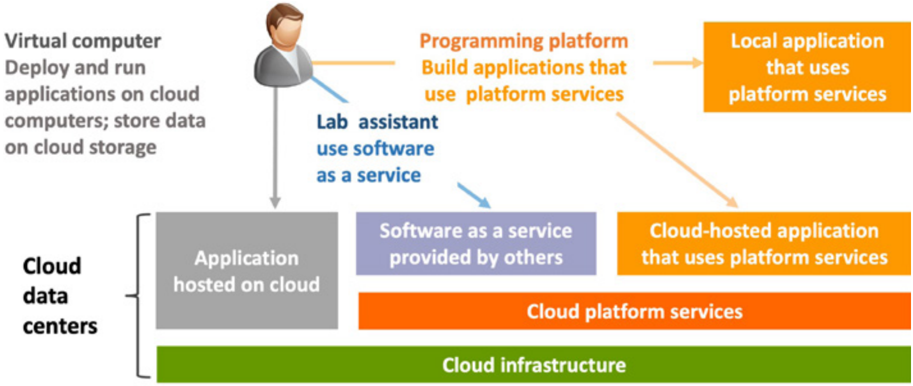

Figure 1: Use of cloud services in science[4]

Furthermore, this method of working requires a completely novel form of communication that is also in a constant state of change.

Finally, the data, digital tools and workflows used by research communities are increasingly being moved to the cloud for joint application and further development. Some large international communities in fields such as Earth observation (MOSES), climate research (Tereno) and seismology (IRIS) have been developing this working method and the services they use together for many years, while other scientific fields are still in their infancy in this regard.

What services are at issue? What are the challenges facing modern digital services and what is the current practice at research institutions and universities? In order to clarify the distinctions between the services and the nuances of the different use types, we will start by establishing categories.

4 Based on I. Foster, D.B. Gannon: Cloud Computing for Science and Engineering (Scientific and Engineering Computation), University of Chicago, 27/11/2017, ISBN: 9780262343992, https://cloud4scieng.org/

Service categories

Generic IT services

The working environment of researchers includes generic IT services, which are mostly provided by universities and research centres for their staff and guests, but increasingly also across institutes. In addition to the conventional IT services of a modern office, many cloud-based services are now also part of the researcher's toolkit, especially in research:

- Cloud storage (e.g. OwnCloud/NextCloud solutions, DropBox, S3 storage):
  Initially introduced primarily for the exchange of files via sync & share, these platforms now enable collaborative work with data and are constantly being expanded with the addition of new applications. Quotas mostly limit the volumes available. Commercial web providers such as Amazon offer on-demand storage that can not only be used interactively but can also be acquired and released by servers on demand.

- Compute power (High Performance Computing, AWS, OpenStack, Docker):
  Academic computing centres have been making computing power available to users for a very long time, mostly through applications. With an allocated quota it is possible to log on to compute clusters and start compute jobs. Today, computing power is also offered on demand in many ways, in the form of CPU computing power, virtual computers or containers. In recent years, computing on graphics cards has also become increasingly important, especially for training machine learning applications.

- Publication and data repositories (e.g. Invenio/Zenodo, DSpace, Fedora):
  In order to manage and offer research data and publications according to FAIR principles[5], these have to be archived in the appropriate repositories. The data is provided with unique identifiers (e.g. DOI) that can be referenced worldwide (e.g. via datacite.org). A great many domain-specific and general solutions have emerged for this in recent years which are to be found in the global registers of open access repositories (OpenDOAR) and research data repositories (re3data).

- Authentication and authorisation:
  In order to use these services, users have to authenticate themselves and be authorised for certain permissions. The German Research Network (DFN) operates the DFN-AAI service, which enables the DFN itself and other service operators to identify users. DFN-AAI applications include the DFN video conference Service, access to national DFG licences and the platforms DARIAH, ELIXIR and LIGO. This system was connected on a European scale via the eduGAIN project. Although the Shibboleth-based procedure was introduced as long ago as 2007, it has not yet become established worldwide. Commercial providers such as Amazon, Google or Microsoft use the OAuth procedure which is simpler and places less emphasis on security and interoperability, thereby accommodating corporate interests. The ORCID-iD allows non-proprietary, unique identification of scientific authors but cannot manage enough attributes to authorise specific permissions.

