

Positionspapier

Digitale Dienste für die Wissenschaft – wohin geht die Reise?

Schwerpunktinitiative Digitale Information
der Allianz der deutschen Wissenschaftsorganisationen

Arbeitsgruppe Digitale Werkzeuge und Dienste

Zitationshinweis

Konrad, Uwe; Förstner, Konrad; Reetz, Johannes; Wannemacher, Klaus; Kett, Jürgen; Mannseicher, Florian (2020): Positionspapier Digitale Dienste für die Wissenschaft.

Herausgegeben von der Arbeitsgruppe Forschungssoftware im Rahmen der Schwerpunktinitiative Digitale Information der Allianz der deutschen Wissenschaftsorganisationen.

Unter Mitarbeit von Bornschein, Mathias; Brems, Björn; Dandekar, Thomas; Dogan, Zeki Mustafa; Eisengräber-Pabst, Dirk; Erben-Russ, Michael; Feulner, Georg; Franke, Michael; Fritsch, Bernadette; Fuhrmann, Jürgen; Goedicke, Michael; Janosch, Stephan

Onlineversion

Die Onlineversion dieser Publikation finden Sie unter
<http://doi.org/10.5281/zenodo.4301924>

Englischsprachige Übersetzung

Eine englischsprachige Übersetzung finden Sie unter
<http://doi.org/10.5281/zenodo.4301947>

Inhaltsverzeichnis

Einleitung und Motivation	5
Landschaft und Handlungsumfeld	8
Hintergrund	8
Kategorien von Diensten	10
Perspektiven auf digitale Wissenschaftsdienste	14
Dienste aus der Sicht der Nutzer	15
Dienste aus der Sicht der Anbieter	17
Fördern von Diensten	21
Diskussion und Handlungsempfehlungen	23
Handlungsempfehlungen für Nutzende	25
Handlungsempfehlungen für Anbieter	26
Handlungsempfehlungen für Förderer	29
Zusammenfassung	31
Impressum	34





Einleitung und Motivation

Aufgrund des informationstechnologischen Fortschritts, der immer stärkeren, domänenübergreifenden Vernetzung in Wissenschaft und Forschung sowie der Notwendigkeit, gemeinsam Dienste und Ressourcen zu nutzen, werden von den Akteuren¹ in Forschung und Wissenschaft zunehmend verteilte digitale Dienste verwendet. Verstärkt wird diese Entwicklung durch Anforderungen der Förderorganisationen (staatliche Akteure sowie Projektförderer), Forschungsprozesse zu optimieren und die Ergebnisse nachhaltig zu sichern und nutzbar zu machen. Die notwendige digitale Transformation beschleunigt diesen Prozess, dabei muss die Forschung den Anspruch haben, diesen selbst zu gestalten und unter Wahrung der digitalen Souveränität zu bewältigen.

Zu den digitalen Diensten zählen neben infrastrukturellen Diensten auch wissenschaftliche Informationsdienste, die häufig in den Forschungs-Communities selbst entstehen oder die von externen Anbietern in Anspruch genommen werden. Die digitalen Dienste in der wissenschaftlichen Praxis können in Kategorien für die Infrastruktur, Kollaboration und Wissenschaft unterschieden werden, wobei wissenschaftliche Communities meist Plattformen von abgestimmten Diensten aller Kategorien benötigen.

Die infrastrukturellen IT-Dienste, zu denen z. B. die Dienste zur Authentifizierung und Autorisierung, für das wissenschaftliche Rechnen sowie für die Speicherung von Forschungsdaten gehören, sind seit langem Gegenstand vielfältiger institutioneller, nationaler und internationaler Aktivitäten und Förderprogramme.

Der Fokus dieser Handreichung liegt auf den wissenschaftlichen Informationsdiensten, zu denen man u.a. Werkzeuge für kollaboratives Arbeiten, für die Aufbereitung und Analyse von Daten sowie Dienste zum wissenschaftlichen Publizieren, aber auch Dienste für die Entwicklung von Forschungssoftware zählen kann.

1 Gemeint sind stets beide Geschlechter. Aus Gründen der Lesbarkeit wird in diesem Dokument auf die Nennung beider Formen verzichtet.

Dieses Positionspapier geht folgenden wesentlichen Fragen nach:

- ◆ Welche Arten von Diensten werden von den Akteuren in der Forschung gefordert und welche werden tatsächlich weitläufig genutzt?
- ◆ Welche Vor- und Nachteile ergeben sich aus der gegenwärtigen Nutzung einerseits und welche Möglichkeiten bieten sich andererseits, notwendige Dienste zusätzlich oder als Alternative zu kommerziellen Angeboten wie z. B. von Google, Amazon, Microsoft, Elsevier oder Digital Science, im nationalen und internationalen Rahmen zu etablieren?
- ◆ Wie können insbesondere unter dem Aspekt der (finanziellen) Ressourcen-Effizienz wissenschaftliche Einrichtungen und Fachcommunities in die Lage versetzt werden, **thematische Dienste** in eigener Verantwortung bereitzustellen? Dabei geht es insbesondere um Dienste, die ursprünglich aus (oft finanziell geförderten) wissenschaftlichen Projekten für einen fachlichen Kontext entwickelt wurden, einschließlich der weiter zu pflegenden Software, Beschreibungen und Inhalte.

Ziele des Positionspapiers sind:

- ◆ Bewusstsein schaffen für die Anforderungen und Notwendigkeiten der Entwicklung, Bereitstellung und des Betriebs digitaler wissenschaftlicher Dienste, vor allem solcher, die direkt aus Forschungs-Communities heraus entstanden sind
- ◆ Betrachtung der nationalen und internationalen Einbettung
- ◆ Potentielle Lösungsansätze und Handlungsempfehlungen vorschlagen und damit eine weitergehende Diskussion des Themas in der Wissenschaft initiieren.

Das vorliegende Positionspapier richtet sich an

- ◆ Forschende und Institutionen, die **digitale wissenschaftliche Dienste** nutzen
 - ◆ Entwickler und Betreiber von digitalen wissenschaftlichen Diensten, dazu gehören institutionelle und kommerzielle Anbieter, aber auch Forschungs-Communities und Forscher, die solche Dienste anbieten
 - ◆ Institutionen, Forschungsorganisationen und Förderer, die für ihre Communities digitale wissenschaftliche Dienste strategisch unterstützen.
- 

Definitionen

- ◆ **Digitaler wissenschaftlicher Dienst:** Informationstechnische Dienstleistung, die Umgebungen, Werkzeuge und Lösungskomponenten für die wissenschaftliche Arbeit von Forschenden und Forschungsgruppen bereitstellt.
- ◆ **Digitaler generischer IT-Dienst:** Infrastrukturell bereitgestellter Dienst für das Identitäts- und Berechtigungsmanagement, das Übertragen, Speichern, Verarbeiten, Teilen, Archivieren sowie Auffinden von Daten und Informationen (Bsp. High Performance Computing, Data Management, sync & share)
- ◆ **Cloud Service²:** IT-Paradigma, das den allgegenwärtigen Zugang zu gemeinsamen Pools von konfigurierbaren Ressourcen und IT-Diensten beschreibt, die dynamisch, in der Regel über das Internet bereitgestellt werden. Die Hauptmerkmale von Cloud-Services sind, dass IT-Ressourcen bei Bedarf in einer **elastischen** (skalierbaren) Form angeboten und Dienstleistungen über eine **Selbstbedienung** angefordert werden. Aufgrund der flexiblen Nutzung und der skalierbaren Architektur lösen Cloud-Services für viele Anwendungen die klassischen IT-Services zunehmend ab.
- ◆ **NFDI:** Nationale Forschungsdateninfrastruktur³ in Deutschland
- ◆ **EOSC:** European Open Science Cloud⁴

2 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

3 <https://www.bmbf.de/de/nationale-forschungsdateninfrastruktur-8299.html>

4 <https://www.eosc-portal.eu/glossary>

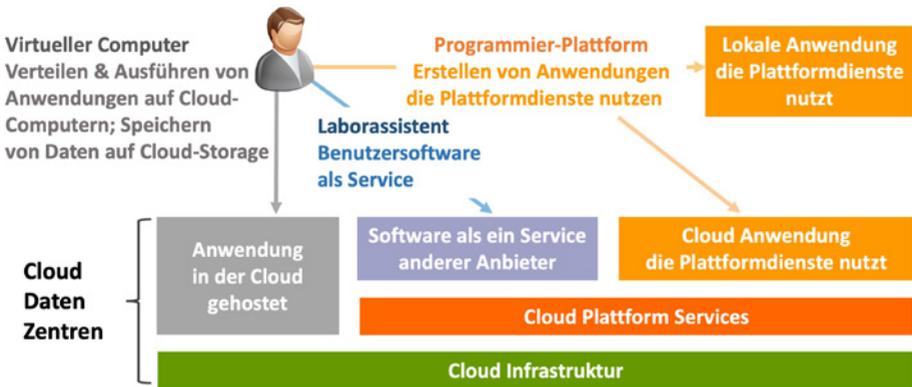
Landschaft und Handlungsumfeld

Hintergrund

Ausgangspunkt der Überlegungen ist die gelebte wissenschaftliche Praxis in den universitären und außeruniversitären Forschungseinrichtungen, die von einer Vielzahl von Diensten geprägt ist, welche oft nicht aufeinander abgestimmt und unter Aspekten der langfristigen Verfügbarkeit, der Sicherheit sowie des Datenschutzes kaum optimiert sind. Nur in wenigen Fällen kann man von einem nachhaltigen Dienstportfolio sprechen.

