

PRACTICE PAPER

Expanding the Research Data Management Service Portfolio at Bielefeld University According to the Three-pillar Principle Towards Data FAIRness

Jochen Schirrwagen, Philipp Cimiano, Vidya Ayer, Christian Pietsch, Cord Wiljes, Johanna Vompras and Dirk Pieper

Bielefeld University, DE

Corresponding author: Jochen Schirrwagen (jochen.schirrwagen@uni-bielefeld.de)

Research Data Management at Bielefeld University is considered as a cross-cutting task among central facilities and research groups at the faculties. While initially started as project “Bielefeld Data Informium” lasting over seven years (2010–2015), it is now being expanded by setting up a Competence Center for Research Data. The evolution of the institutional RDM is based on the three-pillar principle: 1. Policies, 2. Technical infrastructure and 3. Support structures. The problem of data quality and the issues with reproducibility of research data is addressed in the project Conquaire. It is creating an infrastructure for the processing and versioning of research data which will finally allow publishing of research data in the institutional repository. Conquaire extends the existing RDM infrastructure in three ways: with a Collaborative Platform, Data Quality Checking, and Reproducible Research.

Keywords: Research Data Management; Reproducibility; Continuous Integration

1 Introduction

Bielefeld University, founded in 1969, is guided by a high demand for interdisciplinary research and quality of research-oriented teaching. This is reflected in the approach of “transcending boundaries”, i.e. the problem-oriented transcending of subject boundaries, between scientific cultures and between science and society and the support and promotion of openness and transparency in research. Four strategic research priorities are pursued that are interwoven across three cross-cutting themes. In addition, an interdisciplinary approach to the topic of data science is pursued. Bielefeld University¹ hosts about 25,000 students, 1,500 academic staff and 280 professors, and covers a wide range of disciplines across 13 faculties, the Graduate School and the CITEC Cluster of Excellence. Not least a third party funding of more than 20% (60,4 Mio EUR out of a 294,4 Mio EUR total budget in 2017) underlines the need for a structured, institutional research data management approach.

Based on a strong and diverse multidisciplinary research landscape at the university, two evolutions have sharply increased awareness of structured research data management at the institution a decade ago; firstly, a radical shift in research methods from “analogue” experimental laboratories and printed materials to various digital tools and research artefacts that created a demand for new research infrastructures and services. Secondly, the requirements of research funding agencies at the international and national level, which called for digital infrastructure measures that would implement and sustain the preservation of research primary data (HRK, 2014; Steering Committee for the “Digital Information” Initiative of the Alliance of Science Organisations in Germany, 2017). An important impetus for the implementation of a consistent and systematic data management approach and the development of knowledge services for research was the University’s application under the Excellence Initiative of the German Federal Ministry of Education

¹ <https://www.ub.uni-bielefeld.de/english/biblio/profil/>.

and Research and the German Research Foundation (DFG) in 2010 (Kehm, 2013). It was recognized by the university that excellent research requires an excellent research data management infrastructure.

The current concept for research data management at Bielefeld University was developed in the context of the “Bielefeld Data Informium” project in the period from 2010 to 2015, <https://data.uni-bielefeld.de/de/informium>. Aiming for a distributed service concept for research and education, jointly developed by researchers and central facilities such as the library and the computing center, in house-funded pilots were realized. A role model has been created for institutional services that support research data management, which provides central support on the one hand, while taking into account subject-specific practices (Meier zu Verl and Horstmann, 2011) and building on networking with expert groups and initiatives within and outside the university.

2 Policies, Technical Infrastructure and Support

Similar to the practice in other research libraries in Europe (Tenopir et al., 2017) the approach in “Informium” aimed to involve the library in pilots developing subject-specific research infrastructures in order to derive generic requirements for research data management and as a foundation to establish a basic service portfolio that can be offered across the campus and across disciplines. Consequently, the service portfolio is based on the three pillars of technical infrastructure, policies and support, as illustrated in **Figure 1**.

The University of Bielefeld is one of the first universities in Germany to adopt principles and guidelines on the handling of research data (2011, <https://data.uni-bielefeld.de/en/policy>), which were further developed into a resolution on research data management (2013, <https://data.uni-bielefeld.de/en/resolution>).

