

REVIEW

14 Years of PID Services at the German National Library of Science and Technology (TIB): Connected Frameworks, Research Data and Lessons Learned from a National Research Library Perspective

Angelina Kraft¹, Britta Dreyer¹, Peter Löwe² and Frauke Ziedorn¹

¹ Technische Informationsbibliothek (TIB), German National Library of Science and Technology, Welfengarten 1B, Hanover, DE

² Deutsches Institut für Wirtschaftswissenschaften (DIW), German Institute for Economic Research, Mohrenstrasse 58, Berlin, DE

Corresponding author: Angelina Kraft (angelina.kraft@tib.eu)

In an ideal research world, any scientific content should be citable and the coherent content, as well as the citation itself, should be persistent. However, today's scientists do not only produce traditional research papers – they produce comprehensive digital resources and collections. TIB's mission is to develop a supportive framework for a sustainable access to such digital content – focusing on areas of engineering as well as architecture, chemistry, information technology, mathematics and physics. The term digital content comprises all digitally available resources such as audiovisual media, databases, texts, images, spreadsheets, digital lab journals, multimedia, 3D objects, statistics and software code.

In executing this mission, TIB provides services for the management of digital content during ongoing and for finished research. This includes:

- a technical and administrative infrastructure for indexing, cataloguing, DOI registration and licensing for text and digital objects, namely the TIB DOI registration which is active since 2005,
 - the administration of the ORCID DE consortium, an institutional network fostering the adoption of ORCID across academic institutions in Germany,
 - training and consultancy for data management, complemented with a digital repository for the deposition and provision of accessible, traceable and citable research data (RADAR),
 - a Research and Development Department where innovative projects focus on the visualization and the sustainable access to digital information, and
 - the development of a supportive framework within the German research data community which accompanies the life cycle of scientific knowledge generation and transfer. Its goal is to harmonize (meta)data display and exchange primarily on a national level (LEIBNIZ DATA project).
-

Keywords: Information technology; library; database; discoverability; digital resources; research data

Introduction

Academic libraries are experts at identifying, selecting, organizing, describing, preserving, and providing access to information materials, print and digital resources. They have been a safe harbour for research publications of all disciplines since centuries. As a cornerstone of the academic institution, libraries have a demonstrated and sustainable approach for providing services such as collection management, preservation,

and access to a broad variety of information. However, we are experiencing a paradigm shift in the way knowledge is shared. Declining usage of traditional library services, the increased availability of eBooks and e-articles as well as the rise of open access to academic resources are just some examples. Moreover, the present age of digitalization is enabling fundamental changes in how we experience and use the products of research, innovation and personal ideas.

Therefore, it becomes evidently that the academic library's role as a gateway to high quality research information is changing. The need for physical collections and traditional systems for the organization of information must be reconsidered. The challenge lies in how to invest and develop innovation while at the same time continue with selected traditional services. It also includes evaluating services, sharing the results and consequently transforming the workflows toward more digital stewardship. Digital curation services for research data (Keil 2014; Tenpoir et al. 2014; Swanson & Rinehart 2016) and other content such as audio-visual media, images, digital lab journals, multimedia, 3D objects, statistics and software code (Koltay 2016) are new key activities for academic libraries.

Institutions and researchers recognize that the longevity of digital resources may vary depending on the type of data or research. For example, DNA sequences may be outdated after a decade, whilst taxonomy data may be relevant forever. The possibly most important challenge, however, was stated in a white paper by the internationally operating foundation on research data issues, the Research Data Alliance (RDA) (RDA Europe 2014): "*The right minds get the right data at the right time.*" If you slightly change this phrase to '*make sure that the right minds get the right (digital) resources at the right (any?) time*' the future role of research libraries becomes apparent. Data libraries which record the old and new findings of research which we are experiencing in these exciting age of digitization, particularly for research data and other digital research outputs which are not served by big data centres need to be established. But how to achieve this goal? How can we organize it that a digital network, a virtual library can assist to preserve research outputs such as research data, audiovisual media, databases, texts, images, spreadsheets, digital lab journals, images, multimedia, 3D objects, statistics and software code can be found and cited? And how can these resources be re-used to test new hypotheses, combine data and reproduce scientific findings?

When it comes to digital academic resources such as journal articles and the emerging research data, the use of unique and persistent identifiers (PIDs) has become a central aspect of proper data management and access. DOIs (Digital Object Identifiers = DOI® names) are originally contrived for publications of scientific findings as the core technology to refer to the electronic version of an article in a journal. A DOI consists of a unique character string that identifies an entity in a digital environment. Therefore, it identifies the object itself and not the place where it is located. If the object is moved and the location (URL) has changed, the only requirement is to update the URL in the underlying central database. This system ensures that the DOI persistently resolves to the location of the object (Paskin 2006). The Digital Object Identifier System is managed and administrated by the International DOI Foundation (International DOI Foundation 2012).

