Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository PUB-Publications at Bielefeld University complies with the guidelines version 2014-2017 set by the Data Seal of Approval Board.
The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on March 9, 2017.

The Trusted Digital repository is allowed to place an image of the Data Seal of Approval logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the Data Seal of Approval website.

Yours sincerely,

The Data Seal of Approval Board

Assessment Information

Guidelines Version: 2014-2017 | July 19, 2013
Guidelines Information Booklet: DSA-booklet_2014-2017.pdf
All Guidelines Documentation: Documentation

Repository: PUB-Publications at Bielefeld University
Seal Acquiry Date: Mar. 09, 2017

For the latest version of the awarded DSA   http://assessment.datasealofapproval.org/seals/
for this repository please visit our website:

Previously Acquired Seals: None

This repository is owned by: **Bielefeld University Library**

                                                 Bielefeld
                                                 Germany

                                                 T +49 521 1064051
                                                 E data@uni-bielefeld.de
                                                 W http://www.ub.uni-bielefeld.de/

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

"PUB – Publications at Bielefeld University" is used to reflect the work of the university's researchers. PUB is a hybrid institutional repository depositing and disseminating data and publications. The repository is compatible with OpenAIRE which supports and monitors the Open Access mandate in EU Horizon 2020.

Technically, PUB is based on the LibreCat [1] framework developed by the university libraries Gent, Lund and Bielefeld. Through its data processing routines for data oriented applications it facilitates the normalization of metadata and provides plugins for import and export. PUB is well integrated in the international scholarly communication infrastructure, e.g. by importing from large bibliographic databases and thematic repositories. It supports established machine interfaces (OAI-PMH, SRU, CQL) and metadata formats (Dublin Core, DataCite Metadata Kernel, MODS, XMetaDissPlus) to serve aggregative services like BASE [2] OpenAIRE, DataCite, EuropePMC [3] and DNB [4].

PUB uses the URN system of the German National Library (DNB) which operates as an URN-Resolver ensuring the permanent access to the resources. Aditionally, the submission of electronical thesis, dissertations and research literature is legally bound and exclusively subject to the Law regarding the German National Library, passed on 22 June 2006. Since then, the Deutsche Nationalbibliothek received the task (or a legal mandate) of collecting, cataloguing, indexing and archiving non-physical media works (online publications).

To ensure failure safety and reliability, PUB participates in the international distributed preservation repository network, SAFE Private LOCKSS Network [5], with the aim to preserve digital objects for future generations and to minimize the risk of data loss – caused by hardware breakdowns, obsolescence or natural disasters, or even human errors – over the long-term. The overall idea of SAFE-PLN is to make multiple copies (here: seven) as preservation strategy and to disseminate these copies throughout the world, in places considered to be safe. In the event of an unfortunate loss, data can be restored from one of the other preservation nodes, which all act in an autonomous and independent way at both financial and administrative level. In a letter of intent [8], signed in January 2014, the collaboration of the participating institutions has been officialized. The parties have commited themselves to establish and further develop the SAFE PLN network, to join efforts for the long-term preservation of academic publications and data and defined common goals regarding handling of data copies. Furthermore, they have confirmed that the intended goals will lead to a multiyear project.

PUB has been extended by several aspects of data contextualization in the course of the introduction of the institutional research data policy in 2013 [6]. As one measure Bielefeld University qualified as a publication agency for DataCite DOI. In PUB, the DOI registration of research data is part of the publication process. The persistent identification makes sure that the data stays available unchanged over time for later verification and re-use. Thus, the DOI can be used to cite the data in the manuscript. In general, the DOI resolves to a landing page in PUB, except for bilateral agreements made with research groups where the DOI may resolve to databases at their research institute.

For reseachers, PUB provides an easy-to-handle embedding possibilities of different views on the data (e.g publications of a single researcher, research group, department, or whole faculty). The chosen listings can be visualized – optionally with faceted search – on the respective institution's websites.