Infrastrukturelle digitale Dienste werden in der Praxis über die Rechenzentren und IT-Abteilungen allen Nutzern angeboten, jedoch noch oft als klassische, arbeitsplatzbezogene Installationen, die meist auf eine Nutzerschaft der jeweiligen Einrichtung beschränkt sind. Die Mobilität der Forschenden und zunehmende organisationsübergreifende, nationale und internationale Zusammenarbeit erfordert jedoch Dienste, die sowohl von kleinen, dynamischen Gruppen von Forschenden als auch von großen Communities gemeinsam interaktiv genutzt werden können. Dies erfordert Technologien, Verfahren und Richtlinien für eine Verlagerung der Dienste in das Internet, die beispielsweise in Form von Cloud-Services realisiert werden und abgestimmte Verfahren für die Authentifizierung und Autorisierung von Nutzern verlangen. Viele dieser Cloud-Services werden von kommerziellen Anbietern – meist aus den USA – bereitgestellt und von den Forschenden zunehmend genutzt, wobei Aspekte des Datenschutzes, der Nachhaltigkeit und der langfristigen Kosten oft unberücksichtigt bleiben.



Bild 1: Nutzung von Cloud-Services in der Wissenschaft⁵

Weiterhin erfordert diese Arbeitsweise eine völlig neue Art der Kommunikation, die sich ebenfalls in ständigem Wandel befindet.

Schließlich werden zunehmend auch die Daten, digitale Werkzeuge und Arbeitsabläufe (Workflows) der Forschungsgemeinschaften in die Cloud verlagert, um sie dort gemeinsam anzuwenden und weiterzuentwickeln. Große internationale Communities wie z.B. die Erdbeobachtung (MOSES), Klimaforschung (Tereno) oder Seismologie (IRIS) haben diese Arbeitsweise und von ihnen genutzte Dienste schon seit vielen Jahren gemeinsam entwickelt, andere Wissenschaftsbereiche stehen jedoch noch am Anfang.

Um welche Dienste geht es? Welche Herausforderungen sind für moderne digitale Dienste zu beachten und wie ist die gelebte Praxis in den Forschungseinrichtungen und Universitäten? Um die Unterschiede in den Diensten und die Nuancen der unterschiedlichen Nutzung zu verdeutlichen, wird zunächst eine Kategorisierung vorgenommen.

⁵ Basiert auf 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/>

Kategorien von Diensten

Generische IT-Dienste

Zur Arbeitsumgebung der Forschenden gehören generische IT-Dienste, die meist von den Hochschulen und Forschungszentren für ihre Mitarbeiter und Gäste erbracht aber zunehmend auch Instituts-übergreifend bereitgestellt werden. Neben den klassischen IT-Diensten im modernen Office gehören speziell in der Forschung mittlerweile auch viele Cloud-basierte Services zum Handwerkszeug der Forscher:

- ♦ **Cloud-Storage** (z. B. ownCloud/NextCloud Lösungen, DropBox, S3-Storage):
Zunächst vor allem zum Austausch von Dateien über sync & share eingeführt, ermöglichen diese Plattformen mittlerweile ein gemeinsames Arbeiten mit den Daten und werden um immer neue Anwendungen erweitert. Die bereitgestellten Volumina sind meist über Quoten beschränkt.
Die kommerziellen Web-Anbieter wie Amazon bieten Speicher on-demand an, der nicht nur interaktiv genutzt werden kann, sondern auch von Servern bei Bedarf akquiriert und wieder freigegeben werden kann.
- ♦ **Compute-Leistung** (High Performance Computing, AWS, OpenStack, Docker):
Schon sehr lange bieten die akademischen Rechenzentren Rechenleistung für Nutzer an, die zumeist über Anträge vergeben wurde. Mit einem zugewiesenen Kontingent kann man sich auf Compute-Clustern anmelden und Rechenjobs starten. Heute wird Rechenleistung in vielfältiger Form auch on-demand angeboten, als CPU-Rechenleistung, als virtuelle Computer oder als Container. Dabei hat in den letzten Jahren auch das Rechnen auf Grafikkarten stark an Bedeutung gewonnen, speziell auch für das Trainieren von Anwendungen des maschinellen Lernens.



- ◆ **Publikations- und Datenrepositorien** (u. a. Invenio/Zenodo, DSpace, Fedora):
Um Forschungsdaten und Publikationen nach den FAIR Prinzipien⁶ zu verwalten und anzubieten, müssen diese in entsprechenden Repositorien archiviert werden. Die Daten werden dabei mit eindeutigen Identifikatoren (z. B. DOI) ausgestattet, die weltweit referenzierbar sind (z. B. via datacite.org). Dafür sind in den letzten Jahren sehr viele domänenspezifische und allgemeine Lösungen entstanden, die in den globalen Registern von Open Access Repositorien (OpenDOAR) und Forschungsdaten-Repositorien (re3data) zu finden sind.
- ◆ **Authentifizierung und Autorisierung:**
Um diese Dienste nutzen zu können, müssen sich Nutzer authentifizieren und für bestimmte Berechtigungen autorisiert werden. Das Deutsche Forschungsnetz (DFN) betreibt den Dienst DFN-AAI, der es dem DFN selbst und weiteren Service Betreibern erlaubt, Nutzer zu identifizieren. Anwendungen der DFN-AAI sind z. B. der DFN Videokonferenz Service, Zugriffe auf Nationallizenzen der DFG und die Plattformen DARIAH, ELIXIR und LIGO. Über das Projekt eduGAIN wurde dieses System im europäischen Maßstab vernetzt. Obwohl das auf Shibboleth basierende Verfahren bereits 2007 eingeführt wurde, hat es sich weltweit bisher nicht flächendeckend durchsetzen können. Kommerzielle Anbieter wie Amazon, Google oder Microsoft nutzen das Verfahren OAuth, das einfacher ist, weniger Wert auf Sicherheit und Interoperabilität legt und somit Konzerninteressen entgegenkommt. Die ORCID-iD erlaubt eine nicht-proprietäre eindeutige Identifizierung wissenschaftlicher Autoren, kann aber nicht genügend Attribute für die Autorisierung spezieller Berechtigungen verwalten.

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

Dienste für kollaboratives Arbeiten

Da die Informations- und Datenflut durch Verteilen kaum noch beherrscht werden kann, erfolgt die Arbeit zunehmend auf zentral (in der Cloud) verwalteten Daten mit simultanem Zugriff durch viele Nutzer unterschiedlicher Einrichtungen.

Zunächst wurde diese Arbeitsweise durch Services wie Microsoft SharePoint oder Google Docs populär, mittlerweile gibt es eine Vielzahl von Angeboten und Anbietern, die ihr Portfolio an Anwendungen kontinuierlich erweitern. Für Teams von Forschenden sind heute vor allem Services lukrativ, die zugeschnittene Dienstpakete für bestimmte Zielgruppen bereit stellen. Unterstützt wird dies durch die Bereitstellung von integrierten Kommunikationskanälen für Videokonferenzen (u. a. DFNconf, MS Teams/Skype, Zoom, Google Meet/Hangouts), persistenten Chats (wie Slack, Mattermost, RocketChat) und interaktiven Webseiten/Foren (z. B. DeepCode, Eclipse).