Several registration agencies provide DOI services and registration worldwide. One of them is DataCite, a registration agency particularly dedicated to services that support the enhanced search and discovery of research content, especially on research data and grey literature. The DOI community and the DataCite consortium are maintaining a network of services within the DOI system. The relationship of a digital resource to other attributes or resources can be captured in and exposed directly through the service itself as identifier metadata. Identifier metadata may be accessed without visiting the resource itself, thus reducing the load on repositories and catalogues. The usage of persistent identifiers like DOI also enable services such as event tracking of citations that otherwise would not have been realizable for dark archives.¹ To achieve this, though, a PID service must be able to ensure that the types of information needed are available in the metadata. With PID services at the base, libraries are supporting researchers to deposit their research output into disciplinary research data repositories (e.g. ICPSR, GenBank and Pangaea). Additionally, they are establishing and managing institutional data repositories to support the long-tail of data outside of big data disciplines. Furthermore, libraries are helping to prepare data for sharing and re-use much earlier in the research life-cycle (e.g. in the development of a Data Management Plan, a data collection, or a long-time data storage).

While developing these services, research libraries need to be an active player in national, European or international initiatives (Pinfield et al. 2014). As such, the TIB e.g. is active in the Research Data Alliance (RDA), the Force 11 group and ICSTI. Within RDA, the focus is on the topics libraries for research data, long tail of research data, publishing, cost recovery for data centres, legal interoperability, metadata as well as PID services.

¹ Dark archive: Archives storing a data copy to serve local users (non-public) or to function as a failsafe during disaster recovery.

Services Established and Lessons Learned

For more than 50 years, the German National Library of Science and Technology (TIB) in Hannover has been providing scientific information from the disciplines of engineering as well as architecture, chemistry, information technology, mathematics and physics. The TIB has an outstanding stock of core and highly specialized technical and scientific literature and represents the largest specialist library in the world for its subject areas. In the course of transitioning from a traditional library to a modern technical information centre new services, i.e. the DOI service, are implemented and integrated in existing services. Meanwhile, a wide range of new digital contents emerge, using grey literature (i.e. reports (annual, research, technical, project), white papers, government documents, etc.) and as focal point and bridge to comprehensive technical information. Every year, TIB serves customers from around 65 countries. About 55% of these customers belong to the academic research community whereas 45% are commercial clients.

DOI Service: The facts

Research data management offers solutions for the proper storage and curation of datasets and other digital objects and their linking with publications throughout the scholarly research cycle. If authors submit the digital objects that support research papers to (certified) research infrastructures such as repositories, it will make future research studies and information retrieval much easier. In the light of this, TIB took a fundamental step with the innovative STD-DOI project in 2003 (**Figure 1**): In collaboration with scientific institutes, TIB developed an infrastructure model for the DOI registration, establishing a complete workflow for the referencing of research data. Five years later this work led to the foundation of DataCite in 2009. Nowadays, DataCite is a globally oriented non-profit organization operating from local institutions, with 47 members from more than 20 countries. Members include the British Library, the California Digital Library, the Library of ETH Zurich and the Australian National Data Service. DataCite offers an infrastructure that supports simple and effective methods of data citation, discovery and access. Trusted and certified data centres collaborating with the DataCite network and DOI agencies register and supply DOIs for deposited digital resources. Therefore, these objects can be linked to the corresponding publications in a persistent way. Depending on which properties are provided alongside a digital resource, the identifier metadata enables services that support discovery, access, verification of integrity and authenticity and a variety of other use cases. Furthermore, the respective data centres often also provide long-term preservation services for the digital objects. In summary, the use of DOIs enables the scientific community to move beyond journals and make more digital scientific content visible, available and searchable in a citable way. TIB in its role as the German National Library of Science and Technology developed dedicated services around the research data life cycle

- Over 120 academic institutions are provided with administrative, scientific and technical support from the TIB DOI service. Over 1.5 million DOIs for academic content have been registered since 2005.
- Out of the 1.5 million DOIs registered via TIB, 62% have been assigned to research data, 37% to grey literature and 1% to audiovisual media.