The guidelines and resolution emphasize the consistent consideration of research data management aspects throughout the research lifecycle and provide important guidance on how to proceed and provide basic technical and advisory services, such as:

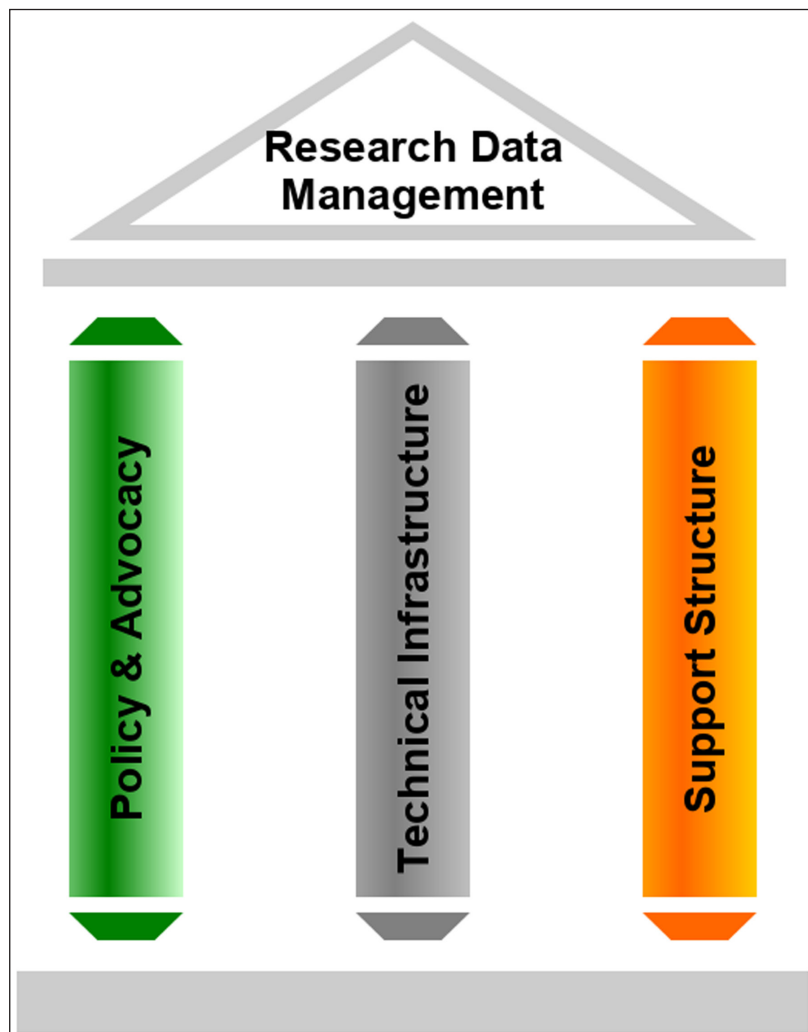


Figure 1: Three pillars of Research Data Management.

- the use of appropriate and relevant subject-specific standards from data acquisition to publication and documentation of research data
- the provision of data management plans, in particular as part of new data-intensive research proposals
- aiming that research data management in institutions and projects is making data widely available and retaining it for the long term in order to facilitate its reuse for research, practice, and the general public while simultaneously balancing the need to protect intellectual property rights, personal data, and obligations to third parties
- to promote a sustained embedding and further development of high-quality research data management by addressing subject-specific methodological training and imparting principles of good academic practice (data literacy)

For permanent support of the implementation of the research data management resolution, the coordination center for research data has been established at the library. It acts as a campus-wide point of contact for all questions on research data management and publishing as well as for concrete, temporary cooperation in pilot projects. One such example is the Data Service Center for Business and Organizational Data (DSZ-BO), developed jointly by researchers at the Faculty of Sociology and the Library, where the library has contributed its expertise in data modeling and microdata documentation in the Data Documentation Initiative (DDI) metadata framework. As an international standard DDI supports the documentation of the entire life cycle of data. A key element of this service is a data catalog for studies encoded in DDI and thus enables an interoperable exchange of studies between data centers (Edler et al., 2012).

The effect of cooperation between discipline-specific research projects and the coordination center for research data is to the mutual benefit of both sides. As part of the DSZ-BO requirements regarding the documentation of studies were implemented and special metadata requirements in organizational research in DDI-3.1 were fed back to the DDI community. But DSZ-BO was also an opportunity to roll-out and adapt a data publishing platform based on DKAN, a Drupal implementation of CKAN. It provides a flexible modular solution that combines bibliography, backup management and data visualization. Experiences gained from this pilot are implemented in an infrastructure service offering, which can now systematically be made available to other researchers and their research projects.