Exemplarisch seien an dieser Stelle zwei Anwendungsfälle genannt:

- ◆ **Internet (Cloud) Office:**

Zum gemeinsamen Arbeiten an Dokumenten werden in der Forschung u. a. NextCloud mit OpenOffice, Google Docs, Microsoft Office365 oder auch Authorea bzw. Overleaf häufig verwendet. Zur Teamarbeit wird dies durch gemeinsame Kalender, Issue Tracker, Projektmanagement-Werkzeugen u. v. m. ergänzt. Auch DropBox bietet heute für Teams mit DropBox Paper und Anwendungen von Drittanbietern eine komplette Arbeitsumgebung an.

- ◆ **Kollaborative Softwareentwicklung**

Softwareentwicklung ist ein bedeutender Bestandteil moderner Forschungsarbeit und findet zunehmend in verteilten Teams statt. Plattformen, die eine kollaborative SW-Entwicklung unterstützen, sind z. B. Microsoft GitHub (github.com), GitLab (gitlab.com) oder Confluence (atlassian.com). GitLab und Confluence werden dabei oft auch als selbst gehostete (jedoch in der professionellen Version kostenpflichtige) Serviceplattformen für Institute oder Communities angeboten. Neben der simultanen Arbeit unterstützen diese Plattformen das Projektmanagement, Versionsmanagement, die systematische Dokumentation, den Test und die Qualitätssicherung von Software. Zur Arbeitsumgebung gehören auch Austauschforen (StackOverflow, StackExchange) und Umgebungen für die entstandenen ausführbaren Anwendungen, die in sog. (Jupyter-)Notebooks gestartet und ausgetauscht werden können (mybinder.org).

Domänenspezifische Wissenschaftsdienste

Bei den domänenspezifischen Diensten kann es sich um analytische Software handeln, die die Prozessierung von Daten durchführt. Spezifische Dienste, die eine klare, fachspezifische Frage international sichtbar adressieren und nur einen kleinen Mitarbeiterstab zur Betreuung haben, sind in der Forschung an Universitäten oder Instituten die Regel. Aber auch Repositorien und Datenbanken, die kuratierte Daten bereitstellen und suchbar machen, fallen in diese Kategorie. Häufig werden diese Funktionalitäten verknüpft, so dass die Ergebnisse einer Datenanalyse mit Einträgen einer Datenbank abgeglichen werden können. Diese Dienste entstehen häufig innerhalb von Forschungsprojekten, die zur Beantwortung spezifischer wissenschaftlicher Fragen Software entwickelten. Hier entstehen häufig erst einmal lokale Lösungen, die bei gegebenem Interesse der jeweiligen Forschungscommunity zu Diensten ausgebaut werden. Domänenspezifische Dienste zeichnen sich durch einen hohen Grad an nötigem Expertenwissen aus, um diese zu betreiben und zu nutzen. Je nach Spezialisierung der Zielgruppe können solche Dienste eine klar definierte und damit enge Nische bedienen (Spezialsoftware für Analysen, z.B. RNA-Analyser für regulatorische Elemente in RNA-Molekülen) oder erhebliche Breite aufweisen (ein Portal mit vielen Datenbanken und Software, z.B. das NCBI).

Prominente Beispiele für solche Dienste aus den Lebenswissenschaften sind Services, die von großen Zentren oder Verbänden wie NCBI, EBI oder de.NBI bereitgestellt werden. Diese umfassen zum Beispiel Werkzeuge wie BLAST, mit denen Sequenzen von Biomolekülen wie DNA, RNA oder Proteinen in Datenbanken gesucht und analysiert werden können. Auch die European Open Science Cloud (EOSC) bietet einen Marktplatz (siehe u.a. <https://marketplace.eosc-portal.eu>) an, auf dem zahlreiche akademische Dienste verschiedener Institutionen zusammengetragen sind.

Perspektiven auf digitale Wissenschaftsdienste

Die beteiligten Interessengruppen (Stakeholder) haben unterschiedliche Perspektiven und Interessen in Bezug auf die wissenschaftlichen Informationsdienste, die in den folgenden Abschnitten diskutiert werden.

- ◆ **Nutzer** (Forschende) benötigen ein optimales Arbeitsumfeld für sich und ihre Partner, das einfach zu bedienen, schnell verfügbar und möglichst kostengünstig ist sowie offene Schnittstellen bietet.
- ◆ **Anbieter** wissenschaftlicher Informationsdienste brauchen klare Anforderungsprofile, nachhaltige Ressourcenausstattung sowie eine (politische) Unterstützung ihrer Einrichtung, als Dienstanbieter für die eigene Einrichtung und auch externe Partner aufzutreten. Neben Anbietern aus dem akademischen Bereich zählen auch kommerzielle Anbieter dieser Dienste zu den Stakeholdern, sie vertreten jedoch stets eigene Interessen, die auf weltweite Exklusivität, strategische Kundenbindung und wirtschaftlichen Erfolg zielen.
- ◆ **Förderinstitutionen** benötigen ebenfalls klare Anforderungsprofile verbunden mit nachhaltigen Lösungs- und Finanzierungskonzepten, die darüber hinaus oft im nationalen und europäischen Umfeld abgestimmt werden müssen.



Dienste aus der Sicht der Nutzer

Viele Forschungstätigkeiten benötigen die Aufzeichnung, Bearbeitung, Analyse, Speicherung oder das Teilen von Daten. Hierfür werden Ressourcen und Kompetenzen benötigt. Online-Dienste, die diese Funktionalitäten abbilden, stellen somit eine wichtige Erleichterung des Forschungsalltags dar bzw. ermöglichen erst bestimmte Tätigkeiten. Dabei werden Leistungen, die prinzipiell durch lokale Anwendungen erbracht werden können, durch eine leistungsstarke Software oder Datenbank (den Dienst) sehr viel schneller und qualitativ hochwertig durchgeführt. Nutzer tauschen für diese Leistung ein Stück Souveränität ein und begeben sich in Abhängigkeit zu den Anbietern der jeweiligen Dienste. Die exponentiell, wachsende Zahl an Zitationen von Informationsdiensten ist eine klares Indiz für ihre wachsende Relevanz und die Auswirkung auf wissenschaftliche Ergebnisse⁷.

Nutzer erwarten von Diensten eine kontinuierliche Verfügbarkeit, gut nutzbare und intuitive Oberflächen, die die richtige Balance an Möglichkeiten und Übersichtlichkeit bieten. Häufig sind Programmierschnittstellen (API – Application Programming Interface) erwünscht, um Anfragen automatisiert durchführen zu können und somit auch größere Mengen an Anfragen durchzuführen. Eine einfach aufgebaute URL trägt zur Sichtbarkeit bei. Für Datenbanken ist die Möglichkeit des Herunterladens großer Datenbestände (Bulk Download) häufig gewünscht. Der Zugang sollte idealerweise keine Registrierung benötigen. Neben den technischen sind auch die rechtlichen Punkte, die eine Nachnutzung ermöglichen, von Relevanz. Hier ist es aus Nutzersicht wünschenswert, Daten möglichst unter permissiven Lizenzen wie der CC0-Lizenz⁸ vorzufinden.

7 <https://galaxyproject.org/galaxy-project/statistics/>

8 <https://creativecommons.org/share-your-work/public-domain/cc0/>

Im akademischen Bereich ist die Nutzung von Diensten im Allgemeinen kostenfrei. Viele Dienste werden dabei von Institutionen ohne Bezahlschranke angeboten. Im Gegensatz zu kommerziellen Angeboten wird hier nicht Werbung geschaltet. Ein häufiges Problem von domänenspezifischen Diensten, besonders solchen, die nur eine kleine Gruppe von Forschenden bedienen, ist die projektbezogene Finanzierung und geringe personelle Ausstattung. Es besteht häufig ein Problem, dass wenn einzelne Personen eine Institution verlassen (z.B. nach Abschluss der Promotion), Dienste nicht mehr betreut und eingestellt werden⁹. Auch durch die hohe Mobilität von Forschungsgruppen müssen Dienste gelegentlich die Institution wechseln, was dysfunktionalen Verlinkungen führen kann.