[1] http://librecat.org
[2] http://base-search.net

[3] http://europepmc.org/LabsLink
[4] http://www.dnb.de/DE/Wir/Kooperation/dissonline/dissonline_node.html
[5] http://www.safepln.org
[6] https://data.uni-bielefeld.de/en/resolution
[7] About PUB: https://pub.uni-bielefeld.de/docs/howto/start
[8] http://safepln.org/safe-pln-partners-officialize-their-collaboration/

[All URLs accessed 21 October 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The institutional repository PUB [1] collects and disseminates research data generated by researchers of Bielefeld University and other institutions and/or organizations that belong to the University's network. The repository's mission is to support the Good Scientific Practice by promoting data publication and enable researchers to secure and store their research data which is the basis for publication.

Quality or trust in the stored data is indirectly ensured in that way, that data provenance is known on the basis of the information about the source institution/organization/projects and the identity of the data submitter and the data creator. Additionally, a detailed description (metadata) of the data and its context (e.g linked publications) can be assessed by external users. In addition, we highly recommend to attach a comprehensible documentation of the methods employed. On the other side, we give researchers the freedom to follow their own discipline-specific rules for acquiring, selecting and processing data. The same applies where it comes to define license conditions -- or taking interests of persons and companies into account.

PUB refers to the institutional policy on research data management and gives guidance in terms of describing the composition of archival packages (e.g. README-file, licenses, documentation, etc.) that should be co-uploaded to facilitate re-use of the data. For this purpose, a key contact person together with repository staff (publication services) gives advice on general recommendations for quality-conscious research data management and corresponding archival and technical questions. Aditionally, advisory services for Data Management Planning are given by the University Library during - or even before applying for funding.

Data submitters using the self-archiving functions of PUB have to confirm that they have read the data release form. According to it, they are required to accept that "*(...) rights of third parties are not violated (...) I am aware, that I am only allowed to publish anonymous data or other data without personal reference. For publishing person-related data I have to seek the agreement of the affected persons.* In certain disciplines (e.g. psychology, epidemiology) it is additionally required that the data collection procedure must have been accepted by an ethics committee. Here, researchers are supported by the Ethics Committee of the University of Bielefeld (EUB) which examines and evaluates research projects according to ethical criteria with regard to the protection of human dignity. The assessment procedure is described in [3].

Data submitted in PUB undergoes formal validation (metadata, links etc.), but is not in the responsibility of PUB to systematically verify whether the data submitted was collected or generated according ethical and quality standards demanded in a particular scientific field. Instead, the repository refers to the University's Research Data Policy [4], according to which researchers at Bielefeld University should treat research data as a valuable academic work to "handle and document it diligently and according to appropriate subject-specific standards" (...). We specifically rely on relevant recommendations and basic rules of good scientific practice of the Deutsche Forschungsgemeinschaft of January 1998, which are part of the context information on the institutional RDM website. Thus, we assume from data stored in PUB to contain sufficient information for others to assess the scientific and scholarly quality of the research and the data submitter to act in compliance with disciplinary and ethical norms.

[1] PUB - Publications at Bielefeld University: https://pub.uni-bielefeld.de
[2] PUB Policy: https://pub.uni-bielefeld.de/docs/howto/policy
[3] Handout for the Assessement of research projects by the Ethics Committee of the University of Bielefeld
[4] Principles and guidelines on handling research data at Bielefeld University:
https://data.uni-bielefeld.de/en/policy

Further Links:
- Resolution on Research Data Management: https://data.uni-bielefeld.de/en/resolution
- DFG, Rules of Good Scientific Practice:
http://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

2. The data producer provides the data in formats recommended by the data repository.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Due to the complexity and wide variations in types and file formats of research data across the active research disciplines at the Bielefeld University, the repository does not restrict types of accepted formats. It is more in the responsibility of the researchers to follow the guidelines given by her/his research community to make sure that data is widely re-usable and the research process comprehensible.

Thus, the PUB policy refers to the University's Guidelines on Research Data Management, which can be found on the institutional RDM website. As stated there "researchers at Bielefeld University should treat research data as valuable academic work" and handle and document them "across the entire data lifecycle – from data collection to publication – (...) diligently and according to appropriate subject-specific standards.

The repository sees its task in encouraging data depositors to use standardized formats, and provides advice for conversion and data documentation (see https://data.uni-bielefeld.de/en/faq/file-requirements).

In case, non-proprietary formats cannot be bypassed, the researcher is advised to precisely document the corresponding software (version) or system which has been used to generate the data. We also recommend that data and descriptions (e.g. abstract) should not be linked to any externally available data, templates, or tools. To facilitate reproducibility, is highly recommended to co-archive these objects into the archival package. If it is not possible, the dependency (e.g. information about the software that is needed to read and process the data) to external data, should be documented in a README file.