Furthermore the coordination center has been participating in Collaborative Research Centers (CRC) including the CRC-882 “From Heterogeneities to Inequalities”,² and the CRC-1288 INF “Practices of comparing. Ordering and changing the world”.³ Since 2007 the German Research Foundation (DFG) supports “Information Infrastructure” (INF) projects to facilitate research data management within a CRC (DFG, 2017). INF-projects provide important impetus for continuous innovation regarding the technical infrastructure for research data.

Experiences from these projects result in a number of cross-disciplinary services which form the technical infrastructure for research data management today, including e.g.

- the Data Management Planning Tool which implements funder-specific templates such as the Open Data Pilot in the EC-Horizon2020 research and innovation framework and templates for research institutes at the university
- the institutional, cross-disciplinary registration of research data and other research outputs with the Digital Object Identifier (DOI)⁴
- the participation in sciebo – the CampusCloud based on ownCloud and support of the follow-up project sciebo-rds aiming to extend sciebo as a low-entry-barrier RDM service⁵ (Vogl et al., 2016; Wilms et al., 2016)
- GitLab,⁶ as a collaborative web-based manager for distributed revision control for software and data (Ram, 2013)

Moreover the functionality of the institutional repository “Publications at Bielefeld University” (PUB) (<https://pub.uni-bielefeld.de/>) has been extended by the deposit of research data and the support of linking

² <https://sfb882.uni-bielefeld.de/de/projects/inf.html>.

³ [https://www.uni-bielefeld.de/\(en\)/sfb1288/projekte/inf.html](https://www.uni-bielefeld.de/(en)/sfb1288/projekte/inf.html).

⁴ Since 2012 the university is publication agent in DataCite.

⁵ <https://www.uni-muenster.de/ZIV/Aktuell/2018/news25042018.html>.

⁶ <https://gitlab.ub.uni-bielefeld.de>.

The screenshot displays a web interface for a research data publication. On the left, the title is "Supplementary Material for 'Infodrops: Sonification for enhanced awareness of resource consumption in the shower'" by Hammerschmidt, J., Tünnermann, R., and Hermann, T. (2013). Below the title are download links for "Infodrops_setup.png", "S3.mp3", and "S5.mp3". A diagram illustrates the experimental setup: a showerhead is connected to a "Resonating body" which is linked to a "Contact Microphone" and a "PC". The PC also receives "Water flow" and "Temperature" data. The right side of the interface shows a "Data Re-Use & License" section with a "Linked Data-Publication" link. It lists the license as "Data Re-Use License The Open Database License (ODbL) v1.0" and provides details for two main files: "Infodrops_setup.png" and "S3.mp3".

Figure 2: PUB as a hybrid repository for publications and data.

of research data with publications (see **Figure 2**) and their record in the Data Citation Index. As a quality assurance measure and to increase the acceptance of the repository by the researchers, the repository team applied successfully for the Data Seal of Approval (DSA) certification.⁷ With nearly 240 data publications the number is still low compared to research publications but with an increasing annual rate of growth of data deposits.

The technical infrastructure support is complemented by comprehensive advisory services on publication workflows and data management planning considering funders' recommendations, policies, and mandates (DFG, 2015; EC, 2016), legal aspects (such as data protection, licensing of software). It also includes support for the consideration of RDM aspects in the application for third-party funding in order to identify any issues with research data at an early stage of the research project.

Furthermore it organizes research data literacy and capacity building for students and researchers. These include biannual seminars in the staff development program for research and teaching staff. The topics are ranging from an introduction into research data management, legal aspects of research data (e.g. on conformance to the EU-General Data Protection Regulation) to managing research data and software under control with GitLab.⁸ The seminars are well attended with 12 to 15 researchers from various disciplines. Sessions of the seminars on "Research Data Management" presenting hands-on GitLab and other Open Science services of Bielefeld University,⁹ are well attended by over 50 students.¹⁰ GitLab has found its way into teaching although we did not anticipate this. When lecturers approached us with this plan, the library asked the computing centre to whitelist all students for use of GitLab.

While the university library set up GitLab in 2013 to support its software development projects it became clear very soon that GitLab has proven crucial for collaborating with researchers from other universities, which is something that many other infrastructures do not cater for. As of today Gitlab at Bielefeld University has 412 active users some of which belong to one or more of 68 groups. Users have created 641 GitLab projects so far. It is used by the DFG collaborative project RATIO¹¹ in the field of computer-assisted decision making. But also a digital humanities project is among our most active users: The Luhmann co-operative effort (<http://www.uni-bielefeld.de/soz/luhmann-archiv/>) with the University of Cologne uses our GitLab instance to annotate digitized index cards of Niklas Luhmann based on the XML language TEI.