5 https://www.forschungsdaten.org/index.php/FAIR_data_principles

Services for collaborative work

Since the flood of information and data can hardly be tackled by distributing it, many users increasingly work on centrally managed data (i.e. in the cloud) with simultaneous access from different institutions.

Initially, this method of working became popular through services such as Microsoft SharePoint and Google Docs, but there are now a multitude of services and providers who are continuously expanding their portfolio of applications. Today, services that provide tailored service packages for specific target groups are particularly lucrative for teams of researchers. This is supported by the provision of integrated communication channels for video conferencing (including DFNconf, MS Teams/Skype, Zoom, Google Meet/Hangouts), persistent chats (such as Slack, Mattermost, RocketChat) and interactive websites/forums (e.g. DeepCode, Eclipse).

Two use cases can be mentioned as examples:

- Internet (Cloud) Office:
  NextCloud with OpenOffice, Google Docs, Microsoft Office365 or Authorea or Overleaf are frequently used in research to collaborate on documents. For collaborative work, this is supplemented with shared calendars, issue trackers, project management tools and much more. DropBox now also offers a complete working environment for teams with DropBox Paper and third-party applications.
- Collaborative software development
  Software development is a significant component of modern research work and increasingly takes place in distributed teams. Platforms that support collaborative software development include Microsoft GitHub (github.com), GitLab (gitlab.com) and Confluence (atlassian.com). GitLab and Confluence are often offered as self-hosted service platforms for institutes or communities (although the professional version is available for a fee). In addition to simultaneous work, these platforms support project management, version management, systematic documentation, testing and quality assurance of software. The working environment also includes exchange forums (StackOverflow, StackExchange) and environments for the resulting executable applications, which can be started and exchanged in so-called (Jupyter) notebooks (mybinder.org).

Domain-specific science services

Domain-specific services includes analytical software that carries out data processing. The norm in research at universities and institutes are specific services that address a clearly defined, subject-specific question in an internationally visible way and have only a small support staff . This category also includes repositories and databases that make curated data available and searchable. Often these functionalities are linked so that the results of a data analysis can be compared with entries in a database. These services are often created within research projects that have developed software to answer specific scientific questions. Here, local solutions are often created first and then expanded into services if the respective research community is interested. Domain-specific services require a high degree of expertise for operation and maintenance. Depending on the specialisation of the target group, such services can serve a clearly defined and therefore narrow niche (specialised software for running analyses such as RNA Analyser for regulatory elements in RNA molecules) or have considerable breadth (a portal with a wide range of databases and software such as the NCBI).

Prominent examples of such services from the life sciences are services provided by large centres or associations such as NCBI, EBI and de.NBI. These include tools like BLAST, which can be used to search for and analyse sequences of biomolecules such as DNA, RNA or proteins in databases. The European Open Science Cloud (EOSC) also offers a marketplace (see for example https://marketplace.eosc-portal.eu) that brings together numerous academic services from various institutions.

Perspectives on digital science services

The stakeholders involved have different perspectives and interests in relation to scientific information services: these are discussed in the sections below.

- Users (researchers) need an optimum working environment for themselves and their partners that is easy to use, readily available and as cost-effective as possible, as well as providing open interfaces.
- Providers of scientific information services need clear requirement profiles, sustainable resourcing and (political) support of their institution to be able to act as a service provider for their own institution as well as external partners. In addition to providers from the academic sector, commercial providers of these services are also among the stakeholders; however, the latter always represent their own interests, seeking to achieve worldwide exclusivity, strategic customer loyalty and economic success.
- Funding institutions also need clear requirement profiles combined with sustainable solution and financing concepts, which often have to be coordinated within the national and European environment, too.

Services from the user's perspective

Many research activities require the recording, processing, analysis, storage or sharing of data. This in turn requires resources and skills. Online services that provide these functionalities are therefore a key factor in facilitating everyday research and indeed making certain activities possible in the first place. Here, services that can essentially be provided by local applications are carried out much more quickly and in high quality by a powerful software or database (the service). Users trade off a degree of self-determination in exchange for this service and are dependent on the providers of the respective services. The exponential growth in the number of citations of information services is a clear indication of their growing relevance and impact on scientific results[6].

Users expect services to provide continuous availability along with easy-to-use and intuitive interfaces that offer the right balance between options and clarity. Programming interfaces (API – Application Programming Interface) are often required so as to be able to carry out requests automatically and also in large numbers. A simply constructed URL contributes to visibility. Regarding databases, the option to be able to download large data sets (bulk download) is often required. Ideally, access should not require registration. In addition to these technical aspects, the legal aspects that enable reusability are also relevant. From the user's point of view, it is desirable to have data available under licences that are as permissive as possible such as the CC0 licence[7].