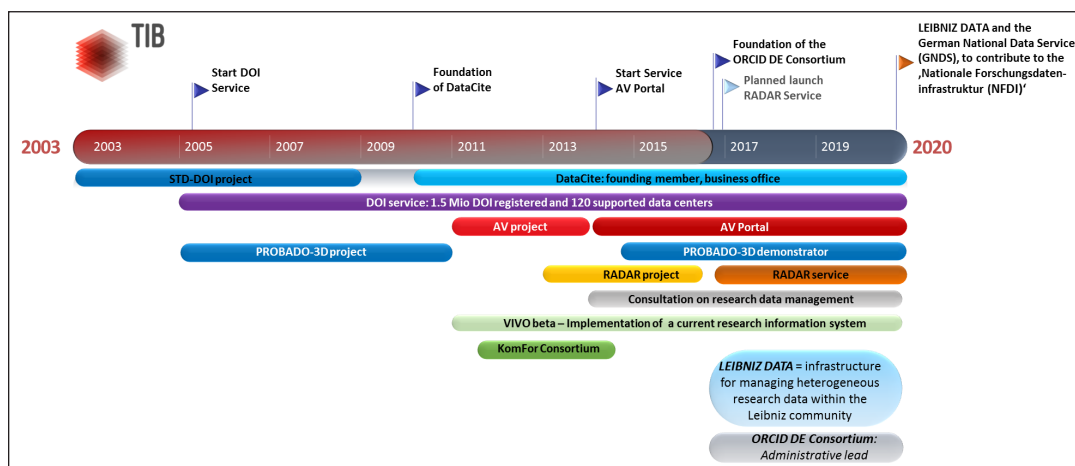


Figure 1: Past, present and planned future management of digital resources at TIB, with a focus on research data.

- As part of our national mandate, the TIB DOI service is free of charge in the EU for publicly funded institutions. Clients include major research centres such as PANGAEA, the World Data Center for Climate (WDCC) and the European Southern Observatory (ESO) as well as 51 universities and university libraries.

Metadata quality of digital resources is of utmost importance for their searchability and citeability, especially in the academic context. The research community assumes that the objects identified by a PID are part of the academic content that might typically be referenced in a journal article. Therefore, PID services such as the ones offered by DataCite have a metadata standard based around the typical attributes of academic records (including parameters such as title, creator, publisher and publication year). However, to ensure that the objects identified via PID service providers become a related part of academic content, metadata standards have to be constantly checked against current and new academic standards and evolve forward. One example is the new DataCite *Metadata Schema Version 4.0.*, published by the DataCite Metadata Working Group (2016). The new schema includes more mandatory fields, including the description of the resource type and the possibility to add a funding reference as well as a funder identifier. The changes increase interoperability with other PID types such as the ORCID for the persistent identification of a researcher and enhance the discoverability of research objects registered via DOI. It also allows PID networks and service providers the implementation of new services such as *DataCite Event Data* (<https://eventdata.datacite.org>), a service which collects events around DataCite DOIs including references to related data, data citations in journal articles, and new versions of a work. The central goal of the TIB research data management service is to pass on such information in a clear and concise manner to our DOI service clients in Germany and Europe. As a research library, we also provide advice, support and in some cases technical services as part of our responsibility to accelerate the current transition to a digitized, data-saturated research system. We prepare for new innovations, such as the recently announced European Open Science Cloud as part of the Digital Single Market (Commission High Level Expert Group on the European Open Science Cloud 2016).

Example from the daily DOI business

In order to reach a functional and (in an ideal world) open network of research data and other digital resources, there are two basic types of challenges to overcome; one being technical (can we agree upon common, interoperable standards for data and the associated metadata?) and the other being social (can we agree upon a similar strategy for data management, e.g. like the Joint Declaration of Data Citation Principles by the Data Citation Synthesis Group (2014)?).

While the technical challenge facing the complexity of data is addressed by multiple working groups across the world (e.g. European Union 2010; Crosas 2011; Lecarpentier et al. 2013; Starr et al. 2015) and seems to make good progress, the social (human) side is still staggering and changing slowly (Data Citation Synthesis Group 2014; Roche et al. 2015). A successful example is the the COPDESS Statement of Commitment (<http://www.copdess.org/statement-of-commitment/>) which includes a recommendation to archive data in public data repositories, and the acceptance of dataset DOIs in reference lists of Journal articles.