For further details see:
- RDM website of the Bielefeld University: https://data.uni-bielefeld.de
- FAQs (mixed in english and german): https://data.uni-bielefeld.de/en/faq

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

3. The data producer provides the data together with the metadata requested by the data repository.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Metadata is attached to the research data "package" during the submission process. At a minimum, the mandatory properties of the DataCite metadata schema must be provided (and in case data is uploaded: licensing information) at the time of research data registration, otherwise data is not accepted by the system. To promote data discovery, submitters may also choose to use several optional properties to identify their data more clearly. For example:

- properties allowing for flexible description of the resource (abstract),
- relationships to other resources (or versions),
- license information,
- keywords, etc.

In addition, the data package can be liked to persons, projects, working fields and publications.

PUB repository registers data with DataCite registry (datacite.org) using the DataCite XML.
In order to provide the basis for interoperability with other data management schemas, PUB supports standard protocols and formats from the Open Repository and Digital Library Community. It is also compatible to the OpenAIRE guidelines (https://www.openaire.eu), and thus compliant to the European Commisions Open Access policy.

As export functionalities, content negotiation API is implemented and gives users the possibility to request a particular representation of the metadata. The several metadata formats can also be manually chosen from the landing page (button export "Open Data PUB") for download.

For a detailed documentation see:
- PUB API documentation: https://pub.uni-bielefeld.de/docs/api
- FAQs: https://data.uni-bielefeld.de/en/faq/metadata-requirements

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. The data repository has an explicit mission in the area of digital archiving and promulgates it.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

PUB - Publications at Bielefeld University and presents the University's repository for permanent storage and access of scientific research. Its focus is on establishing national and international visibility of the research output of Bielefeld University.

PUB offers rapid, global communication within the scientific community and encourages an open and technically and legally barrier-free access to scientific information and knowledge. It explicitly refers to the Open Access Resolution and the Resolution on Research Data Management of the Bielefeld University.

Other activities to promote PUB's mission and create awareness for a good data management are:

- Visibility of data sharing is done by making data prominent on the PUB start page (parallel to classical publications). It helps promoting data archiving and data sharing within the University's researcher groups.

- Our data publication workflows have also been communicated in several publications, e.g.:
  Repository workflow for interlinking research data with grey literature
  Vompras J, Schirrwagen J (2015)
  In: 8th Conference on Grey Literature and Repositories. 8th Conference on Grey Literature and Repositories, 12. Prague: National Library of Technology: 21-28.

  Research in Context
  Schirrwagen J, Jahn N (2013)
  Presented at the Seminar on Providing Access to Grey Literature 2013, Prague.

  Towards Linked Research Data: An Institutional Approach
  Wiljes C, Jahn N, Lier F, Paul-Stueve T, Vompras J, Pietsch C, Cimiano P (2013)
  In: 3rd Workshop on Semantic Publishing (SePublica). García Castro A, Lange C, Lord P, Stevens R (Eds); CEUR Workshop Proceedings, Aachen: 27-38.

- The Rektorat of the Bielefeld University voices its commitment to preserve and provide access to research data by adopting guidelines and comunicating its mission statement. In a resolution adopted by the Rektorat in November 2013, Bielefeld University calls on its scholars and scientists to improve the discoverability of their research data and, if possible, make them re-usable. Bielefeld University thereby became the first German university to adopt such a resolution. To introduce the service, newsletters have been published (rektorat.info). A special issue of BI.research News (No. 45, 2015), which has been published by the University -- presented different ideas in different disciplines on how to share data, which synergies might arise and how the incentives look like.

- Guidelines and recommendations are published on the research data management website (data.uni-bielefeld.de).

- Library staff of "Publishing Services" organizes teachings for promote the PUB system.

- Contact person for research data works closely with researchers to incorporate new requirements (e.g. for archiving, visualization, etc.) into the PUB system.

  Further references:
  Open Access and Research Data: http://oa.uni-bielefeld.de/en/forschungsdaten.html
  Open Access Resolution at Bielefeld University (in german): http://oa.uni-bielefeld.de/resolution.html

  [All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The repository PUB is no legal entity in its own right. It is a central service unit at the Bielefeld University which itself is governed by public law.