3 Data Quality and Analytical Reproducibility

The data disclosure policies of publishers and funders (Peng, 2009; Stodden et al., 2013) require scientific research data to be published alongside the research publication to enable **scientific reproducibility** and **Open Access**. However, the scientific and academic community does not have a single or precise definition for the term *reproducibility*.

⁷ https://assessment.datasealofapproval.org/assessment_141/seal/html/.

⁸ <https://www.uni-bielefeld.de/pep/fortbildung/ub/gitlab.html>.

⁹ <http://www.uni-bielefeld.de/open-science/>.

¹⁰ <https://ekvv.uni-bielefeld.de/sinfo/publ/modul/7925150>.

¹¹ <http://www.spp-ratio.de>, <https://www.cit-ec.de/en/news/feeding-systems-arguments>.

The DFG-funded project *Continuous quality control for research data* (Conquaire, 2016–2019)¹² (Cimiano et al., 2015; Wiljes et al., 2013) at Bielefeld University supports analytical reproducibility (Ayer et al., 2017a) via a generic infrastructure architecture (Ayer et al., 2017b) that supports collaboration and easy data sharing with access to different versions of research data over time.

We consider three factors important for scientific reproducibility (see **Figure 3**):

- availability of data
- analytical reproducibility
- quality & compliance

Furthermore, by adopting the principle of continuous integration of data, we ensure regular quality checks that produces high quality scientific data within a scientific project; thereby validating the data that can be reused (Pasquetto et al., 2017) to reproduce the published analytical result.

The design and development of RDM services must maximize reuse and visibility of research data for the benefit of data producers (researchers) and data infrastructure service providers. To facilitate this process, we structured our data quality management service around Gitlab, a repository management tool built around the *Git* Distributed Version Control System (DVCS) (Ram, 2013). It is a powerful collaborative research tool with a web and command-line functionality and an in-built continuous integration (CI) system.

3.1 Data Irreproducibility Analyzer (DIRA)

The **Data Irreproducibility Analyzer (DIRA)** is the first minimal proof-of-concept-implementation of our data quality framework for checking the analytical reproducibility status of CSV files.

As shown in the Conquaire system architecture (see **Figure 4**), when a researcher commits their CSV file (s) along with the corresponding format (schema) file into their GitLab repository, the GitLab CI runners are triggered via the YAML file stored in each repository. If the format (*.fmt) file is missing, researchers receive a notification via email asking them to declare and commit the format file, which is a human-readable schema file describing their research data.

Then, when the researcher recommits the data files, they receive a status report of the data quality checks, via an email communication with a link to a HTML page displaying the results. On the front-end page, we use a triple colour encoding system to highlight the data errors on a “red”, “yellow” or “green” shade background for easy visibility.

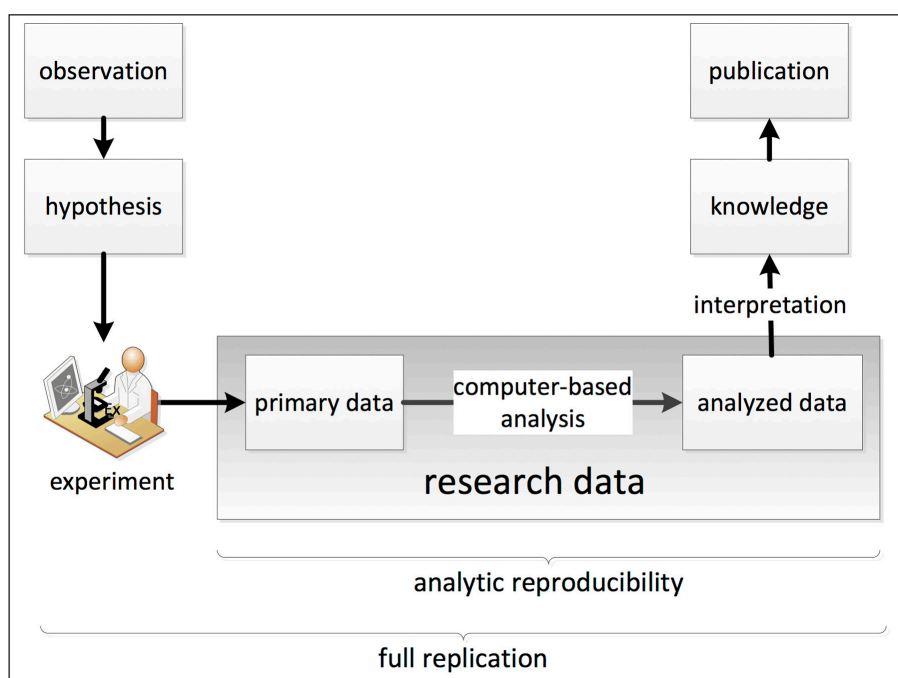


Figure 3: Research Data Lifecycle.