6 https://galaxyproject.org/galaxy-project/statistics/

7 https://creativecommons.org/share-your-work/public-domain/cc0/

In the academic sector, the use of services is generally free of charge. Many services are offered by institutions without a payment barrier. Unlike commercial services, there is no advertising. A common problem facing domain-specific services, especially those that serve only a small group of researchers, is project-based funding and limited staffing. The problem often arises that when individuals leave an institution (e.g. after completing their doctorate), services are no longer supervised and are discontinued[8]. Due to the high mobility of research groups, services occasionally have to change institutions, too, which can result in dysfunctional links.

Good examples are the portals operated by the NCBI (National Center for Biotechnology Information) and EMBL EBIs (European Bioinformatics Institute)[9], where numerous services and databases can be found that are interlinked. Analyses can be carried out of biological macromolecules such as DNA and proteins, for example, as well as searches in various databases. The tools on which the services are based on are generally open source and – like the data – available under open licences.

8 Use it or lose it: citations predict the continued online availability of published bioinformatics resources https://doi.org/10.1093/nar/gkx182

9 https://www.ebi.ac.uk/

Services from the perspective of providers

All the services mentioned in chapter 1 are offered by different providers with varying interests and portfolios. These are

- Commercial providers
- European and national institutions
- Community-specific providers with national or even international target groups

Commercial services

The global providers of IT services, mainly based in the US, have been very successful in establishing their generic, infrastructural services in the European science system in recent years. They are temptingly easy for researchers to use:

- entry is usually free of charge and without delay,
- the services are easy to use, stable and available worldwide,
- authentication and the involvement of third parties are not a hurdle and
- increasingly, a package of helpful services is available for collaborative work (see 1.2) in addition to the generic services (see 1.1).

However, there are also serious disadvantages to using commercial services:

- Data privacy and the protection of intellectual property are not guaranteed.
- Sustainability and availability of data and functions are critical in that the conditions can change at any time (costs, functionality, target group).
- Globally operating commercial providers ensure interface compatibility within their services but rarely support open standards which would easily enable users to enrich the services themselves or even change providers.

Typical providers of such services are companies such as Google, Microsoft and Amazon as well as DropBox, Atlassian, GitLab or Elsevier. In recent years, it has become clear that partnership approaches hardly have any role to play in the field of commercial cloud services. Longstanding and as yet futile discussions with providers such as Microsoft, Elsevier and most recently also Zoom show that in the field of web services, commercial providers have little interest in openness and equal partnership. The German science system has to react to this before a dependency develops similar to that of the major science publishers.

Services provided by European and national institutions

A lot has happened in the science system at European and national level in recent years, too. The European Open Science Cloud (EOSC) envisaged by the EU Commission and the Member States is taking shape with the EOSC portal (http://eosc-portal.eu) which has an integrated service catalogue (http://marketplace.eosc-portal.eu). The EOSC aims to provide a reliable environment that enables researchers to share and analyse research data across technologies, disciplines and borders in order to increase efficiency, productivity and plausibility in research and science.

In addition to generic IT services, scientific information services are also offered via the EOSC portal. The portal enables providers to introduce their own digital services and make them accessible. The prerequisite for this is that newly added services run through an on-boarding process based on a checklist (http://providers.eosc-portal.eu/). The next step for EOSC is to develop a sustainability concept and create outreach to the huge user base, however.

Another European initiative is GAIA-X[10], a non-profit association currently consisting of 22 German and French companies and organisations. The project aims to create an open digital ecosystem to strengthen European companies and business models in global competition. This ecosystem should enable both the digital self-determination of cloud service users and the scalability of European cloud providers.

10 https://www.data-infrastructure.eu/GAIAX

The National Research Data Infrastructure (NFDI) in Germany is also still in the early stages of development and is geared towards developing a community structure. The portfolio of services with which the NFDI will position itself and whether or how it will be integrated in the EOSC has yet to be defined. The numerous institutional and also community-specific research data repositories in Germany and the existing service structures relating to information supply and data standardisation (e.g. via standards data) form a solid basis for this.