Data deposits are mainly done under the Open Data Licenses. The submission of the Open Data Release Form is electronically submitted by the data owner during the process of data publication. The wording of the agreement is as follows:

"I declare, that I am allowed to publish and to release the data under an Open Data License. Rights of third parties are not violated. Any joint author has given his consent to publish and to release the data under an Open Data License. I am aware, that I am only allowed to publish anonymous data or other data without personal reference. For publishing person-related data I have to seek the agreement of the affected persons. I declare, that I do not violate these rules. I am aware that this release might limit exploitation potential in terms of vending the data or achieving patents."

Thus, PUB requests confirmation from data depositors that data collection or creation was carried out in accordance with legal and ethical criteria prevailing in the data producer's research discipline (e.g. data protection legislation, ethical committees, etc.). The depositors themselves are responsible for compliance with any legal regulations in the research field data is collected. In respect to the content of the data, PUB itself does not provide – and they are not applicable – any data anonymization services or procedures to review disclosure risk in data.

Technically, PUB enables the depositors to restrict access to selected data (e.g data with disclosure risk). The data depositor can assign a visibility status to each created archival package. The options are: 1) open access, 2) university internal, or 3) restricted access (or no data upload) requiring contact information of the data creator or responsible person in the department. In case no data is added to the archival package, just metadata (including the corresponding data usage conditions) is openly accessible on the landing page.

Since the third option – (assigning permission to to data, implementing a data usage contract) is in the responsibility of the researcher, thus a breach policy for PUB is not needed.

There are a number of specific codes of conduct that PUB repository refers to, e.g.

- Rules of Good Scientific Practice [1]
- Legal Framework (Deposit Policy) as part of PUB Policy
- Principles and guidelines on handling research data at Bielefeld University

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

6. The data repository applies documented processes and procedures for managing data storage.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

PUB is running on a high-powered Linux server equipped with an internal RAID storage for the repository software and a network attached storage (iSCSI-Raid Storage) for the data.

The PUB software is running as a virtualization service built on a scalable, high-performance open-source virtualization platform (Xenserver). It provides all needed server management, monitoring (e.g. notification in case of deterioration of storage media) and administration interfaces for the implementation of the PUB system virtualization. Moreover, it offers no-downtime maintenance by allowing virtual machines and associated storage to be moved while they are running.

The server administration, backups, and monitoring is collaboratively carried out by the IT department of the University Library (LibTec) and the University's Data Center (HRZ). The repository stores its resources on its own RAID compliant server in protected by a firewall which is a standard service of the University Data Center. By using RAID, the system's reliability is automatically given by the existing redundancy.

IBM Tivoli archive service is implemented for both the PUB Software and for the data. In order to optimize data recovery processes (e.g. service provider can react more quickly), backups are done on two levels: In the first place, LibTec provides its own backup server (first backup) performing incremental backup. The procedures are described in [1]. In addition to this, PUB is integrated to the general backup strategy provided by the University Data Center (HRZ). This additional strategy in documented in [2] (in German).

The access control fulfills high security standards. The entrance to the storage location is controlled by a two-stage authentication system (an access card for the Data Center and an access token for the cage in the machine hall). Monitoring of the hardware (e.g. hard disks) is done by LibTec staff. Global monitoring of the Data Center is done by the Facility Management of the University (e.g. air-conditioning or temperature monitoring).

To ensure failure safety and reliability, PUB is connected to an uninterruptible power supply (USV system). Both, the servers and the external RAID are equipped with redundant power supply units (at least two), being connected on separate power circuits.

Moreover, PUB participates in the international distributed preservation repository network, SAFE Private LOCKSS network (SafePLN). That means in practice, that PUB repository (metadata, full texts – excluding research data) is daily mirrored to other nodes in the SafePLN network in order to disseminate copies of PUB data to places considered to be safe. In the event of an unfortunate loss (e.g. caused by a natural disaster), data could be restored from one of the other preservation nodes.

In addition, data protection procedures are documented in an internal document "Verfahrensverzeichnis" (directory of procedures) which is needed according to § 4d, § 4e BDSG (Federal Data Protection Act – BDSG) are given. They include information about:

- collection of individual-related data,
- the usage of automatic processes to that data,

- purposes of collecting, processing or using data,
- and which data protection measures are met.