¹² <https://conquaire.uni-bielefeld.de>.

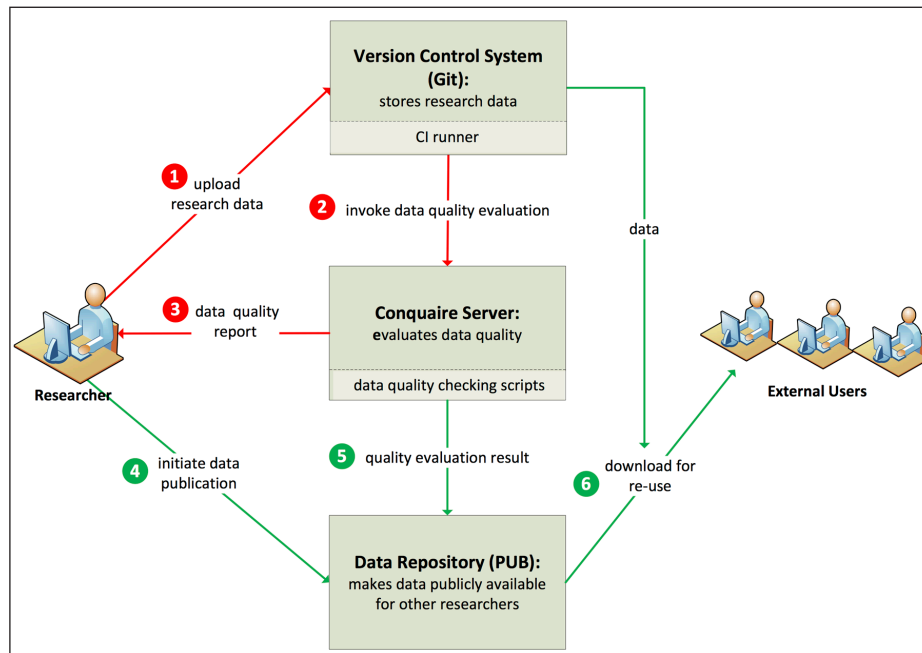


Figure 4: Conquaire System Architecture and Workflow.

The important quality check features that we have implemented in DIRA are:

- The textual definition of the column is as per the format (*.fmt) file.
- Each column in the data record has a name that appears in the format schema file, else, a warning notification is issued.
- Check whether a value is out of range or Null /NAN values.
- The range of values is specified in the format file.
- Data is plain text using a character set like ASCII/ Unicode (UTF-8).
- The file contains one record per line with the same sequence of fields.
- A data type for each value is declared and all values match the datatype.

It is proposed that the quality warnings for PUB consider some important caveats regarding notifications, viz., the researcher must be able to ignore those warnings that don't significantly hamper either the data analysis, or the data quality.

The above described research workflow based on continuous integration principles has numerous advantages, namely:

- Potential re-users of data sets have a strong starting point to assess the formal value of research data.
- Data sets published via a quality verification process using GitLab would enable researchers to establish a systematic research workflow.
- Reduced data errors via an automated workflow ensures value of data, as opposed to a data dump in the form of a *.zip file (where data contents are unknown).
- Git and the inbuilt CI system reduce the ambiguity that accompanies the traditional model of sharing data at the end of the research lifecycle.

As part of our technical support services, we provided training and conducted 6 Git workshops for all the members of our nine partner groups that provided the participants hands-on learning experience in learning to use the Gitlab collaboration tool. This training served as a foundation for the Conquaire system testing workshop that was later conducted in March 2018. Since a few research groups were already using Git, the Conquaire infrastructure structured around the DVCS is considered a useful collaboration tool.

Currently we are in the process of connecting PUB with GitLab which will allow the researcher to link their data for each publication with an assigned DOI. Thus, the data quality framework (Cimiano et al., 2015)

defined in the Conquaire project supports the FAIR data principles (Wilkinson et al., 2016) by making data findable, archivable and reusable.