In the German federal states, the universities have developed platforms for generic services in cooperation with the German Research Network (DFN). These include the sync&share platforms sciebo in NRW, bwSync&Share in Baden-Württemberg, TU Berlin's Collab Cloud of and the LRZ sync&share cloud in Bavaria. While the services for affiliated institutions are free of charge in terms of limited basic use, external users are charged fees of approximately €1.50 to €3 per user and year plus fees for storage space.

Some of the non-university research organisations also organise central services for their research institutes. For the Max Planck Society (MPG), the Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG) in Göttingen offers various generic cloud services, including ownCloud, RocketChat, GitLab and CodiMD.

The Helmholtz Association's Incubator Initiative Information & Data Science goes one step further. This initiative supports not just pilot projects but also several service platforms designed for sustainability. One of these platforms for scientific information services is HIFIS (Helmholtz Federated IT Services). In addition to a broad portfolio of cloud services, the latter platform also provides backbone services (including basic uniform AAI services) and software services for high-quality, sustainable software development.

Services offered by community-specific providers

A large number of scientific services are established within the research communities themselves, initiated by personal, national and international initiatives or projects. Some services are able to take the step to becoming a sustainable infrastructure based on years of funding. This is exemplified by the development of TextGrid and DARIAH-DE/CLARIAH-DE into a digital infrastructure for researching textual and linguistic sources in the humanities and cultural studies. On such platforms, not only data but also software tools, publications and expert networks can be brought together to provide a virtual research environment[11].

One successful example is de.NBI (German Network for Bioinformatics Infrastructure) platform, which has been funded by the BMBF since 2013. This network of eight institutions maintains and develops nearly 100 software tools and four internationally recognised databases as well as providing cloud computing, training and other services. Among other things, it operates successfully as part of the global Galaxy project, an open, web-based platform for accessible, reproducible and transparent computational biomedical research.

However, the vast majority of scientific services are the result of individual working groups, some with a limited number of staff, for which sustainable further development and maintenance is hardly possible. In particular, the critical phase of the transition from developer to provider requires the will and support of academic institutions and funding agencies, as well as professional monitoring of software projects. A key factor during this phase is integration in an open source community and the use of professional software development processes from the outset[12].

11 Virtuelle Forschungsumgebungen – Ein Leitfaden, https://doi.org/10.2312/ALLIANZOA.026

12 Handreichung zum Umgang mit Forschungssoftware, https://doi.org/10.5281/zenodo.1172970

Funding of services

Essentially, the services provided by funding agencies are oriented either towards the needs expressed directly by the scientific disciplines or towards the integration of science policy expertise in decision-making processes on future funding lines. Scientific digital services are funded through different organisations at all levels – from local to international and from discipline-specific to generic.

The creation and further development of these services should not only benefit one working group in the medium term after initial use, however. From the point of view of the funding organisation therefore, the multiple creation of services with similar functionalities should be avoided, while the diversity of sometimes competing services should not be subjected to excessive restrictions, especially in early development phases. As such, consolidation into individual, advanced services with established user bases and stable functionality appears to be an objective for funding agencies, especially in the medium to long term. This consolidation is reflected in the fact that as soon as a service has reached a certain size/relevance, the aim is to make the transition from a niche service used (sometimes heavily) within a limited community to a supra-regional, long-term infrastructure service available to a broad community.

One example of a successful transfer to the entire community is the Protein Motif Collection founded by the scientist Amos Bairoch. After consolidation into the service PROSITE, it finally gained community-wide importance by bundling numerous such protein sequence analysis services and establishing the ExPASy (Expert Protein Analysis System) portal at the Swiss Bioinformatics Institute (SIB, Lausanne). In the case of this transfer, which came about after initial project funding, the commitment and vision of this particular researcher resulted in continuous funding from the Swiss government as part of the research infrastructure.

The desire for permanence gives rise to new needs and challenges. For example, structures for supra-regional use have to be created to ensure operation of one or more services in the long term. In this context, one key factor often mentioned is the sustainability of a service. The structure to be created through an appropriate funding process is confronted with having to establish an organisational, legal and financial framework so as to be capable of being restructured to achieve long-term sustainability. The challenges to be solved in this context, including the long-term operation of digital services across state or national borders, cannot be solved by either individual institutions or temporary funding. The more complex and cost-intensive a service is, the more evident this aspect becomes.