Gute Beispiele stellen die Portale des NCBI (National Center for Biotechnology Information) und EMBL EBIs (European Bioinformatics Institute)¹⁰ dar, unter denen sich zahlreiche Dienste und Datenbanken finden lassen, die untereinander verknüpft sind. Hier können zum Beispiel Analysen biologischer Makromoleküle wie DNA und Proteinen, aber auch Suchen in verschiedenen Datenbanken durchgeführt werden. Die Werkzeuge, auf denen die Dienste basieren, sind i.A. quell-offen und – wie auch die Daten – unter offenen Lizenzen verfügbar.

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

10 <https://www.ebi.ac.uk/>



Dienste aus der Sicht der Anbieter

Alle im Kapitel 1 genannten Dienste werden von verschiedenen Anbietern mit unterschiedlichen Interessen und unterschiedlichem Portfolio angeboten. Dies sind

- ◆ Kommerzielle Anbieter
- ◆ Europäische und nationale Institutionen
- ◆ Community-spezifische Anbieter mit nationalen oder auch internationalen Zielgruppen

Kommerzielle Dienste

Die weltweiten, vor allem aus dem amerikanischen Raum kommenden Anbieter von IT-Diensten konnten in den vergangenen Jahren sehr erfolgreich ihre generischen, infrastrukturellen Services im europäischen Wissenschaftssystem etablieren. Die Nutzung ist für Forschende verlockend einfach:

- ◆ Der Einstieg ist meist kostenlos und ohne Zeitverlust möglich,
- ◆ die Dienste sind einfach in der Handhabung, stabil und weltweit verfügbar,
- ◆ eine Authentifizierung und das Einbeziehen von Dritten stellen keine Hürde dar und
- ◆ zunehmend gibt es neben den generischen Diensten (siehe 1.1) ein Paket von hilfreichen Services für die kollaborative Arbeit (siehe 1.2).

Allerdings gibt es auch gravierende Nachteile bei der Nutzung kommerzieller Dienste:

- ◆ Datenschutz und die Wahrung intellektuellen Eigentums sind nicht gewährleistet.
- ◆ Nachhaltigkeit und Verfügbarkeit der Daten und Funktionen sind insofern kritisch, als sich Angebote jederzeit ändern können (Kosten, Funktionalität, Zielgruppe).
- ◆ Weltweit agierende kommerzielle Anbieter achten auf Schnittstellenkompatibilität innerhalb ihrer Angebote, aber unterstützen selten offene Standards, mit denen man die Dienste einfach selbst anreichern oder auch den Anbieter wechseln könnte.

Typische Anbieter solcher Dienste sind Firmen wie Google, Microsoft, Amazon aber auch DropBox, Atlassian, GitLab oder Elsevier. In den letzten Jahren hat sich deutlich gezeigt, dass partnerschaftliche Ansätze auf dem Gebiet von kommerziellen Cloud-Services kaum eine Rolle spielen. Die jahrelangen und bisher vergeblichen Diskussionen u. a. mit Microsoft, Elsevier und zuletzt auch der Firma zoom zeigen, dass auf dem Gebiet der Web Services kommerzielle Anbieter kaum ein Interesse an Offenheit und gleichberechtigter Partnerschaft haben. Darauf muss das deutsche Wissenschaftssystem reagieren, bevor eine ähnliche Abhängigkeit entsteht, wie sie bei den großen Wissenschaftsverlagen entstanden ist.

Dienste europäischer und nationaler Institutionen

Auch im Wissenschaftssystem auf europäischer und nationaler Ebene hat sich in den letzten Jahren viel getan. Die von der EU-Kommission und den Mitgliedsstaaten angestrebte Europäische Cloud für eine offene Wissenschaft (**European Open Science Cloud, EOSC**) nimmt mit dem EOSC-Portal (<http://eosc-portal.eu>) und einem darin integrierten Servicekatalog (<http://marketplace.eosc-portal.eu>) Gestalt an. Die EOSC soll eine zuverlässige Umgebung bereitstellen, die es Forschenden über Technologien, Fachgebiete und Grenzen hinweg ermöglicht, Forschungsdaten auszutauschen und zu analysieren, um die Effizienz, Produktivität und Nachvollziehbarkeit in der Forschung und Wissenschaft zu erhöhen.

Neben den generischen IT-Diensten werden auch wissenschaftliche Informationsdienste über das EOSC-Portal angeboten. Das Portal ermöglicht Anbietern, eigene digitale Dienste einzubringen und zugänglich zu machen. Voraussetzung ist, dass neu hinzukommende Dienste anhand einer Checkliste einen **onboarding**-Prozess durchlaufen (<http://providers.eosc-portal.eu/>). Im nächsten Schritt muss die EOSC jedoch ein Nachhaltigkeitskonzept entwickeln und die Reichweite zu der riesigen Nutzerbasis schaffen.

Eine weitere europäische Initiative ist **GAIA-X**¹¹, ein non-profit Verein aus derzeit 22 deutschen sowie französischen Unternehmen und Organisationen. Das Projekt soll ein offenes digitales Ökosystem schaffen, um europäische Unternehmen und Geschäftsmodelle im globalen Wettbewerb zu stärken. Dieses Ökosystem soll sowohl die digitale Souveränität der Nutzer von Cloud-Diensten als auch die Skalierbarkeit der europäischen Cloud-Anbieter ermöglichen.

11 <https://www.data-infrastructure.eu/GAIA-X>



Auch die [Nationale Forschungsdateninfrastruktur \(NFDI\)](#) in Deutschland steht noch am Anfang der Entwicklung, sie ist auf die Entwicklung einer Community-Struktur ausgerichtet. Mit welchem Portfolio an Diensten sich die NFDI positioniert und ob bzw. wie eine Integration in die EOSC erfolgt, muss noch definiert werden. Die zahlreichen institutionellen aber auch Community-spezifischen Forschungsdaten-Repositorien in Deutschland und die bestehenden Servicestrukturen rund um die Informationsversorgung und die Datenstandardisierung (z.B. über Normdaten) bilden dafür eine solide Basis.

In den Ländern wurden durch die Hochschulen in Zusammenarbeit mit dem Deutschen Forschungsnetz (DFN) Angebote für generische Dienste entwickelt. Dazu gehören z.B. die sync&share Plattformen [sciebo](#) in NRW, [bwSync&Share](#) in Baden-Württemberg, die [Collab Cloud](#) der TU Berlin oder die [LRZ sync&share-Cloud](#) in Bayern. Während die Angebote für angeschlossene Einrichtungen in der begrenzten Basisnutzung kostenlos sind, werden für [externe Nutzer](#) Gebühren von ca. 1,5 bis 3 € pro Nutzer und Jahr zuzüglich Gebühren für Speicherplatz verlangt.

Auch die außeruniversitären Forschungsorganisationen organisieren zum Teil zentrale Services für ihre Forschungsinstitute. Für die Max-Planck-Gesellschaft (MPG) bieten die Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG) und die Max Planck Computing and Data Facility (MPCDF) verschiedene generische Cloudservices an, u. a. [ownCloud](#), [RocketChat](#), [GitLab](#) oder [CodiMD](#).

Noch einen Schritt weiter geht die [Incubator-Initiative Information & Data Science](#) der Helmholtz-Gemeinschaft. Mit dieser Initiative werden nicht nur Pilotprojekte gefördert, sondern auch mehrere auf Nachhaltigkeit angelegte Serviceplattformen. Eine dieser Plattformen für wissenschaftliche Informationsdienste ist [HIFIS](#) (Helmholtz Federated IT Services). Mit dieser Plattform werden neben einem breiten Portfolio von Cloud-Services auch Backbone-Services (u.a. Basisdienste einer einheitlichen AAI) sowie Software-Services für eine qualitativ hochwertige, nachhaltige Softwareentwicklung bereitgestellt.

Dienste Community-spezifischer Anbieter

Eine große Anzahl an wissenschaftlichen Diensten entstehen in den Forschungs-Communities selbst, initiiert durch persönliche, nationale und internationale Initiativen oder Projekte. Einzelne Dienste schaffen durch langjährige Förderung den Sprung zu einer nachhaltigen Infrastruktur. Exemplarisch zeigt dies die Entwicklung von TextGrid bzw. DARIAH-DE/CLARIAH-DE zu einer digitalen Infrastruktur für die Erforschung textlicher und sprachlicher Quellen der Geistes- und Kulturwissenschaften. In solchen Plattformen können nicht nur Daten, sondern auch Softwarewerkzeuge, Publikationen und Expertennetzwerke zusammengeführt werden und stellen dann virtuelle Forschungsumgebungen¹² dar.