Beside the shortage of experts on the handling of digital resources, there are still major questions concerning the 'art of resource management' within and between the scientific communities: One of the historic but still omnipresent questions is, whether the scientific community expects the object being identified via a PID to contain exclusively academic content, and by which standards the content is evaluated as such. The following describes one example from the practice of TIB, being a DataCite member for seven years: In early 2014 we received an inquiry of DOIs for field stations, a topic which has been discussed for some time. The foundation of DataCite was a specific initiative to broaden the definition of academic output, focusing on the persistent identification of research data. Therefore, many questions and discussions have taken place regarding the scope of 'academic content'. Questions asked were '*Should field laboratories be identified persistently?*' and '*Which identifier should be used?*' Sometimes it is hard to tell where the line of academic content should be drawn, and when to cross it. From a research library point of view, we strongly recommended to select an identifier based on a globally unique scheme for a field station, while considering that not every PID may fit every purpose. In order to identify a field station, a whole new metadata approach and much more insight in geographic location description and temporal resolution are needed. If field stations are located close to each other, detailed spatial and temporal resolution as well as information about the hard- and software of each field station may be of great interest to distinguish between them. The usage of varying spatial reference systems in the different disciplines is another factor that needs to be considered. All these aspects cannot be sufficiently represented by the DataCite metadata schema in its present form. In

conclusion, it was recommended to the inquirer that other identifier initiatives such as the International Geo Sample Number (IGSN) or geocoding might be more suitable here, because it includes metadata designed to describe sampling sites (<http://www.geosamples.org/help/vocabularies/#object>) and can be used for the description of field stations. An example of a field station can be found here: <https://app.geosamples.org/sample/igsn/JPLOOZM00>.

Another example is the adopted use of DOIs for scientific audiovisual media such as recordings of conferences, lectures and experiments, reports and presentation of research work. These media are made available via the TIB AV-Portal, a cooperative project of both TIB and the Hasso-Plattner Institute. The portal was launched in April 2014 (av.tib.eu) and focuses on the automatic generation of metadata, a semantic search and cross-lingual retrieval (German and English). Content-based filter facets for search results enable the exploration of the increasing number of video assets. Search terms are not only searched for in the films metadata, such as author, title or abstract but also in the spoken texts, text overlays and image information. These technologies allow the users to search more efficiently by locating the relevant video segment even if the search term could not be found in the film metadata such as title or abstract. By using the open standard Media Fragment Identifier (MFID) in addition to the DOI, individual segments of a video can also be cited as easily as a chapter or a page in a book. In order to cite a video or a video segment the provided DOI link – which can be enhanced by the MFID – is simply copied and pasted into a document. By connecting video assets to dynamic digital content such as datasets, the scientific value can be increased and contribute to a better understanding of the content life-cycle, from acquisition and preservation to access.

Research and Development Department

The German National Library of Science and Technology underpins its PID and framework services with research on sustainable access to digital information, its visualization and by highlighting its connectivity and potential use for other research areas. Thus, this research is a derivate of the main mission and is focused on innovation while continuously improving the existing services. A major goal of TIB is to establish its own research capacity and support the trust in and (future) role of libraries as preservers of (digital) information and knowledge.

Present research and development projects at TIB focus on

- innovative, media-specific portals enabled by e.g. an automated video analysis with scene, speech, text and image recognition (an additional service upgrading the TIB AV-Portal).
- bibliometric and linked open data (LOD) services for research products such as author information, full texts and research data as part of the KomFor Consortium² and the VIVO beta project (<https://vivo.tib.eu/vivo/>).
- the development of tools and the prototypical operation of a service for automated indexing, storage and presentation of non-textual types of documents, e.g. 3D models of CAD applications (PROBADO-3D project, http://www.probado.de/en_3d.html).
- the area of visual analytics by developing automatic processes that recognize patterns in the data and generate condensed graphic representations of high-dimensional data.
- the development of new ways of software citation and software management tools (e.g. using the Jupyter Notebook, a web application).
- establishing a generic Research Data Repository (RADAR), a collaborative project to preserve research results up to 15 years and assign well-graded access rights, or to publish data with a DOI assignment for an unlimited period of time (www.radar-service.eu/en).

Potential clients for the project results include libraries, research institutions, publishers and open platforms which require an adaptable digital infrastructure to archive, publish, analyze and explore digital resources while considering their institutional requirements and workflows. In the following, three exemplary outputs of TIB's participation in PID services and supporting frameworks are described: the collaborative project for

² KomFor Consortium: The Competence Center was established in a project from 2011 to 2015. It serves as link between Geoscientific facilities and an existing archive network for earth and environmental data in Germany. KomFor generally aims at improving the overall availability and quality of data in a sustainable way. KomFor is based on ICSU World Data Centers and Services located in Germany, which are collaborating in a national cluster since 2003 – the WDC Climate (WDC-C, DKRZ), WDC for Remote Sensing (WDC-RSAT, DLR), the German Research Centre for Geosciences (GfZ), and the Data Publisher for Earth & Environmental Science (PANGAEA, AWI/Marum), Springer, AGU etc.)

research data preservation and archiving, RADAR, the planned framework LEIBNIZ DATA and the recently established ORCID DE Consortium.