Funding for the establishment of scientific information services is not always successful. Many services widely used by researchers are offered by commercial providers – mostly based in the USA – on the internet. However, the use of such services involves risks in terms of data privacy and data self-determination and the non-mandatory redirection of public funds to commercial actors. One example is the service Github used by many scientific software developers which was purchased by Microsoft at the end of 2018. It is now unclear, for example, whether the new owners will maintain free use of the service. For this reason, funding agencies systematically aim to ensure the connectivity (e.g. through interfaces) and interoperability (e.g. through standards) of scientific services.

The example of Github shows that it is possible for a dependency to arise in terms of digital services offered by dominant providers (so-called vendor lock-in) in subfields or indeed in the sciences as a whole. The involvement of commercial providers can also mean that the rigorous requirements in terms of the openness and accessibility of scientific results (scientific services as a result of public funding) cannot be fully met. For this reason, funding agencies regularly reflect critically on such developments.

Discussion and recommendations for action

Digital scientific information services can greatly increase the productivity of a community in terms of efficiency and quality of research results. At the same time, the creation of a service constitutes a high-risk investment for the research group involved. Even though there are some services with enormously high numbers of users and therefore citations, however, later successful use is not guaranteed.

As such, the situation of scientific information service developers is similar to that of tech start-ups but without the underlying commercial context. Just as with start-ups, the development of a service requires venture capitalists who have to boldly promote initiatives, grant creative leeway and provide basic funding. A focus on short-term scientific output is not helpful in the development of scientific services in the initial phase.

Against the background of the open science movement and the Big Data hype, a great deal of energy and considerable resources have flowed into the establishment and expansion of research data repositories in Germany in recent years. The research data landscape is set to become consolidated in the years to come, not least under the guidance of the NFDI consortia.

As time goes on, it will be important to provide equal support for both the software needed for research and the scientific information services so as to be able to scientifically evaluate and sustainably develop this treasure trove of data. This will have a significant impact on the international standing of German research and on scientific output in general.

As the previous chapters have made clear, the landscape of scientific information services is extraordinarily diverse and subject to constant change. There are universities and research institutions that are still in their infancy in terms of (cloud-based) scientific information services, while others have already advanced considerably. The same applies to the research communities: here, too, there are very innovative groups and others that are still operating in a more conventional fashion. This position paper aims to contribute to shaping this landscape through concrete recommendations for action.

A survey conducted by the HIFIS platform among users and IT experts in the Helmholtz Association at the end of 2019 concluded that a total of around 300 services of some 50 different types were requested at the 19 Helmholtz Centres. By contrast, about 100 services were also offered internally, albeit not always sustainably. However, a single HGF survey is not representative of the entire science system in Germany, and this will have changed within just a few months.

The reasons for this are certainly manifold and might include the following, depending on the specific field of research:

- State of digitalisation
- National and international networking
- Culture and political will for change and shared services
- Financial and human resources

In any case, there is a need to catch up in terms of the quality and sustainability of research services themselves; commercial service providers of generic services have an advantage here. As set out in chapter 2, however, there are good examples of successful research services that have managed to reach large user groups as well as creating the necessary quality and sustainability. From this, recommendations can be derived to help ensure that not only individual lighthouse projects are able to overcome this hurdle but as many science services as possible.

Recommendations for action for users

As with all scientific activities, researchers should value reproducibility when using services. Ideally, analyses should be carried out using programming interfaces and the workflows should be stored so that they can be re-used. If different services are available with the same range of functions, those services guaranteeing long-term availability should be selected. Preference should also be given to services that are based on open standards, meet minimum European standards for data protection and copyright and guarantee a high degree of sustainability.

The use of services should be documented after completion of a project, including citation in publications. For many services, citations are the basis for further funding, so this is essential to their long-term existence. In addition, as everywhere else, feedback is good for service – providers of scientific services are open to constructive criticism, bug reports and suggestions. As a user, you can contribute positively to improving services in this way.

If services are subject to intense use, it makes sense to network with the service providers in order to ensure longer-term financing with them if necessary, as well as to communicate clear requirement profiles. Institutions also act as users by representing their researchers vis-à-vis the providers, entering into cooperations or commissioning external providers. As such, institutions have a considerable influence on the selection and development of scientific information services and must make the most of this to provide their scientific communities with up-to-date, sustainable and broadly networked services.

Recommendations for action for providers

In the life cycle of scientific work, researchers/research groups or infrastructure operators often become service providers. This requires intrinsic motivation but also external incentives to gradually open up one's own scientific services to a broad audience.