[1] All documents (access control agreement with Data Center), and relevant backup and recovery procedures are documented in an internal WIKI (accesible by LibTec staff).

[2] Backup Strategy of the University's Data Center:
http://www.uni-bielefeld.de/hrz/services/backup/Sicherungssystem-Software-Dokumentation.pdf

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. The data repository has a plan for long-term preservation of its digital assets.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

Bielefeld University has a strategic plan for ensuring the long-term availability of research data and publications as part of its overall IKM strategy. We understand by long term preservation the challenge to preserve data, metadata, and documentation in a way that ensures accessibility and comprehension in the future. According to the rules of good scientific practice, the guidelines of educational policy makers [1], and research funding agencies (eg. [2]), research data should be archived in the researcher's own institution or an appropriate nationwide infrastructure for at least 10 years. PUB not only offers the organizational and technical infrastructure for implementing this requirement, but makes further efforts for long-term preservation of its digital assets.

As a bistream-preservation approach, documents and research data that are uploaded to PUB are additionally secured on the servers of the SAFE Private LOCKSS Network, with the aim to preserve digital objects for future generations and to minimize the risk of data loss caused by hardware breakdowns, obsolescence or natural disasters, or even human errors. The overall idea of SAFE-PLN is to make multiple copies (here: seven) as preservation strategy and to disseminate these copies throughout the world, in places considered to be safe. In the event of an unfortunate loss, data can be restored from one of the other preservation nodes, which all act in an autonomous and independent way at both financial and administrative level.

PUB as technical infrastructure, already supports the following preservation actions: documents and research data added in PUB are not allowed to be changed in retrospect (except metadata). The deletion for all objects is not intended; it is carried out only exceptionally and documented in writing. Since protective measures for files might interfere with long-term archiving strategies (e.g. migration or emulation), the data accepted by the repository are free of any Digital Rights Management (DRM), password protection, or limitations regarding the use of the document (copy and paste, printing).

The long-term preservation of dissertations and publication series is guaranteed by the submission of the documents to the German National Library (DNB) in the context of compulsory deliveries of online publications according to § 15 ff of DNB law. The DNB archives reported publications permanently in its sole discretion and legal mandate. The documents obtain persistent identifiers in the form of URNs ensuring ongoing access via the URN resolving service of the German National Library. In addition to PUB's bit-stream preservation we can rely on the preservation activities carried out by the DNB. Until now, there is no technical infrastructure for archiving research data at the DNB, but if it will be the case in the future, we will use this service.

In addition, we raise awareness across scientists by encouraging them to use non-proprietary file formats whenever possible and to follow the guidelines published in the FAQs (e.g. information on file formats) on the institutional RDM website. Additional services like data format migration or emulation will be planned and executed according to discipline-specific needs. The CONQUAIRE pilot projects already examines those requirements and implements preservation actions in order to ensure the comprehensibility, future readability and the semantic interpretation of the data. Together with the Semantic Computing Group at Bielefeld University the University Library has recently begun the development of a modified infrastructure which is able to store research data in that way, that continuous integration (CI) can be applied to predefined data management processes. The overall goal is to monitor the data quality at each step in the research cycle, by automatically checking if the data fulfills a number of predefined tests. PUB will be gradually enhanced by such procedures and modules implemented during this project (time frame: the next 2 years) for the data types "code", "textual" and "tabular data". This will include the monitoring activities for the technology and community needs, for example regular checks for software compilation and regular automatic checks if data formats can be still be read and interpreted. If this would be no longer the case, flags are raised to indicate that manual intervention is required.

Through a variety of active partnerships with scientific communities and data creators across the entire university, the infrastructure facilities (University Library, University's Data Center) are in a position to identify relevant changes of both technical and scientific nature and to translate them into relevant preservation actions for the PUB repository.

[1] HRK: How university management can guide the development of research data management. Orientation paths, options for action and scenarios.
https://www.hrk.de/uploads/tx_szconvention/Empfehlung_Forschungsdatenmanagement_10112015_EN_02.pdf

[2] DFG Guidelines on the Handling of Research Data:
http://www.dfg.de/en/research_funding/proposal_review_decision/applicants/submitting_proposal/research_data/

[3] CONQUAIRE: Continuous quality control for research data to ensure reproducibility: an institutional approach: http://conquaire.uni-bielefeld.de/about

[All URLs accessed 18 January 2017]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

The long term preservation of publications has been implemented. In the case of research data, PUB's long term preservation actions are mainly theoretical and developing. However, since there are preservation actions for the pilots in the CONQUAIRE project and since preservation actions will be applied for data types "code", "textual" and "tabular" during next two years, this can be seen as being in implementation phase.