RADAR

Globally resolvable, persistent digital identifiers have become an essential tool to enable unambiguous links between published research results and their underlying digital resources. One particular problem for the management of data originating from (collaborating) research infrastructures is their dynamic nature in terms of growth, access rights and quality. On a global scale, systems for access and preservation are in place for the big data domains (e.g. environmental sciences, space, and climate). However, the stewardship for disciplines without a tradition of data sharing, including the fields of the so-called long tail, remains uncertain.

RADAR – Research Data Repository – is an interdisciplinary end-point research data repository which provides both preservation and publication services. The project focuses on the so-called 'long tail' of research disciplines and will serve as an addition to established 'big data' and/or domain specific repositories, with a complementary function rather than being competitive to them. RADAR started in March 2017 and provides data services for customers without own data repository infrastructures or storage capacities. The repository was developed in the course of a project funded by the German Research Foundation from 2013 to 2016: <http://www.radar-projekt.org>. The project is placed within the program 'Scientific Library Services and Information Systems (LIS)' on restructuring the national information services in Germany. RADAR welcomes data from specialized research disciplines of all areas, i.e. natural, economic, social and cultural sciences. The heterogeneity of research data is a serious issue for many research data repositories, especially when they provide storage and publication services for a wide range of scientific disciplines. RADAR is facing this problem by focusing on real scientific workflows and elaborates a generic best practice approach that will be evaluated and tested with data provided by scientific partners from different research areas.

RADAR is developed as a cooperation project of five research institutes from the fields of natural and information sciences (Razum et al. 2014). The technical infrastructure is provided by the FIZ Karlsruhe – Leibniz Institute for Information Infrastructure and the Steinbuch Centre for Computing (SCC) of the Karlsruhe Institute of Technology (KIT). The Ludwig-Maximilians-Universität Munich (LMU), Faculty for Chemistry and Pharmacy, and the Leibniz Institute of Plant Biochemistry (IPB) provide the scientific knowledge and specifications and ensure that RADAR services can be implemented in the actual scientific workflows of academic institutions and universities. The sustainable management and publication of research data with DOI assignment is provided by the German National Library of Science and Technology. The partners aim to establish an interdisciplinary research data repository based on a stable business model. The data management processes and tools needed to achieve this goal include

- guidelines for researchers to introduce and facilitate research data management in general and to store/publish data in RADAR in particular,
- secure data preservation in compliance with required storage periods (including permanent storage) by the use of distributed data storage mechanisms,
- (optional) data publication with DOI-assignment to secure traceability, access and citeability, and
- technical support for institutions, including an optional provision of a review link that may be sent to reviewers/editors during the peer-review of a corresponding paper and frontend-branding.

Being the proverbial “transmission belt” between data producers and data consumers, RADAR specifically targets researchers, scientific institutions, libraries and publishers. In the data lifecycle, RADAR services are placed in the “Persistent Domain” of the conceptual data management model described in the “domains of responsibility” (Treloar et al. 2008). These domains of responsibility are used to demonstrate duties and responsibilities of the stakeholders involved in research data management. Simultaneously, the domains outline the contexts of shared knowledge about data and metadata information, with the goal of a broad reuse of preserved and published research data. Depositing research data in RADAR ensures that the requirements of funding agencies and of Good Scientific Practice are met. As a generic service RADAR accepts all types of digital data that are collected in the course of scientific research studies. A dataset deposited in RADAR may comprise raw data, primary data (intermediate working data), secondary data and files describing the data and documenting the research process. RADAR accepts both data underlying scientific articles and standalone data publications, e.g. 'negative data'.³ Data may be submitted in any file format, however, format recommendations reflecting

³ Negative data: When scientific experiments do not generate the expected outcome, contradict the initial hypothesis results or do not support the expected model, negative results and negative data are created. The data are indeed present, but different than expected and therefore often not considered 'worthy' of publication.

the requirements for long-term accessibility of digital content will be provided in the author guidelines (e.g. the use of PDF/A or XML based formats for text files). The online service can be used in a collaborative way. RADAR enables clients to upload, edit, structure and describe their (collaborative) data in an organizational workspace. In such a workspace, administrators and curators can manage access and editorial rights before the data enters the preservation and optional publication phase. RADAR applies different PID strategies for closed vs. open data. For closed datasets, RADAR uses handles as identifiers and offers format-independent data preservation between 5 and 15 years, which can also be prolonged. By default, preserved data are only available to the respective data curators, which may selectively grant other researchers access to preserved data. For open datasets, RADAR provides a DOI to enable researchers to clearly reference and reuse data and to guarantee data accessibility. RADAR offers the publication service of research data together with format-independent data preservation for an unlimited time period. Each published dataset can be enriched with discipline-specific metadata and an optional embargo period can be specified. Workflows and detailed services of RADAR, including a pricing model, are available online (www.radar-service.eu/en).