Such incentives can derive from positive feedback provided by a community (successful open source projects, references in publications), from career paths in the organisation (tenure track, agile teams) or from the granting of creative leeway and resources. Bold project funding can also be the starting point for a successful service, even if a business model has yet to be established.

In order to develop such incentive systems in a targeted manner, the aforementioned work (software and service development, open source projects, services for external users) has to be given visibility. In addition to publication figures and third-party funded projects, facts and figures on such activities should also appear regularly in the reports – this alone can enhance appreciation of such services!

If there is the intention and possibility to offer a service, the following questions should be clarified as early as possible:

- What is the target group and what are their requirements?
- Are there international initiatives with which development can be networked?
- What are the requirements (resources) for service provision?
- How can sustainable operation be financed and guaranteed?
- Which licences, standards and interfaces have to be observed?
- What service and support processes are needed?
- Is monitoring or accounting necessary?

For central services that are already in high demand, it is advisable to secure redundancy. NCBI and EBI complement each other in their services, for example, while GenBank and EMBL carry out a very concrete exchange and comparison of data. Other good examples are PubMed (central service) and, a later initiative, Europe PMC[13]. Users, providers and funding bodies should look carefully at where such redundancies are still lacking and where secure or complementary services would be important in order to achieve sustainability. One current example is the OMIM database (genetic diseases), which has launched a private donation initiative for this purpose. Expert associations can help identify and close such gaps in good time. Services deserve special attention that have emerged from research consortia where there is no sustainable funding but which are nevertheless important in providing the community with a service which could then suddenly disappear.

The NCBI, EMBL EBI and DDBJ have jointly established the International Nucleotide Sequence Database Collaboration[14] to harmonise services based on standards and guarantee long-term availability based on redundant storage of sequence data.

In order to consolidate the services, it has proved advantageous to offer them as a federation. In doing so, cooperating organisations take on certain services under their own responsibility and concentrate on their respective areas of expertise using resources available. This approach is successfully implemented on many platforms, including EOSC, Sciebo, de.NBI and HIFIS: it is an excellent way to use resources efficiently, especially in research and teaching.

13 https://europepmc.org/

14 http://www.insdc.org/

One increasingly important aspect that providers of digital services for science should consider is the integration of functionalities for structured knowledge representation in the form of standards data and other metadata, as well as for semantic networking of data and entities. Annotations improve the searchability of results while also promoting cross-domain and metadata-driven research approaches.

Although databases (e.g. in the form of the Linked Data Cloud) and services (Knowledge Graph, Wikidata, GND, ORCID, Annotation Services) already exist that allow annotation and (cross-domain) networking, there is still a lack of broad integration in scientific processes and services. The aim must be to establish the structured and standards-oriented annotation of research data and publications with internationally linked vocabulary as part of the research process and to create incentives for this. An important basis for this is also the development and establishment of standards and best practices (ontologies, coordination processes, etc.) in the subject-specific communities as well as across the disciplines in order to achieve standardisation and broad acceptance with regard to the annotation vocabulary. Collaborative platforms for knowledge modelling such as the Wikibase software are suitable for this. Such a service also includes organising coordination processes (working groups, committees, etc.) and establishing quality assurance mechanisms.

In the area of structured knowledge representation, too, preference should be given to offerings that allow permanently free access to data, flexible integration in services and opportunities for participation. This can be done by using existing open platforms (e.g. wikidata, GND) and by offering internal services based on open software, for example. In the long term, a machine-readable, semantic network of culture and science could be created in this way.

Recommendations for action for funding bodies

There are many ways in which funding agencies can influence the scientific IT services landscape. As shown in chapter 2, IT services are now essential building blocks of scientific work (e.g. PubMed, GitHub). In view of this, clear commitments should be made by funding agencies to recognise IT services in general and successful individual services in particular as an essential part of the scientific infrastructure. Explicit commitments should be made with the aim of enabling the permanent financial and organisational safeguarding of as many successful scientific services as possible. These commitments should ultimately help to ensure that digital services for science are not left entirely to commercial providers.

Both existing IT services and the creative creation of innovative services are indispensable for the continuous development of science. Providers and funding agencies have already done much of the groundwork and established incentives to ensure that services are used as much as possible and maintained consistently (see chapter 2). In terms of further development, however, it is crucial that a high-quality service is provided. The quality of a service should be used as a key assessment criterion for funding decisions. Furthermore, in addition to other quality standards, it is imperative that recognised standards of data protection and security of services are set, required and recognised as a second key evaluation criterion.