3.2 Reproducibility Experiments

In order to develop a generic framework it was important to understand the research workflow of each of our case study partner groups. With nine groups from varied disciplines collaborating on the Conquaire project, we were well aware that to create a technical infrastructure geared to solve the “reproducibility” problem, we must obtain a deeper understanding of their research workflow.

To facilitate this, we undertook an intensive **Reproducibility Experiment** that entailed *analytically reproducing* one result from a paper already published by our partner groups. Our goal was to understand:

- their data analysis process
- their data collection process and the storage process
- the experiment tools used by the research groups
- software analysis and tools used during their research process
- the level of open access and open science adoption within each group

Each reproducibility experiment was conducted independently and with minimal input from the original researcher(s) after the pre-processed data and research artefacts were handed over to us. For the duration of the experiment, we worked independently and spent considerable effort in understanding their data management process. Also, we were clear that we would not change the existing workflow of the groups drastically, but rather strive for finding a collaborative way to steer the group towards Open Science in an open-ended manner.

During the course of our reproducibility experiments, we discovered numerous positive and negative practices that were prevalent within the research groups. A few research groups were already aware of Open Science methodologies and had taken measures to use Free and Open Source Software (FOSS), had some documentation about their data and the scripts, and had also chosen a license for their data and software reuse.

However, they were exceptions to the norm, and we discovered some common issues and shortcomings that hinder Open Science, as listed below (in brief):

1. **File formats:** Obscure file formats which don't always have an opensource library implementation to read the file contents. Some examples are: *.opj, *.fdt, *.sps, *.set, *.spv, *.sav, etc.
2. **Proprietary software:** Used for experiment workflows, statistical analysis and visualizations. Some examples are: Origin software, SPSS, etc.
3. **File IO issues:** Final results stored in Excel sheets where the column names were missing and the datatype changed in the middle of the columns or there were spaces and special characters in file header names. All these make it difficult to implement a generic CI system that can automatically validate the data quality.
4. **Non-homogeneous Data storage platform:** The lack of a single university-wide cloud storage platform to manage research data storage has resulted in an ad hoc data management by each research group. Some groups use private servers or sciebo which does not provide a reliable backup service, hence the *risk of data loss* is quite high.
5. **Documentation:** Many research projects did not document their data, their analysis process nor the scripts used for the statistical analysis and visualization. With respect to reproducibility, if a non-domain researcher were to be given the entire data set, there was no information available regarding the *reproducibility* aspect.
6. **Licensing:** Most research projects did not state clearly the licensing terms for their data, the software scripts and the documentation (if any). The lack of licences renders their data unusable by another researcher.
7. **Metadata & Ontologies:** Many research groups lack a standardised ontology or metadata schema. The non-availability of machine-readable metadata that describes a digital resource for their research data is a hurdle for the library in creating a semantically linked resource for the research data.

Currently, we are in the process of documenting these research findings in the form of a book that will be publicly released towards the end of 2018.

4 Next steps and prospects – the Competence Center for Research Data Management

From the perspective of Bielefeld University Library, research data management has become one of several strategically important fields of activity in recent years. This manifests itself both in the strategy paper of the library (Knorn, 2017), which is coordinated with the university administration, and on the national level in the position paper of the German Library Association which shapes objectives up to 2025 (Deutscher Bibliotheksverband, 2018).

As a result, the previous achievements in research data management at the university are well recognized. There is a growing demand for research data management in terms of human resources, coordination, and technical and subject-specific infrastructure. This is in line with the requirements of the research funding agencies, which both require subject-specific solutions for research data management and research data curation, and demand a statement in funding applications on quality assurance, management and long-term availability of research data.

Therefore by turning new challenges into opportunities the library, the computing center “BITS” and the Bielefeld Center for Data Science (BiCDaS)¹³ jointly developed a concept to build a competence center for research data that extends the already existing coordination center for research data. It is understood that RDM is a cross-cutting task of central facilities and research groups in the faculties. An inventory of previously published research data and current research projects highlights the heterogeneity of file formats and types, the variability of storage infrastructure requirements, and the bandwidth between generic and discipline-specific infrastructures. Additionally, comparisons with surveys at other academic institutions revealed that there is a primary need for:

- collaborative tools and infrastructures, such as GitLab, sciebo, wikis with a flexible and configurable rights management
- collaboratively usable databases and working platforms for authoring
- workflows for the provision of data via the Web (not necessarily publication of the data)
- structured metadata capture
- research data literacy
- long-term archiving

It is planned that by the end of the year 2018 the Competence Center for Research Data will start its work, although it will also be necessary to ensure its continuity in the long run. The main areas of work will be in the following areas, see **Figure 5**.