8. Archiving takes place according to explicit work flows across the data life cycle.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

Since PUB' s role as an institutional and interdisciplinary repository is providing services for the University's research, we have defined clear responsibilities more for the repository workflows then for stages of the research life cycle - which vary substantially according to research disciplines.

- Data selection is mainly done by the depositor itself, with the aim to make decisions about data which should be preserved, either for use/reuse or to validate research results. We highly encourage researchers to make this decision at the time of the data creation, or if possible even earlier in accordance with a pre-established data management plan. To support researchers to find out which data might be valuable resources for publication and sharing, we have developed a set of practical recommendations (in German: https://data.uni-bielefeld.de/de/faq/data-qualify-for-publication). In this FAQ, aspects like value of the data, obligations, or re-use are taken in account.

- Data documentation: done by the depositor by providing additional information about the data (e.g. README, technical information, codebooks etc.).

- Interoperability: advice for applying standards is done by the repository staff

- Data ingest: done by the depositor
  This in an automatic process (e.g. linkage of data with publications, entry of metadata)

- Data Validation: formal consistency checks (completeness of the metadata, selected licenses, duplicate detection) are automatically carried out by the system during the self-archiving process. After that, the repository staff checks the created data packages to make sure that provenance and contextual metadata needed for an adequate discovery has been provided (e.g. linkage author to "Persons", references to publications). As

needed, the staff contacts the submitter and assist in properly formatting or re-ingesting the data.

- Data re-use options (access rights, licenses) are defined by the depositor itself during the self-archiving.

- Data Update: Existing data in the repository is modified at irregular intervals, in general as the result of error corrections (metadata can be modified by depositor after publication) or supplement of additional contextual metadata (linkage to a recently published publication).

- Data security/availability and system stability is ensured by the PUB repository

- Legal issues: depositor/PI is responsible for protecting the privacy or any subjects identifiable in the data, see: Open Data Release Form

In the future, it is planned to provide an interface to our institutional Gitlab (versioning system for software engineering) installation: Then validation and test procedures of empirical research data and evaluations will become integral part of the data ingest and data-reuse workflows in PUB. In addition, it would also be conceivable to provide a peer review data validation.

Other procedures and decision-making processes are documented in our internal University Library Wiki. The documentation includes amongst other things:

- assignment of DOI in case of special versioning requirements which vary across disciplines,
- naming conventions (e.g. discipline-specific agreements),
- handling missing data,
- missing linkage to publications
- workflows for a pre-ingest DOI assignment (in case of urgent need for submitting a paper), etc.

Staff members regularly take part in internal and external trainings on data management, metadata, long-term preservation and other relevant fields, like:

- Participation on the "NESTOR Praktiker Tag",
- GESIS Workshop: Looking after your Research Data 27.–29.06.2012
- GESIS Workshop "Digital Preservation Management - Impelemeting Short-term Strategies for Long-term Problems" (28 - 30th June 2011)
- DataCite Conferences: "Möglichkeiten und neue Lösungen im Forschungsdatenmanagement" (Cologne, 12.12.2012)
and other discipline specific national and international activities, like: IASSIST, DINI Forschungsdaten, scientific data infrastructure projects (like SFB882 INF), or participation in OpenAIRE and RDA.

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The data producer, i.e. the depositor will always remain the proprietor. In general, PUB's policy is to prefer resources that will be available under an Open Data Licence. All resources stored in PUB are available through the data portal (pub.uni-bielefeld.de/data).

PUB is using the PID system (DataCite DOIs) ensuring the continued access, availibility and validity of digital objects (e.g. PID reference of data can be used in a research paper). The corresponding landing page for the data is part of the PUB system, thus the maximum system stability is ensured. In case that external landing pages are used as PID target, we refer to a list of trustworthy repositories (re3data.org), which is a global registry of research data repositories that covers research data repositories from different academic disciplines. In this case, PUB is used to make data visible. The University's Resolution on Research Data Management says: "Bielefeld University Library supports faculties and academia to interlink existing data services with the worldwide network of data archives. At the same time, the institutional repository PUB – Publikationen an der Universität Bielefeld offers services for the publication of research data. Directories such as the DFG-funded "Registry of Research Data Repositories" also list appropriate locations for depositing data.".