A repository service to publish and preserve research outputs takes up a significant part of an institutional strategy and budget planning process. To aid institutions in such processes, various cost models have become available over the last years. Examples include the European 4C project (C Project Collaboration to Clarify the Costs of Curation 2016) and the APARSEN project, which maps and compares the various models (Kaur et al. 2014). For RADAR, a “cost by service” approach was selected. The calculation sheets are dependent on the three central phases (ingest, curation and access) of the repository service. The RADAR pricing model includes yearly payment plans based on institutional contracts depending on required storage volume and duration. Additionally, the pricing model of RADAR is (indirectly) subsidized by governmental funds (directed to the academic, non-profit institution operating the service, FIZ Karlsruhe), in addition to the service fee which is charged for its use. While this mixed business model of a subsidized service fees is still a novel structure in Germany, such models are well known and used (i.e. in repositories such as Dryad, <http://datadryad.org/>). With RADAR, academic institutions pay the fee for services including long term preservation without any additional cost for the individual researcher being a member of the respective institution.

RADAR aims to meet demands from a broad range of specialized research disciplines: To provide a secure, citable data storage and citeability for researchers which need to retain restricted access to data on one hand, and an e-infrastructure which allows for research data to meet the FAIR principles (Wilkinson 2016), meaning for research data to be findable, accessible, interoperable and re-useable in a digital platform available 24/7, on the other.

Planned frameworks: LEIBNIZ DATA and the ORCID DE Consortium

Libraries, IT services and research offices at an institution are increasingly required to collaborate locally, nationally and globally in order to jointly build data libraries that support diverse data and research in the long-tail. A positive trend is the increased communication and formation of connections between research libraries and their engagement with research councils, funding councils and other key national and international stakeholders such as the Research Data Alliance (e.g. the Libraries for Research Data Interest Group) to gain and pass on best practices and guidelines of digital (data) curation, e.g. the project *23 Things: Libraries for Research Data* (Witt & Libraries for Research Data Interest Group 2016). As such, they may promote the interoperability of disciplinary, national, and international infrastructures, provide a communication platform for the exchange of knowledge and work together to address central challenges. An exemplary project for such a communication and outreach framework concerning research data management is the Leibniz Network for Open Research Data (LEIBNIZ DATA). To better support researcher networking, TIB hosts the business office of the newly established ORCID DE Consortium.

LEIBNIZ DATA

LEIBNIZ DATA is designed as an infrastructure which offers a reliable and long-term service to those research infrastructures of the Leibniz Roadmap⁴ and beyond, which work with heterogeneous research data. It provides expertise on the cataloguing, archiving and subsequent use of these diverse and in some instances

⁴ Leibniz Roadmap: The Leibniz Roadmap for Research Infrastructures is a plan of the German Leibniz Association, of which the TIB is a member. It maps out how the Leibniz Association can sustainably consolidate and shape the German scientific system over the next 10 to 15 years, including the Association's own institutes. The Leibniz Roadmap contains concepts for research infrastructures which the Leibniz Association has prioritized in an internal process – with priority going to concepts which require a larger consortium of Leibniz partners and external partners. The roadmap also contributes to the incorporation of four concepts in the national prioritizing process – the National Roadmap for Research Infrastructures in Germany.

unique digital research data, and thus ensures that they remain available and usable as a central information source for scientific research and development. The data archives of the Leibniz Association's specialized and established research data centres are networked with, and made visible within the association, using international metadata standards. To enhance the handling of research data in the institutions and among the researchers, common guidelines and standards on research data management will be established. As a network, LEIBNIZ DATA thus works towards the common and collaborative (further) development of sustainable solutions for the integration of heterogeneous data (<https://www.leibniz-gemeinschaft.de/en/infrastructures/leibniz-roadmap-for-research-infrastructures/leibniz-data/>).

ORCID DE Consortium

In October 2016, the German ORCID DE Consortium was launched, establishing a new alliance to make research more accessible through the adoption of ORCID researcher identifiers (orcid.org). The ORCID (Open Researcher and Contributor ID) is an open, non-profit, community-driven effort with the goal to create and maintain a registry of unique researcher identifiers. Researchers may obtain an ORCID to distinguish themselves from other researchers while at the same time managing their records of activities (e.g. publications of papers, dissertation, research data, and attended conferences) and search for others in the registry. Another core function of ORCID is the provision of APIs that support system-to-system communication and authentication. ORCID makes the code available under an open source license, and posts an annual public data file under a CCO waiver for free download. Organisations may become members to link their records to ORCID identifiers, to update ORCID records and to register their employees and students for ORCID identifiers.