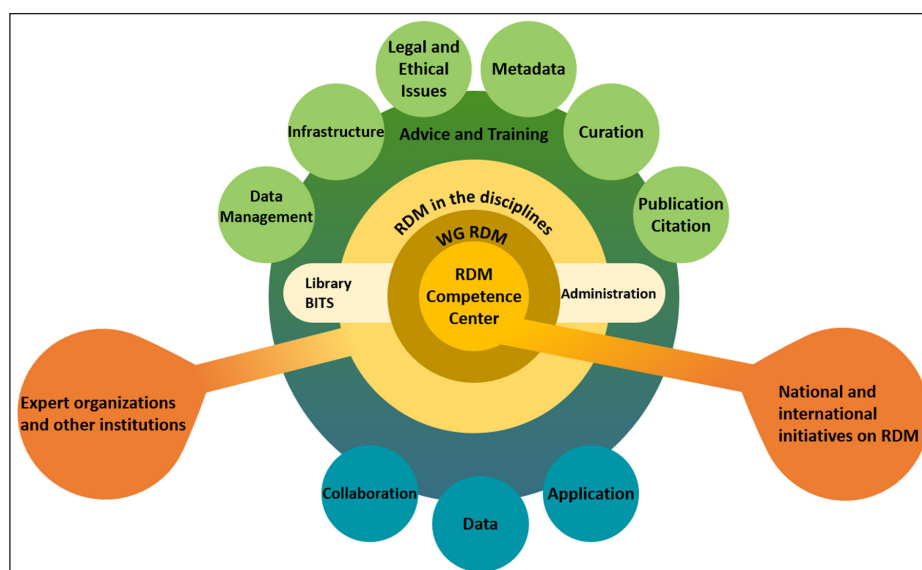


Figure 5: Joint infrastructure and support services of the RDM Competence Center.

¹³ <https://www.uni-bielefeld.de/datascience/>.

- organization and management
- maintenance and expansion of the RDM service portfolio
- operation and expansion of the technical infrastructure
- advisory support (RDM aspects in project applications, recommendations of suitable technical infrastructure, data publication, legal aspects)
- cooperation in the university, with national and international initiatives, both technical and disciplinary

Acknowledgements

This research/work was supported by the German Research Foundation (DFG) and the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

Competing Interests

The authors have no competing interests to declare.