PUB gives external users the possibility to search for and access the deposited files and adopts responsibility for the scientific good practice to handle the data. Data submitters using the self-archiving functions of PUB have to confirm that they have read the data release form. According to it, they are required to accept that "*(...) rights of third parties are not violated (...) I am aware, that I am only allowed to publish anonymous data or other data without personal reference. For publishing person-related data I have to seek the agreement of the affected persons.* For a successful completion of a data deposit, the selection of a license is mandatory.

Both, the participation in SafePLN Network, and the established in-house long-term data storage and data security procedures constitute an important building block for our crisis management.

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. The data repository enables the users to discover and use the data and refer to them in a persistent way.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The repository PUB provides an interface for retrieving and downloading the data in formats commonly used by the research communities. An advanced metadata search utility is provided, as well as a simple search tool for textual content. All visible metadata are indexed and searchable. Queries can contain special operators, fieldnames, wildcards etc. and results can be refined using facets by the user.

All metadata can be harvested via the OAI-PMH protocol. The OAI interface supports incremental harvesting correctly, so external service providers are enabled to update their data without having to harvest all metadata records.

Interlinks between data and publications are modeled within DataCite XML metadata schema and are trackable and globally available. See PUB API documentation for reference. Data registered with PUB (primary publication) is assigned a persistent identifier (DOI).

The landing pages of the data and publications provide a block with chosable citation styles to be copied to clipboard. This ensures an easy way to cite and refer to the published research data. Example: https://pub.uni-bielefeld.de/data/2900912#contentnegotiation

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. The data repository ensures the integrity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The integrity of the data is ensured by the version control by MD5 checksums. Checksum tests are done regularly, especially when data and metadata has to be updated by the depositor. In addition, to track the history of changes to files, metadata that is recorded and versioned with each changeset.

The availability of the files, web, and application servers is monitored continuously. We consider all objects deposited in our repository as fixed and immutable. In case a digital object is has to be updated (e.g. data changes, error correction) a new digital object is created as update and the old versions are kept in the PUB repository. However, updates of metadata for existing resources are possible without considering the result to be a new version (conform with the DataCite PID guidelines). Assigning and tracking versions is implemented on the metadata level by defining relations between assigned PUB IDs (internal identifier) of the objects.

Internal procedures and practices (like versioning procedures, organizational issues, metadata curation) are listed in the internal part of a WIKI which is accessible to appointed Bielefeld Library stuff. The link to openly accessible part of the WIKI (with some overall information) about PUB is: http://www.ub.uni-bielefeld.de/wiki/PUB.

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

12. The data repository ensures the authenticity of the digital objects and the metadata.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

The repository PUB makes the original deposited objects available in an unmodified way. Different versions of the same data publication (or further publications containing the same data objects) are not automatically compared by the PUB system itself. In the pre-ingest stage, data submitted to the repository is formally examined and checked for duplicates.

A new version of a resource gets a new persistent identifier (DOI) and the old version keeps the original one. The data depositor has the possibility to define semantic relations between them (like is-cited-by etc.). General metadata (DataCite XML) might change if the depositor or data librarian considers it necessary, e.g. in the case of misspelling or missing information. Changes to the metadata are not logged.

The identity of the depositor is formally checked when he/she logs into the PUB system by using their institutional credentials (LDAP) and upload their data.

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

Minimum Required Statement of Compliance:

3. In progress: We are in the implementation phase.

Applicant Entry

Statement of Compliance:

3. In progress: We are in the implementation phase.

Self-assessment statement:

PUB is based upon the LibreCat framework which is co-developed by University Library Bielefeld, Univeristy of Ghent and Lund. With defined workflows supported by the repository's interface, the repository aims to be as conformant to OAIS as possible.

According to the DINI Certificate, PUB supports by now the following OAIS functional entities:

- Ingest
- Archival Storage
- Data Management
- Administration
- Preservation Planning
- Access: direct access to the archived objects via the web

The OAIS model serves as a reference model for criteria catalogues for the assessment of the trustworthiness of digital archives.