The project ORCID DE has been initiated by the German Initiative for Networking Information (DINI). It intends to promote the Open Researcher and Contributor ID for the persistent identification of researchers across academic institutions in Germany including research institutes and universities. Currently more than 40 German institutions expressed their interest in joining the consortium. Since ORCID is considered to be implemented in many universities and research institutions in Germany, the project ORCID DE aims to support the sustainable implementation of ORCID by an integrated approach. This is done by including international experience, important supra-regional infrastructures such as the Gemeinsame Normdatei (GND) and the Bielefeld Academic Search Engine (BASE), as well as publication formats beyond conventional text-based publications. The project promotes the ORCID standard by using related international groundwork by the Knowledge Exchange network or by the Confederation of Open Access Repositories (COAR) that will be edited and adapted during the project to the specific requirements of the German institutions.

Outlook

TIB as DataCite founding member and service provider for research data, audio-visual media and ORCID in Germany has learned that researchers and institutions working outside the 'big data' disciplines principally

- need assistance, on both a technical and a social level, concerning new PID standards, digital resource management and research data policies,
- need assurance and support to submit their data to data centres and publishers; presently they still keep valuable research resources in many different ways without the appropriate measures for preservation, access and re-use,
- locate research data and other digital resources in a patchy way: via colleagues, search portals and via scholarly literature. They would like to have this digital content (e.g. data and publications) linked, and
- treat digital research objects underlying scientific publications in terms of their validation, linking and accessibility in a wide variety of ways. As such, these processes often do not follow general standards or conventions and need individual consultancy of clients.

As a consequence of the COPDESS Statement of Commitment several new data policies have been published by large publishing houses including Springer-Nature and Copernicus. They explicitly recommend archiving data in public data repositories (and not "together with the paper", i.e. as classical data supplement"). In addition, new publishing forms are introduced, including data journals, which publish peer-reviewed data descriptions. The data which are published via public data repositories (and not in the article) can be cited in such data papers.

To ensure a maximum transparency and re-usable scientific content, researchers first of all need to decide what kind of digital content is produced during a scientific project, and secondly which digital content (and

data) should be shared with the respective community and the general public. Following these decisions, questions on storage, organization, metadata description and citeability need to be addressed in order to guarantee the re-use and citation of the shared resources. Research libraries such as the TIB (in cooperation with data centres and other infrastructures, e.g. DataCite) can provide support and act as multipliers for reproducible science by addressing the demands of the scientific communities on one hand and bring them together with the appropriate technical e-infrastructures and data centres on the other hand. To take further steps in the digital landscape, research libraries such as the TIB will additionally provide

- further development of skills, especially in applying metadata standards to digital objects,
- persistent identifier (PID) support around digital resources (DOI), information infrastructure and person identifiers (ORCID),
- advocating the importance of linked information for user-friendliness and ease of use of new and developing digital infrastructures,
- complimentary research data management, preservation and publication services and know-how,
- recommendations and infrastructures which securely store, curate and preserve research data, as well as,
- access to research data while providing the appropriate rights for its reuse.

Acknowledgements

We thank the information scientists, directorate and marketing officers at TIB for their contribution to this work. The RADAR project was funded by the German Research Foundation (DFG) (<http://gepris.dfg.de/gepris/projekt/237143194>). We also thank the scientific advisory board of RADAR for their contributions to our discussions on data management, research data services and infrastructures.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

All authors contributed equally to this work.