References


- Ayer, V, Pietsch, C, Vompras, J, Schirrwagen, J, Wiljes, C, et al.** 2017a. Conquaire: Towards an architecture supporting continuous quality control to ensure reproducibility of research. *D-Lib Magazine, CNRI (USA)*, 23(1/2). DOI: <https://doi.org/10.1045/january2017-ayer>
- Ayer, V, Pietsch, C, Vompras, J, Schirrwagen, J, Wiljes, C, et al.** 2017b. Enabling Git based research data quality control for institutional repositories. <http://nbn-resolving.de/urn:nbn:de:0070-pub-29161886>. Presented at Research Data Alliance (RDA), Repository Platforms for Research Data IG, 9th Plenary meeting, Barcelona.
- Bibliotheksverband, D.** 2018. Wissenschaftliche Bibliotheken 2025. https://www.bibliotheksverband.de/fileadmin/user_upload/Sektionen/sektion4/Publikationen/WB2025_Endfassung_endg.pdf.
- Cimiano, P, McCrae, J, Jahn, N, Pietsch, C, Schirrwagen, J, et al.** 2015. CONQUAIRE: Continuous quality control for research data to ensure reproducibility: An institutional approach. DOI: <https://doi.org/10.5281/zenodo.31298>
- Deutsche Forschungsgemeinschaft.** 2015. DFG Guidelines on the Handling of Research Data. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/guidelines_research_data.pdf.
- Deutsche Forschungsgemeinschaft.** 2017. Guidelines Collaborative Research Centres. http://www.dfg.de/formulare/50_06/50_06_en.pdf.
- Edler, S, Meyermann, A, Gebel, T, Liebig, S and Diewald, M.** 2012. The German Data Service Center for Business and Organizational Data (DSC-BO). *Schmollers Jahrbuch*, 132(4): 619–634. DOI: <https://doi.org/10.3790/schm.132.4.619>
- EUROPEAN COMMISSION Directorate-General for Research & Innovation.** 2016. Guidelines on FAIR Data Management in Horizon 2020. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- German Rector's Conference (HRK).** 2014. Management of research data a key strategic challenge for university management. *Recommendation by the 16th General Meeting of the HRK on 13 May 2014 in Frankfurt/Main*. https://www.hrk.de/fileadmin/vmigrated/content_uploads/HRK_Empfehlung_Forschungsdaten_13052014_EN.pdf.
- Kehm, BM.** 2013. To be or not to be? The impacts of the excellence initiative on the German system of higher education. In: *Institutionalization of world-class university in global competition*, 81–97. Springer DOI: https://doi.org/10.1007/978-94-007-4975-7_6
- Knorn, B.** 2017. Strategieprozess der Universitätsbibliothek Bielefeld und seine Folgen. *Pro Libris*, 22(3). https://www.bibliotheken-nrw.de/fileadmin/Dateien/Daten/Aktuelles/170921-Pro_Libris-web_Einzelseiten.pdf.
- Meier zu Verl, C and Horstmann, W.** (eds.) 2011. Studies on Subject-Specific Requirements for Open Access Infrastructure. Universitätsbibliothek. DOI: <https://doi.org/10.2390/PUB-2011-1>
- Pasquetto, I, Randles, B and Borgman, C.** 2017. On the reuse of scientific data. *Data Science Journal*, 16: 8. DOI: <https://doi.org/10.5334/dsj-2017-008>
- Peng, RD.** 2009. Reproducible research and biostatistics. *Biostatistics*, 10(3): 405–408. DOI: <https://doi.org/10.1093/biostatistics/kxp014>
- Ram, K.** 2013. Git can facilitate greater reproducibility and increased transparency in science. *Source Code for Biology and Medicine*, 8(1): 7. ISSN: 1751-0473. DOI: <https://doi.org/10.1186/1751-0473-8-7>

- Steering Committee for the “Digital Information” Initiative of the Alliance of Science Organizations in Germany.** 2017. Shaping digital transformation in science. “Digital Information” Initiative by the Alliance of Science Organizations in Germany. Mission statement 2018–2022. DOI: <https://doi.org/10.2312/allianzoa.016>
- Stodden, V, Guo, P and Ma, Z.** 2013. Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals. *PLoS one*, 8(6): e67111. DOI: <https://doi.org/10.1371/journal.pone.0067111>
- Tenopir, C, Talja, S, Horstmann, W, Late, E, Hughes, D,** et al. 2017. Research data services in European academic research libraries. *Liber Quarterly*, 27(1): 23–44. ISSN: 2213-056X. DOI: <https://doi.org/10.18352/lq.10180>
- Vogl, R, Rudolph, D, Thoring, A, Angenent, H, Stieglitz, S,** et al. 2016. How to Build a Cloud Storage Service for Half a Million Users in Higher Education: Challenges Met and Solutions Found. In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 5328–5337. ISSN: 1530-1605. DOI: <https://doi.org/10.1109/HICSS.2016.658>
- Wiljes, C, Jahn, N, Lier, F, Paul-Stueve, T, Vompras, J,** et al. 2013. Towards Linked Research Data: An Institutional Approach. In: Garca Castro, A, Lange, C, Lord, P and Stevens, R (eds.), *3rd Workshop on Semantic Publishing (SePublica)*, 994: 27–38.
- Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, Appleton, G, Axton, M,** et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3. DOI: <https://doi.org/10.1038/sdata.2016.18>
- Wilms, K, Meske, C, Stieglitz, S, Rudolph, D and Vogl, R.** 2016. How to Improve Research Data Management. In: Yamamoto, S (ed.), *Human Interface and the Management of Information: Applications and Services*, 434–442. Cham: Springer International Publishing. ISBN: 978-3-319-40397-7. DOI: https://doi.org/10.1007/978-3-319-40397-7_41

How to cite this article: Schirrwagen, J, Cimiano, P, Ayer, V, Pietsch, C, Wiljes, C, Vompras, J and Pieper, D. 2019. Expanding the Research Data Management Service Portfolio at Bielefeld University According to the Three-pillar Principle Towards Data FAIRness. *Data Science Journal*, 18: 6, pp. 1–10. DOI: <https://doi.org/10.5334/dsj-2019-006>

Submitted: 30 June 2018 **Accepted:** 14 December 2018 **Published:** 15 January 2019

Copyright: © 2019 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 