Ein erfolgreiches Beispiel ist die Plattform de.NBI (German Network for Bioinformatics Infrastructure), die seit 2013 durch das BMBF gefördert wird. Das Netzwerk von acht Einrichtungen pflegt und entwickelt fast 100 Software-Tools, vier international anerkannte Datenbanken und bietet Cloud-Computing, Training und weitere Services. Es arbeitet u.a. erfolgreich in dem weltweiten Projekt Galaxy, einer offenen, webbasierten Plattform für zugängliche, reproduzierbare und transparente rechnergestützte biomedizinische Forschung.

Die überwiegende Anzahl an wissenschaftlichen Diensten ist jedoch das Ergebnis einzelner Arbeitsgruppen mit z.T. wenigen Mitarbeitern, für die eine nachhaltige Weiterentwicklung und Pflege kaum möglich ist. Besonders die kritische Phase des Übergangs vom **Entwickler** zum **Anbieter** erfordert den Willen und die Unterstützung der akademischen Einrichtungen und Förderorganisationen sowie eine professionelle Begleitung der Softwareprojekte. Wertvoll sind in dieser Phase die Einbettung in eine open source Community und die Nutzung professioneller Softwareentwicklungsprozesse von Beginn an¹³.

12 Virtuelle Forschungsumgebungen – Ein Leitfaden, <https://doi.org/10.2312/ALLIANZOA.026>

13 Handreichung zum Umgang mit Forschungssoftware, <https://doi.org/10.5281/zenodo.1172970>

Fördern von Diensten

Grundsätzlich orientieren sich die Angebote von Förderorganisationen entweder an dem aus den wissenschaftlichen Disziplinen direkt geäußerten Bedarf oder an der Einbindung wissenschaftspolitischer Expertise in Entscheidungsprozesse zu zukünftigen Förderlinien. Die Förderung wissenschaftlicher digitaler Dienste findet durch unterschiedliche Organisationen auf allen Ebenen statt: von lokalen bis internationalen Diensten und von disziplinspezifischen bis generischen Diensten.

Die Neu- und Weiterentwicklung dieser Dienste soll nach anfänglicher Erstützung jedoch mittelfristig nicht nur einer Arbeitsgruppe zu Gute kommen. Die mehrfache Erzeugung von Diensten mit ähnlichen Funktionalitäten sollte aus Sicht von Förderorganisation deshalb zwar grundsätzlich vermieden, die Diversität von mitunter konkurrierenden Diensten insbesondere in frühen Entwicklungsphasen jedoch nicht zu sehr beschränkt werden. Eine Konsolidierung auf einzelne, fortgeschrittene Services mit etablierten Nutzerbasen und stabiler Funktionalität erscheint deshalb vor allem mittel- bis langfristig für Förderorganisationen als Förderziel zu bestehen. Diese Konsolidierung zeigt sich darin, dass, sobald ein Dienst eine bestimmte Größe/Relevanz erreicht hat, in der Regel die Überführung von einem in einer Community (mitunter auch stark) genutzten Nischendienst zu einem überregionalen, langfristig verfügbaren Infrastrukturdienst für eine gesamte Community angestrebt wird.

Ein Beispiel für eine gelungene Überführung auf die gesamte Community ist die durch den Wissenschaftler Amos Bairoch begründete Protein-Motif-Sammlung. Nach der Konsolidierung zum Dienst [PROSITE](#) erlangte er schließlich Communityweite Bedeutung durch Bündelung vieler solcher Proteinsequenzanalysedienste und durch Etablierung des Portals [ExPASy](#) (Expert-Protein-Analysis-System) am Schweizer Bioinformatik-Institut (SIB, Lausanne). Bei diesem Transfer kam es durch den Einsatz und die Vision des genannten Forschers nach Projektförderungen zu einer kontinuierlichen Förderung durch den Schweizer Staat als Teil der Forschungsinfrastruktur.

Mit dem Wunsch zur Verstetigung ergeben sich neue Bedarfe und Herausforderungen. Zum Beispiel müssen Strukturen für eine überregionale Nutzung geschaffen werden, die den Betrieb eines oder mehrerer Dienste langfristig sicherstellt. In diesem Zusammenhang wird häufig von der Nachhaltigkeit eines Dienstes gesprochen. Die durch einen geeigneten Förderprozess zu schaffende Struktur sieht sich damit konfrontiert, einen organisatorischen, rechtlichen und finanziellen Rahmen für eine Umstrukturierung auf Nachhaltigkeit schaffen zu müssen. Die in diesem Zusammenhang zu lösenden Herausforderungen, wie z. B. der langfristige Betrieb von digitalen Diensten über (Bundes-)Ländergrenzen, kann weder durch einzelne Einrichtungen noch durch zeitlich begrenzte Fördermittel gelöst werden. Dieser Aspekt wird umso deutlicher, je komplexer und kostenintensiver ein Dienst ist.

Nicht immer gelingt die Förderung zum Aufbau wissenschaftlicher Informationsdienste. Viele von Forschenden in breitem Maße genutzte Dienste werden von kommerziellen Anbietern – meist aus den USA – im Internet bereitgestellt. Diese Nutzung birgt jedoch Gefahren bezüglich des Datenschutzes und der Datenhoheit und der nicht zwingend notwendigen Umwidmung von öffentlichen Mitteln auf kommerzielle Akteure. Ein Beispiel stellt der von sehr vielen wissenschaftlichen Softwareentwicklern genutzte Dienst github dar. Dieser wurde Ende 2018 durch Microsoft gekauft und kann somit im besten Fall nicht mehr als gemeinnützig angesehen werden. Unklar ist nun z. B., ob die neuen Eigentümer die kostenlose Nutzung des Dienstes beibehalten werden. Förderorganisationen zielen deshalb systematisch darauf, Anschlussfähigkeit (z. B. durch Schnittstellen) und Interoperabilität (z. B. durch Standards) der wissenschaftlichen Dienste sicherzustellen.

Das Beispiel github zeigt, dass möglich ist, dass in Teilbereichen oder den gesamten Wissenschaften eine Abhängigkeit von digitalen Diensten dominanter Anbieter entsteht (sog. vendor-lock-in). Die Einbindung kommerzieller Anbieter kann dann auch zur Folge haben, dass die hohen Ansprüche der Offenheit und Zugänglichkeit von wissenschaftlichen Ergebnissen (wissenschaftliche Dienste als Ergebnis einer öffentlich-rechtlichen Förderung) nicht in vollem Umfang erfüllt werden können. Diese Entwicklungen werden deshalb von Förderorganisationen regelmäßig kritisch reflektiert.



Diskussion und Handlungsempfehlungen

Digitale wissenschaftliche Informationsdienste können die Produktivität einer Community mit Blick auf Effizienz und Qualität der Forschungsergebnisse enorm erhöhen. Gleichzeitig stellt die Erstellung eines Services eine risikobehaftete Investition der generierenden Forschungsgruppe dar. Aber wenngleich es einige Services mit enorm hohen Nutzer- und damit verknüpften Zitationszahlen gibt, ist eine spätere, erfolgreiche Nutzung nicht garantiert.

So ähnelt die Situation der Entwickler wissenschaftlicher Informationsdienste derer von Tech-Startups, allerdings ohne den kommerziellen Ansatz. Genau wie bei den Startups erfordert die Entwicklung eines Dienstes [Risiko-Kapitalgeber](#), die Initiativen mutig fördern, Gestaltungsspielräume einräumen sowie eine Basisausstattung bereitstellen müssen. Eine Fokussierung auf kurzfristigen wissenschaftlichen Output ist bei der Entwicklung von wissenschaftlichen Diensten in der Anfangsphase nicht hilfreich.

Im Zuge der Open-Science-Bewegung und des Big-Data-Hypes sind in den letzten Jahren in Deutschland sehr viel Energie und erhebliche Ressourcen in den Auf- und Ausbau von Forschungsdaten-Repositoryn geflossen. Die Forschungsdaten-Landschaft wird sich in den kommenden Jahren konsolidieren, sicherlich auch mit Steuerung durch die NFDI-Konsortien.

In den nächsten Jahren wird es darauf ankommen, auch die für die Forschung notwendige Software und die wissenschaftlichen Informationsdienste in ähnlichem Maße zu fördern, um diesen Datenschatz wissenschaftlich auswerten und nachhaltig weiterentwickeln zu können. Dies wird eine wesentliche Auswirkung auf die Stellung der deutschen Forschung im internationalen Kontext und den wissenschaftlichen Output im Allgemeinen haben.