PUB also is compliant to the DINI Certificate 2013 [1] (see logo on the repository website), which provides a catalog of criteria checking if a set of goals is reached by the repository. The certification process is based on aim at Open Access Repositories and Publishing Services and their inherent core components and processes.

[1] Deutsche Initiative für Netzwerk Information (DINI) is a leading certification effort to establish quality of service, visibility, interoperability and reliance on standards within institutional document and publication repositories. The DINI certificate, launched in 2003 by the Electronic Publishing working group established a minimum set of requirements for repositories and their administering institutions, covering, among others, issues of server policy, legal issues and long term availability. Although restricted to just document formats, the DINI effort represents one of the only fully implemented digital repository certification schemes. The certification requirements can be found here:
http://edoc.hu-berlin.de/series/dini-schriften/2013-3-en/PDF/dini-zertifikat-2013-en.pdf

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. The data consumer complies with access regulations set by the data repository.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

For some data sets (embargo or confidential data), the data consumer has the possibility to directly contact the data creator and to request the data with the corresponding data usage contract. In this case, just metadata is openly accessible on the landing page. Thus for the repository itself, there is no need for providing contracts or workflows for granting access to confidential data.

The mission of PUB is to promote data sharing and re-use of data carried out by the University's research. Thus, most data published in PUB is licensed with the Open Database License and is available openly to the public. The explicit licensing information can be found on each data landing page on the tab "Files" (example: http://doi.org/10.4119/unibi/2901280), so the data consumer is informed about the usage restrictions of the data she/he wants to download. Special licenses, like Creative Commons or GPL are assigned upon request to the data after consultation with the RDM contact person. See FAQ ("Other Licenses") on the institutional RDM website.

Since data is openly available, the repository does not need to provide End User License(s) with data consumers. No explicit misuse checks are done by the repository. The only thing that authors can practically do when they uncover an infringing use of their data is to make the research community aware of the misuse.

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Even though the data are openly available, a breach policy would be useful.

15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

There are a number of specific codes of conduct that PUB repository refers to, e.g.
- Rules of Good Scientific Practice [1]
- Legal Framework (Deposit Policy) as part of PUB Policy
- Principles and guidelines on handling research data at Bielefeld University

Any data user is bound by the terms and conditions of use of the repository, as soon as repository services are used. The repository provides guidance on how to use and publish data. These guidelines are available on the institutional RDM website of the Bielefeld University.

According to data deposit, it is in the responsibility of PUB users, who are uploading and publishing data to make sure that they have the necessary rights and permissions to upload and distribute it (see Open Data Declaration Form). Persons who reuse data deposited in PUB are required to respect any copyright or license related to the data when re-using or re-distributing it (see FAQ "Data Usage").

The repository itself does not carry out any structured control of compliance.

[1] http://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

16. The data consumer respects the applicable licences of the data repository regarding the use of the data.

Minimum Required Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Applicant Entry

Statement of Compliance:

4. Implemented: This guideline has been fully implemented for the needs of our repository.

Self-assessment statement:

The mission of PUB is to promote data sharing and re-use of data carried out by the University's research. Thus, most data published in PUB is licensed with the Open Database License and is available openly to the public. The explicit licensing information can be found on each data landing page on the tab "Files" (example: http://doi.org/10.4119/unibi/2901280), so the data consumer is informed about the usage restrictions of the data she/he wants to download.

If there are legal regulations or discipline-specific needs for restricting access to the data (e.g. data protection, data security, non-anonymized primary data), an access request can be done by directly contacting the data creator or the person shown as "contact person" in the metadata. The conditions for data usage (or information that data is "confidential") and contact person are described in the metadata. After an email inquiry, the depositor need to decide whether or not access is granted. For some data, explicit statements are need to be made by the data consumer about the use of the data before he/she receives the data. You can find the corresponding example for this second case by resolving the DOI presented at the following landing page: https://pub.uni-bielefeld.de/data/2767323.

No explicit measures are in place if licenses are not complied with. The only thing that authors can practically do when they uncover an infringing use of their data is to make the research community aware of the misuse.

[All URLs accessed 18 August 2016]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Even though the data are openly available, a breach policy would be useful.