Widely used IT services offered by commercial providers sometimes achieve a very high degree of market penetration, also in the field of science (e.g. Google). This makes it much more difficult to establish new or little-used scientific IT services. So in addition to the often existing opportunity for innovative (new/further) development of services by individuals or small teams, a successful strategy for consolidation should also always offer opportunities to facilitate the growth of promising and successful services based on suitable funding formats. These funding formats should emphasise the following characteristics of the services right at the beginning of a funding period: use of open interfaces and standards, use of open source licences and the greatest possible openness and FAIRness in terms of user acquisition/involvement.

In addition to applying regulatory leverage, funders should also create opportunities for positive feedback by strengthening incentive systems. In particular, the involvement of users should be required by funding agencies, e.g. in order to receive regular feedback and/or community support. Best-practice examples of desirable project developments should be selected and appropriately recognised as role models (e.g. by means of prizes or by ensuring projects are given prominent mention). Incentive schemes targeting individuals should also be improved. The following are conceivable in order to achieve such improvements: the inclusion of data and software publications in the evaluation of a person's scientific work or the establishment of career paths for scientific software developers.

The development and maintenance of a coordinated, supra-regional and sustainable information infrastructure has to be achieved to a large extent through cooperative efforts. Funding bodies should therefore create suitable frameworks to strengthen exchange and cooperation in the field of digital services beyond projects and institutions at both the national and international level. The goals of the cooperative processes should be high performance, accessibility and usability of the services. For this reason, it is important for measures to be formulated, coordinated and implemented primarily from within the communities. For example, it would be conceivable for funding agencies to provide the financial framework for such negotiation processes, thereby ensuring that the organisational framework is generated from within the communities themselves. Depending on the intensity and scope of the cooperation, however, far-reaching support should be possible, too.

Summary

As digitalisation advances, the world of science is undergoing profound change. The aim of this position paper is to raise awareness of the current requirements and necessities for the development, provision and operation of digital services for science, consider the national and international embeddedness of these services and discuss potential solutions and recommendations for action. The aim for these ideas is to provide a starting point for a broad discussion within the research landscape, as has now become necessary.

In terms of the key questions referred to in the introduction, a differentiated pattern emerges with regard to the landscape of digital services used for science. Nonetheless, it is possible to identify numerous exemplary practices and tendencies that allow initial conclusions and recommendations for action. These show that it is possible and advantageous in European research to design and sustain digital services by pooling resources and focusing on the specific requirements and workflows of the scientific work itself.

Which types of services are in demand among research actors and which of these really are subject to widespread use?

For a variety of reasons, there is a significant discrepancy between services demanded by research stakeholders and those which really are subject to widespread use. Easy accessibility, performance, initial low costs and to some extent also a range of services tailored to scientific needs contribute to the pragmatic use of global commercial solutions by scientific communities. In recent years, however, many services have also emerged from the scientific infrastructure providers as well as the communities themselves, some of which have already become excellently established, acting as collaborative platforms that bring together not only data but also software tools, publications and expert networks.

What are the advantages and disadvantages of current use practices, and what opportunities are there to establish the necessary services in addition to or as an alternative to commercial offerings such as those provided by Google, Amazon, Microsoft, Elsevier or Digital Science, on a national and European scale?

While the advantages to science of using the currently widespread digital services are obvious, the use of commercial services in particular also involves a number of drawbacks, including the lack of interest on the part of commercial providers in openness and equal partnership, the frequent re-use of data for commercial interest and the risk of vendor lock-in. The possibilities in terms of establishing the necessary services in addition or as an alternative to commercial offerings within the national and European setting have by no means been fully exploited yet.

Scientific communities should respond to these drawbacks and challenges. A solution might be to build and consistently strengthen sustainable, non-commercial, cooperative structures before a dependency develops in the area of scientific information services that is similar to the dependency on major scientific publishers in the supply of scientific information.

How can scientific institutions and subject-specific communities be enabled to provide thematic services under their own responsibility, especially from the point of view of (financial) resource efficiency?

IT services in general and individual digital scientific services in particular should be recognised by funding bodies as an essential part of the scientific information infrastructure. The quality of a service should be used as a key assessment criterion for funding decisions. Benchmarks such as FAIR principles, sustainability, data protection and security for services should be set, required and recognised as further evaluation criteria. In addition to the opportunity for innovative (new/further) development of services by individuals or small teams, a successful strategy for consolidation should also offer opportunities to facilitate the sustainable creation of promising services based on suitable funding formats.