Wie in den vorangegangenen Kapiteln deutlich wurde, ist die Landschaft der wissenschaftlichen Informationsdienste außerordentlich vielfältig und in stetiger Veränderung. Es gibt Hochschulen und Forschungseinrichtungen, die auf dem Gebiet der (Cloud-)basierten wissenschaftlichen Informationsdienste noch am Anfang stehen, andere sind schon sehr weit. Gleiches gilt für die Forschungs-Communities, auch hier gibt es sehr innovative Gruppen und andere, die noch eher klassisch arbeiten. Dieses Positionspapier hat den Anspruch, zur Gestaltung dieser Landschaft durch konkrete Handlungsempfehlungen beizutragen.

Eine Umfrage der Plattform HIFIS bei Nutzern und IT-Experten in der Helmholtz-Gemeinschaft Ende 2019 kam zu dem Ergebnis, dass in den 19 Helmholtz-Zentren insgesamt ca. 300 Services nachgefragt wurden, davon ca. 50 verschiedene Arten. Demgegenüber wurden auch ca. 100 Services selbst angeboten, wenn auch nicht immer nachhaltig. Eine Umfrage in der HGF ist allerdings nicht repräsentativ für das gesamte Wissenschaftssystem in Deutschland, und schon in wenigen Monaten wird sich dies verändert haben.

Die Ursachen dafür sind sicher vielfältig und können für das jeweilige Forschungsgebiet z. B. sein:

- ◆ Stand der Digitalisierung
- ◆ Nationale und internationale Vernetzung
- ◆ Kultur und politischer Wille zur Veränderung und für [shared Services](#)
- ◆ Finanzielle und personelle Ausstattung

Nachholbedarf besteht auf jeden Fall bei der Qualität und Nachhaltigkeit von Angeboten der Forschung selbst, hier sind kommerzielle Diensteanbieter für generische Dienste im Vorteil. Aber wie im Kapitel 2 exemplarisch erläutert, gibt es gute Beispiele für erfolgreiche Services der Forschung, die es geschafft haben, große Nutzergruppen zu erreichen sowie die notwendige Qualität und Nachhaltigkeit zu schaffen. Daraus lassen sich Empfehlungen ableiten, die helfen sollen, damit nicht nur einzelne Leuchtturmprojekte, sondern möglichst viele Dienste der Wissenschaft diese Hürde meistern können.

Handlungsempfehlungen für Nutzende

Wie bei allen wissenschaftlichen Tätigkeiten sollten Forscher bei der Nutzung von Diensten Wert auf Reproduzierbarkeit legen. Idealerweise sollten dazu Analysen durch Nutzung von Programmierschnittstellen durchgeführt werden und die Workflows nachnutzbar abgelegt werden. Sollte es verschiedene Services mit gleichem Funktionsumfang geben, sollten solche Services gewählt werden, bei denen eine langfristige Verfügbarkeit gewährleistet ist. Auch sollten solche Dienste bevorzugt werden, die sich an offenen Standards orientieren, europäischen Mindeststandards für Datenschutz und Urheberrecht genügen und ein hohes Maß an Nachhaltigkeit verbürgen.

Die Nutzung von Diensten sollte nach Abschluss eines Projektes inklusive Zitation in Publikationen dokumentiert werden. Zitationen sind für viele Dienste Grundlage für weitere Finanzierungen und somit essentiell für ein langfristiges Bestehen. Zudem gilt wie überall [Feedback ist gut für den Service](#) – Anbieter wissenschaftlicher Dienste sind für konstruktive Kritik, Bug-Reports und Vorschläge offen. Als Nutzer kann man durch solche Hinweise einen positiven Beitrag leisten, die Dienste zu verbessern.

Wenn Dienste intensiv genutzt werden, ist eine Vernetzung mit den Dienst-anbietern sinnvoll, um ggf. mit diesen für eine längerfristige Finanzierung zu sorgen, aber auch um klare Anforderungsprofile vermitteln. Auch Institutionen treten als Nutzer auf, in dem sie ihre Forscher gegenüber den Anbietern vertreten, Kooperationen eingehen oder externe Anbieter beauftragen. Damit haben Institutionen einen erheblichen Einfluss auf die Auswahl und Entwicklung von wissenschaftlichen Informationsdiensten und müssen diesen nutzen, um ihre wissenschaftlichen Communities mit zeitgemäßen, nachhaltigen und breit vernetzten Diensten zu versorgen.

Handlungsempfehlungen für Anbieter

Im Lebenszyklus der wissenschaftlichen Arbeit werden häufig Forscher/Forschergruppen oder Infrastrukturbetreiber zu Anbietern von Diensten. Dafür braucht es eine intrinsische Motivation, aber auch externe Anreize, um eigene wissenschaftliche Dienste schrittweise für ein breites Publikum zu öffnen.

Solche Anreize können durch z.B. positives Feedback einer Community entstehen (erfolgreiche open source Projekte, Referenzen in Publikationen), durch Karrierepfade in der Organisation (tenure track, agile Teams) oder durch gewährte Gestaltungsspielräume und Ressourcen. Auch eine mutige Projektförderung kann der Startpunkt für einen erfolgreichen Dienst sein, auch wenn anfangs ein **Geschäftsmodell** erst noch gefunden werden muss.

Um solche Anreizsysteme gezielt zu entwickeln, müssen die genannten Arbeiten (Software- und Dienste-Entwicklung, open source Projekte, Services für externe Nutzer) sichtbar gemacht werden. In den Berichten sollten neben den Publikationszahlen und Drittmittelprojekten auch Zahlen und Fakten zu solchen Aktivitäten regelmäßig erscheinen – allein dadurch schon können diese Angebote aufgewertet werden!

Wenn die Absicht und Möglichkeit besteht, einen Dienst anzubieten, sind folgende Fragen möglichst frühzeitig zu klären:

- ◆ Was ist die Zielgruppe und welche Anforderungen hat diese?
 - ◆ Gibt es internationale Initiativen, mit denen sich die Entwicklung vernetzen lässt?
 - ◆ Welche Voraussetzungen (Ressourcen) für eine Servicebereitstellung gibt es?
 - ◆ Wie kann ein nachhaltiger Betrieb finanziert und gewährleistet werden?
 - ◆ Welche Lizenzen, Standards und Schnittstellen sind zu beachten?
 - ◆ Welche Service- und Supportprozesse werden benötigt?
 - ◆ Ist ein Monitoring bzw. Accounting erforderlich?
- 

Für bereits sehr stark nachgefragte zentrale Dienste ist es ratsam, auf eine gesicherte Redundanz zu achten. So ergänzen sich NCBI und EBI in ihren Services. Einen ganz konkreten Datenaustausch und -abgleich gibt es etwa bei Genbank und EMBL. Weitere gute Beispiele sind Pubmed (zentraler Dienst) und, später erschienen, Europe PMC¹⁴. Nutzer, Anbieter und Förderer sollten darauf achten, wo solche Redundanzen noch fehlen und für die Nachhaltigkeit sichere oder komplementäre Dienste wichtig wären. Ein aktuelles Beispiel ist etwa die OMIM Datenbank (genetische Krankheiten), die deshalb eine private Spendeninitiative gestartet hat. Fachgesellschaften können dazu beitragen, rechtzeitig solche Lücken zu erkennen und zu schließen. Besonders Augenmerk haben Dienste verdient, die aus Forschungskonsortien hervorgegangen sind, für die keine nachhaltige Förderung gegeben ist, die aber dennoch einen wichtigen Service für die Community bereitstellen, der dann plötzlich wegfallen könnte.

Das NCBI, EMBL EBI und DDBJ haben gemeinsam die International Nucleotide Sequence Database Collaboration¹⁵ ins Leben gerufen, um die Dienste durch Standards abzustimmen und durch redundante Speicherung von Sequence-Daten ein dauerhafte Verfügbarkeit zu garantieren.

Um die Dienste zu konsolidieren, hat es sich als vorteilhaft erwiesen, diese als Föderation anzubieten. Dabei übernehmen kooperierende Organisationen bestimmte Angebote eigenverantwortlich und können sich mit den verfügbaren Ressourcen auf ihre jeweiligen Kompetenzen konzentrieren. Dieser Ansatz wird in vielen Plattformen erfolgreich realisiert, u.a. EOSC, Sciebo, de.NBI und HIFIS und ist speziell in Forschung und Lehre eine hervorragende Möglichkeit, Ressourcen effizient einzusetzen.