References


- Commission High Level Expert Group on the European Open Science Cloud** 2016 A Cloud on the 2020 Horizon: first report and recommendations on realising the European Science Cloud. Retrieved from: [http://ec.europa.eu/research/openscience/pdf/hleg/hleg-eosc-first-report_\(draft\).pdf](http://ec.europa.eu/research/openscience/pdf/hleg/hleg-eosc-first-report_(draft).pdf).
- C Project Collaboration to Clarify the Costs of Curation** 2016 D4.5 from Costs to Business Models. Retrieved from: <http://4cproject.eu/d4-5-from-costs-to-business-models>.
- Crosas, M** 2011 The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data. *D-Lib Magazine*, 17(1/2). DOI: <https://doi.org/10.1045/january2011-crosas>
- Data Citation Synthesis Group** 2014 Joint Declaration of Data Citation Principles. San Diego CA: FORCE11. Retrieved from: <https://www.force11.org/datacitation>.
- DataCite Metadata Working Group** 2016 DataCite Metadata Schema for the Publication and Citation of Research Data. Version 4.0. *DataCite e.V.* DOI: <https://doi.org/10.5438/0012>
- European Union** 2010 Riding the Wave: How Europe can Gain from the Rising Tide of Scientific Data. *Final Report of the High Level Expert Group on Scientific Data*, 1–40. Retrieved from: <https://www.fosteropenscience.eu/sites/default/files/pdf/831.pdf>.
- International DOI Foundation** 2012 The DOI system concept. In: *DOI Handbook*. DOI: <https://doi.org/10.1000/182>
- Kaur, K, Herterich, P, Dallmeier-Tiessen, S, Schmitt, K, Schrimpf, S, Tjalsma, H, Lambert, S and McMeekin, S** 2014 D32.1 Report on Cost Parameters for Digital Repositories. Retrieved from: http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D32_1-01-1_0_incURN.pdf.
- Keil, D** 2014 Research Data Needs from Academic Libraries: The Perspective of a Faculty Researcher. *Journal of Library Administration*, 54: 233–240. DOI: <https://doi.org/10.1080/01930826.2014.915168>
- Koltay, T** 2016 Are you ready? Tasks and roles for academic libraries in supporting research 2.0. *New Library World*, 117(1/2): 94–104. DOI: <https://doi.org/10.1108/NLW-09-2015-0062>

- Lecarpentier, D, Wittenburg, P, Elbers, W, Michelini, A, Kanso, R, Coveney, P and Baxter, R** 2013 EUDAT: A New Cross-Disciplinary Data Infrastructure for Science. *International Journal of Digital Curation*, 8(1): 279–287. DOI: <https://doi.org/10.2218/ijdc.v8i1.260>
- Paskin, N** 2006 Digital Object Identifiers for scientific data. *Data Science Journal*, 4: 12–20. DOI: <https://doi.org/10.2481/dsj.4.12>
- Pinfield, S, Cox, A M and Smith, J** 2014 Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLoS ONE*, 9(12): e114734. DOI: <https://doi.org/10.1371/journal.pone.0114734>
- Razum, M, Neumann, J and Hahn, M** 2014 RADAR – Ein Forschungsdaten-Repository als Dienstleistung für die Wissenschaft. *Zeitschrift für Bibliothekswesen und Bibliographie*, 61(1): 18–27. DOI: <https://doi.org/10.3196/186429501461150>
- RDA Europe** 2014 The Data Harvest How sharing research data can yield knowledge, jobs and growth. Retrieved from: <https://rd-alliance.org/sites/default/files/attachment/The%20Data%20Harvest%20Final.pdf>.
- Roche, D G, Kruuk, L E B, Lanfear, R and Binning, S A** 2015 Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLOS Biology*, 13(11): e1002295. DOI: <https://doi.org/10.1371/journal.pbio.1002295>
- Starr, J, Castro, E, Crosas, M, Dumontier, M, Downs, R R, Duerr, R, Haak, L L, Haendel, M, Herman, I, Hodson, S, Hourclé, J, Kratz, J E, Lin, J, Nielsen, L H, Nurnberger, A, Proell, S, Rauber, A, Sacchi, S, Smith, A, Taylor, M and Clark, T** 2015 Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science*, 1: e1. DOI: <https://doi.org/10.7717/peerj-cs.1>
- Swanson, J and Rinehart, A K** 2016 Data in context: Using case studies to generate a common understanding of data in academic libraries. *The Journal of Academic Librarianship*, 42(1): 97–101. DOI: <https://doi.org/10.1016/j.acalib.2015.11.005>
- Tenopir, C, Sandusky, R J, Allard, S and Birch, B** 2014 Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 36(2): 84–90. DOI: <https://doi.org/10.1016/j.lisr.2013.11.003>
- Treloar, A and Harboe-Ree, C** 2008 Data management and the curation continuum. How the Monash experience is informing repository relationships. In: 14th Victorian Association for Library Automation, 2008, Conference and Exhibition, Melbourne, Australia. Retrieved from: <http://arrow.monash.edu.au/hdl/1959.1/43940>.
- Wilkinson, M D** 2016 The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3: 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>
- Witt, M and Libraries for Research Data Interest Group** 2016 23 Things: Libraries for Research Data. *Be2Share*. DOI: <https://doi.org/10.15497/RDA00005>

How to cite this article: Kraft, A, Dreyer, B, Löwe, P and Ziedorn, F 2017 14 Years of PID Services at the German National Library of Science and Technology (TIB): Connected Frameworks, Research Data and Lessons Learned from a National Research Library Perspective. *Data Science Journal*, 16: 36, pp. 1–10, DOI: <https://doi.org/10.5334/dsj-2017-036>

Submitted: 21 October 2016 **Accepted:** 10 July 2017 **Published:** 17 July 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 