In addition to applying regulatory leverage, scientific institutions and funding bodies should also create opportunities for positive feedback by strengthening incentive systems for the development of sustainable services and software. In particular, incentives must be set based on citation and inclusion in scientific reporting as well as inclusion as a criterion for job placements and appointments. Last but not least, funding bodies should establish targeted framework conditions to strengthen exchange and cooperation at national and international level.

The currently on-going profound changes in the course of digitalisation can lead to free and democratic access to almost unlimited resources for science, but they can also result in great dependencies and upheavals. We have the capacity to influence this process! In a virtual retrospect from the year 2050 to the year 2019, Dennis Gannon describes the present time as a turning point towards a positive development:

e-Science 2050: A Look Back[15]

- In 2019, scientists were at a major turning point in terms of the technology they were able to use in science. The cloud became a huge heterogeneous online supercomputer, available on demand. Long considered a topic of purely theoretical interest, quantum computing emerged as a service in the cloud.
- One of the most remarkable features of computer science was the development of software. Programming tools had evolved into very deep stacks that used AI methods to enable scientists to achieve more with, say, a few lines of Julia in a Jupyter notebook than was possible with elaborate mainframe programming in 1980.
- The role of AI was no longer limited to programming and running e-science experiments. The first simple AI research assistants emerged that, as an intelligent system, were able to read and learn science information and answer simple (scientific) questions.

15 Based on Gannon, Dennis: eScience 2050: A Look Back, 2019/08/08, DOI 10.13140/RG.2.2.14835.07206

Imprint

Publisher

Digital Tools and Services Working Group of the Digital Information Priority Initiative (https://www.allianzinitiative.de/) of the Alliance of Science Organisations in Germany.

The members of the working group include the following (in alphabetical order): Mathias Bornschein, Björn Brembs, Thomas Dandekar, Zeki Mustafa Dogan, Dirk Eisengräber-Pabst, Georg Feulner, Konrad Förstner, Michael Franke, Bernadette Fritzsch, Jürgen Fuhrmann, Michael Goedicke, Stefan Janosch, Jürgen Kett, Uwe Konrad, Florian Mannseicher, Johannes Reetz, Klaus Wannemacher

Lead authors

Dr. Uwe Konrad
Helmholtz-Zentrum Dresden-Rossendorf e. V.
Bautzner Landstraße 400
01326 Dresden
E-Mail: u.konrad@hzdr.de
ORCID: 0000-0001-8167-9411

Jürgen Kett
Deutsche Nationalbibliothek
Adickesallee 1
60322 Frankfurt am Main
E-Mail: j.kett@dnb.de

Dr. Florian Mannseicher
Deutsche Forschungsgemeinschaft
Kennedyallee 40
53175 Bonn
E-Mail: florian.mannseicher@dfg.de
ORCID: 0000-0002-0563-1763

Prof. Dr. Konrad Förstner
ZB MED – Informationzentrum Lebenswissenschaften
Gleueler Straße 60
50931 Köln
E-Mail: foerstner@zbmed.de
ORCID: 0000-0002-1481-2996

Dr. Johannes Reetz
Max Planck Computing and Data Facility
Gießenbachstraße 2
85748 Garching
E-Mail: johannes.reetz@mpcdf.mpg.de
ORCID: 0000-0001-8183-846X

Dr. Klaus Wannemacher
HIS-Institut für Hochschulentwicklung e. V.
Goseriede 13 a
30159 Hannover
E-Mail: wannemacher@his-he.de
ORCID: 0000-0003-4810-3693

Digital Object Identifier
Original German version
http://doi.org/10.5281/zenodo.4301924

English version
http://doi.org/10.5281/zenodo.4301947

Producer
Gabriele Wicker, Fraunhofer IRB, Stuttgart

Last revised
December 2020

DFG Deutsche Forschungsgemeinschaft

Fraunhofer

HELMHOLTZ SPITZENFORSCHUNG FÜR GROSSE HERAUSFORDERUNGEN

HRK Hochschulrektorenkonferenz
Die Stimme der Hochschulen

Leibniz-Gemeinschaft

Leopoldina
Nationale Akademie der Wissenschaften

MAX-PLANCK-GESELLSCHAFT

WR | WISSENSCHAFTSRAT