¹⁴ <https://europepmc.org/>

¹⁵ <http://www.insdc.org/>

Ein zunehmend wichtiger Aspekt, den Anbieter digitaler Dienste für die Wissenschaft berücksichtigen sollten, ist die Integration von Funktionalitäten zur strukturierten Wissensrepräsentation in Form von Normdaten und anderen Metadaten sowie zur semantischen Vernetzung von Daten und Entitäten. Annotationen verbessern zum einen die Recherchierbarkeit von Ergebnissen, fördern insbesondere aber auch domänenübergreifende und metadatengetriebene Forschungsansätze.

Zwar existieren bereits Datenbanken (u.a. in Form der Linked Data Cloud) und -Services (Knowledge Graph, Wikidata, GND, ORCID, Annotation Services), die eine Annotation und (domänenübergreifende) Vernetzung erlauben, aber es fehlt noch an einer breiten Integration in wissenschaftliche Abläufe und Dienste. Ziel muss es sein, das strukturierte und an Standards orientierte Annotieren von Forschungsdaten und Publikationen mit international verknüpftem Vokabular als Teil des Forschungsprozesses zu etablieren und dafür Anreize zu schaffen. Eine wichtige Grundlage hierfür ist auch die Entwicklung und Etablierung von Standards und Best-Practices (Ontologien, Abstimmungsprozesse, etc.) in den Fachcommunities sowie fächerübergreifend, um eine Vereinheitlichung und breite Akzeptanz bezüglich des Annotationsvokabulars zu erreichen. Hierfür sind kollaborative Plattformen für die Wissensmodellierung, wie z.B. die Software Wikibase geeignet. Bestandteil eines solchen Services ist dabei auch die Organisation von Abstimmungsprozessen (Arbeitsgruppen, Gremien, etc.) und der Aufbau von Mechanismen zur Qualitätssicherung.

Auch im Bereich der strukturierten Wissensrepräsentation sind Angebote zu bevorzugen, die einen dauerhaft freien Zugriff auf Daten, eine flexible Einbindung in Services und Partizipationsmöglichkeiten erlauben. Dies kann z.B. durch die Nutzung bereits bestehender offener Plattformen (u.a. wikidata, GND) und durch das Angebot eigener Services basierend auf offener Software erfolgen. Langfristig könnte auf diese Weise ein maschinenlesbares, semantisches Netz der Kultur und Wissenschaft entstehen.



Handlungsempfehlungen für Förderer

Förderorganisationen können auf vielfältige Weise auf die Gestaltung der Landschaft wissenschaftlicher IT-Dienste einwirken. Wie aus Kapitel 2 hervorgeht, sind IT-Dienste mittlerweile essentielle Bausteine des wissenschaftlichen Arbeitens (z.B. PubMed, GitHub). Vor diesem Hintergrund sollten von Seiten der Förderorganisationen klare Bekenntnisse vorgelegt werden, **IT-Dienste im Allgemeinen und einzelne, erfolgreiche Dienste im Speziellen als essentiellen Teil der wissenschaftlichen Infrastruktur anzuerkennen**. Es sollten konkrete Verpflichtungen genannt werden, die zum Ziel haben, die dauerhafte finanzielle und organisatorische Absicherung möglichst vieler erfolgreicher wissenschaftlicher Services zu ermöglichen. Diese Verpflichtungen sollen schließlich dazu beitragen, dass digitale Dienste für die Wissenschaft nicht gänzlich kommerziellen Anbietern überlassen werden.

Für die ständige Fortentwicklung der Wissenschaft sind sowohl bestehende IT-Dienste, als auch die kreative Neuschöpfung innovativer Dienste unverzichtbar. Anbieter und Förderorganisationen haben bereits viele Grundsteine und Anreize gesetzt, damit Dienste möglichst stark genutzt und beständig gewartet werden (s. Kapitel 2). Entscheidend für die weitere Entwicklung ist jedoch, dass ein qualitativ hochwertiger Dienst bereitgestellt wird. Die **Qualität eines Dienstes** sollte als zentrales Bewertungsmerkmal für Förderentscheidungen herangezogen werden. Außerdem sollten neben anderen Qualitätsmaßstäben unbedingt **anerkannte Maßstäbe an Datenschutz und Sicherheit der Dienste** gesetzt, eingefordert und als zweites zentrales Bewertungskriterium anerkannt werden.

Viel genutzte IT-Dienste kommerzieller Anbieter erreichen teils eine sehr hohe Marktdurchdringung, auch im wissenschaftlichen Betrieb (z.B. Google). Die Etablierung neuer oder wenig genutzter wissenschaftlicher IT-Dienste ist dadurch deutlich erschwert. Eine **erfolgreiche Strategie zur Konsolidierung** sollte deshalb neben der häufig bestehenden Möglichkeit zur innovativen (Neu-/Weiter-)Entwicklung von Services durch Einzelpersonen oder kleinen Teams auch immer Gelegenheiten bieten, vielversprechende und erfolgreiche Dienste durch geeignete Förderformate **wachsen** zu lassen. Diese Förderformate sollten bereits zu Beginn einer Förderperiode auf folgende Eigenschaften der Dienste Wert legen: Nutzung von offenen Schnittstellen und Standards, Nutzung von Open-Source-Lizenzen und eine möglichst große Offenheit und **FAIRness** in Bezug auf Nutzergewinnung/-bindung.

Neben regelnden Stellschrauben sollten Förderer auch Möglichkeiten positiver Rückkopplung durch die **Stärkung von Anreizsystemen** schaffen. Insbesondere die Einbindung der Nutzerinnen und Nutzer sollte durch Förderorganisationen eingefordert werden, um z. B. regelmäßiges Feedback und/oder **Community Support** zu erhalten. Best-Practice-Beispiele für wünschenswerte Projektentwicklungen sollten ausgewählt und als Vorbild angemessen gewürdigt werden (z. B. durch Preise oder durch gezielte Nennungen der Projekte an prominenten Stellen). Außerdem sollten Anreizsysteme, die sich an Einzelpersonen richten, verbessert werden. Für diese Verbesserung sind denkbar: die Einbeziehung von Daten- und Softwarepublikationen bei der Bewertung der wissenschaftlichen Arbeit einer Person oder die Etablierung von Karrierepfaden für wissenschaftliche Softwareentwickler.

Der Aufbau und die Erhaltung einer abgestimmten, überregionalen und nachhaltigen Informationsinfrastruktur muss zu einem großen Teil durch kooperative Anstrengungen erreicht werden. Förderer sollten deshalb geeignete Rahmen schaffen, um im Bereich digitale Dienste über Projekte und Einrichtungen hinaus **Austausch und Zusammenarbeit auf nationaler sowie internationaler Ebene** zu verstärken. Als Ziele der kooperativen Prozesse sollten Leistungsfähigkeit, Zugänglichkeit und Nutzbarkeit der Services im Vordergrund stehen. Wichtig ist deshalb, dass Maßnahmen vor allem aus den Communities formuliert, koordiniert und durchgeführt werden können. Denkbar ist z. B., dass Förderorganisationen den finanziellen Rahmen für solche Aushandlungsprozesse zur Verfügung stellen und somit bereits der organisatorische Rahmen durch die Communities selbstverwaltet erzeugt wird. Je nach Intensität und Umfang der Zusammenarbeit sollte jedoch auch weitreichende Unterstützung möglich sein.



Zusammenfassung

Im Zuge der fortschreitenden Digitalisierung befindet sich die Welt der Wissenschaft in einem tiefgreifenden Wandel. Das Ziel dieses Positionspapiers ist, das Bewusstsein für die aktuellen Anforderungen und Notwendigkeiten der Entwicklung, Bereitstellung und des Betriebs von digitalen Diensten für die Wissenschaft zu schärfen, die nationale und internationale Einbettung dieser Dienste zu betrachten und potentielle Lösungsansätze und Handlungsempfehlungen zu diskutieren. Das soll der Ausgangspunkt für eine jetzt notwendige, breite Diskussion in der Forschungslandschaft sein.

Für die in der Einleitung genannten Schlüsselfragen ergibt sich im Hinblick auf die Landschaft der genutzten digitalen Dienste für die Wissenschaft ein differenzierter Befund. Es sind aber viele gute Praxisbeispiele und Tendenzen erkennbar, die erste Rückschlüsse und Handlungsempfehlungen erlauben. Diese zeigen, dass es in der europäischen Forschung möglich und vorteilhaft ist, digitale Dienste durch Bündelung von Ressourcen und Fokussierung auf die spezifischen Anforderungen und Workflows der wissenschaftlichen Arbeit selbst zu gestalten und nachhaltig zu betreiben.

Welche Arten von Diensten werden von den Akteuren in der Forschung gefordert und welche werden tatsächlich weitläufig genutzt?

Aus vielfältigen Gründen besteht eine erhebliche Diskrepanz zwischen Diensten, die von den Akteuren in der Forschung gefordert werden und den tatsächlich weitläufig genutzten. Die leichte Zugänglichkeit, Performanz, zunächst geringen Kosten und z. T. auch ein auf wissenschaftliche Bedarfe zugeschnittener Leistungsumfang tragen zu einer pragmatischen Nutzung globaler kommerzieller Lösungen durch wissenschaftliche Communities bei. In den letzten Jahren entstanden jedoch auch viele Dienste der wissenschaftlichen Infrastrukturanbieter sowie der Communities selbst, von denen sich einige bereits hervorragend etabliert haben und als kollaborative Plattformen nicht nur Daten, sondern auch Softwarewerkzeuge, Publikationen und Expertennetzwerke zusammenführen.

Welche Vor- und Nachteile ergeben sich aus der gegenwärtigen Nutzung einerseits und welche Möglichkeiten bieten sich andererseits, notwendige Dienste zusätzlich oder als Alternative zu kommerziellen Angeboten wie z.B. von Google, Amazon, Microsoft, Elsevier oder Digital Science, im nationalen und europäischen Rahmen zu etablieren?

Während die Vorteile der Nutzung der gegenwärtig verbreiteten digitalen Dienste für die Wissenschaft offenkundig sind, ist insbesondere die Nutzung kommerzieller Angebote zugleich mit vielfältigen Nachteilen verbunden, darunter das mangelnde Interesse kommerzieller Anbieter an Offenheit und gleichberechtigter Partnerschaft, die häufige Weiterverwertung von Daten in kommerziellem Interesse und das Risiko eines Vendor-Lock-in. Die Möglichkeiten, notwendige Dienste zusätzlich oder als Alternative zu kommerziellen Angeboten im nationalen und europäischen Rahmen zu etablieren, werden bislang noch keinesfalls umfassend ausgeschöpft.

Wissenschaftliche Communities sollten auf diese Nachteile und Herausforderungen reagieren. Eine Lösung könnte im Aufbau und der konsequenten Stärkung nachhaltiger, nicht-kommerzieller, kooperativer Strukturen bestehen, bevor im Bereich wissenschaftlicher Informationsdienste eine ähnliche Abhängigkeit entsteht wie im Fall der wissenschaftlichen Informationsversorgung mit den großen Wissenschaftsverlagen.

Wie können insbesondere unter dem Aspekt der (finanziellen) Ressourcen-Effizienz wissenschaftliche Einrichtungen und Fachcommunities in die Lage versetzt werden, thematische Dienste in eigener Verantwortung bereitzustellen?

IT-Dienste im Allgemeinen und einzelne digitale wissenschaftliche Dienste im Speziellen sollten von Förderern als essentieller Teil der wissenschaftlichen Informationsinfrastruktur anerkannt werden. Die Qualität eines Dienstes sollte als zentrales Bewertungsmerkmal für Förderentscheidungen herangezogen werden. Es sollten Maßstäbe wie FAIR-Prinzipien, Nachhaltigkeit, Datenschutz und Sicherheit für Dienste gesetzt, eingefordert und als weitere Bewertungskriterien anerkannt werden. Eine erfolgreiche Strategie zur Konsolidierung sollte neben der Möglichkeit zur innovativen (Neu-/Weiter-)Entwicklung von Services durch Einzelpersonen oder kleine Teams auch Gelegenheiten bieten, vielversprechende Dienste durch geeignete Förderformate nachhaltig voranzubringen.

Neben regelnden Stellschrauben sollten wissenschaftliche Einrichtungen und Förderer auch Möglichkeiten einer positiven Rückkopplung durch die Stärkung von Anreizsystemen für die Entwicklung von nachhaltigen Diensten und Software schaffen. Insbesondere durch Zitation und Einbeziehung in das wissenschaftliche Reporting wie auch als Kriterium bei Stellenvergaben und Berufungen müssen Anreize gesetzt werden. Nicht zuletzt sollten Förderer gezielt Rahmenbedingungen

schaffen, um Austausch und Zusammenarbeit auf nationaler und internationaler Ebene zu verstärken.

Die aktuellen tiefgreifenden Veränderungen im Zuge der Digitalisierung können zu einem freien und demokratischen Zugang zu nahezu unbegrenzten Ressourcen für die Wissenschaft führen, sie können aber auch große Abhängigkeiten und Verwerfungen zur Folge haben. Wir können diesen Prozess beeinflussen! In einem virtuellen Rückblick aus dem Jahr 2050 auf das Jahr 2019 beschreibt Dennis Gannon diese Zeit als Wendepunkt zu einer positiven Entwicklung:

e-Science 2050: A Look Back¹⁶

- ◆ Die Wissenschaftler befanden sich 2019 an einem wichtigen Wendepunkt in Bezug auf die Technologie, die sie in der Wissenschaft einsetzen konnten. Die Cloud hat sich zu einem riesigen heterogenen Online-Supercomputer entwickelt, der auf Abruf zur Verfügung steht. Lange Zeit als ein Thema von rein theoretischem Interesse betrachtet, tauchte das Quantencomputing als ein Dienst in der Wolke auf.
- ◆ Eines der bemerkenswertesten Merkmale der Informatik war die Entwicklung von Software. Die Programmierwerkzeuge hatten sich zu sehr tiefen Stacks entwickelt, die KI-Methoden verwendeten, um Wissenschaftler in die Lage zu versetzen, z.B. mit ein paar Zeilen Julia in einem Jupyter-Notebook mehr zu erreichen, als es 1980 mit der aufwendigen Programmierung von Großrechnern möglich war.
- ◆ Die Rolle der KI beschränkte sich nicht mehr auf die Programmierung und Ausführung von e-Science-Experimenten. Es entstanden erste einfache KI-Forschungsassistenten, die als intelligentes System Informationen der Wissenschaft lesen, lernen und einfache (wissenschaftliche) Fragen beantworten konnten.

16 Basiert auf Gannon, Dennis: eScience 2050: A Look Back, 2019/08/08, DOI 10.13140/RG.2.2.14835.07206

Impressum

Herausgeber

Arbeitsgruppe Digitale Werkzeuge und Dienste der Schwerpunkttinitiative Digitale Information (<https://www.allianzinitiative.de/>) der Allianz der deutschen Wissenschaftsorganisationen.

Zu den Mitgliedern der Arbeitsgruppe gehören (in alphabetischer Reihenfolge):

Mathias Bornschein, Björn Brembs, Thomas Dandekar, Zeki Mustafa Dogan, Dirk Eisengraber-Pabst, Georg Feulner, Konrad Förstner, Michael Franke, Bernadette Fritzsich, Jürgen Fuhrmann, Michael Goedicke, Stefan Janosch, Jürgen Kett, Uwe Konrad, Florian Mannseicher, Johannes Reetz, Klaus Wannemacher

Federführende Autoren

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 Lebens-
wissenschaften
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 13a
30159 Hannover
E-Mail: wannemacher@his-he.de
ORCID: 0000-0003-4810-3693

Digital Object Identifier

Die Onlineversion dieser Publikation finden Sie unter

<http://doi.org/10.5281/zenodo.4301924>

Engelsprachige Übersetzung

Eine englischsprachige Übersetzung finden Sie unter

<http://doi.org/10.5281/zenodo.4301947>

Herstellung

Gabriele Wicker, Fraunhofer IRB, Stuttgart

Stand

Dezember 2020

Lizenz

Alle Texte dieser Veröffentlichung, ausgenommen Zitate, sind unter einem Creative Commons Attribution 4.0 International (CC BY 4.0) Lizenzvertrag lizenziert.

<http://creativecommons.org/licenses/by/4.0/>



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