



Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository: Institute for Educational Quality Improvement (IQB) - The Research Data Centre (FDZ) at IQB
Website: <https://www.iqb.hu-berlin.de/>
Certification Date: 15 May 2020
This repository is owned by: **Humboldt-Universität zu Berlin**



Institute for Educational Quality Improvement (IQB) - The Research Data Centre (FDZ) at IQB

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

Comments

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Brief Description of the Repository's Designated Community.

The Research Data Centre (FDZ) at the Institute for Educational Quality Improvement (IQB) archives and documents datasets from large national and international school assessment studies in Germany that are part of the national strategy for educational monitoring (e.g., IGLU, PISA, TIMSS, IQB National Assessment Study) as well as other national educational studies that include performance tests (e.g., ELEMENT, BiKS, ASCOT, BiLieF, BiWiss, PHONO, BiSpra) and makes them available to researchers for secondary analyses. Interested researchers may gain access to the datasets after successful application.

The FDZ at IQB's Designated Community consists of researchers from various disciplines in which educational processes are studied such as educational science, psychology, sociology, economics, and political science belonging either to universities or to non-university research institutions, students being equally welcome.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

Level of Curation Performed. Select all relevant types from:

B. Basic curation – e.g. brief checking; addition of basic metadata or documentation, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation, D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Outsource Partners. If applicable, please list them.

- Humboldt-Universität zu Berlin (1) (organizational): The IQB is affiliated with the Humboldt-Universität zu Berlin. Humboldt-Universität zu Berlin provides the FDZ at IQB with its IT infrastructure (2) (such as network, server and backup technology).
- da|ra – Registration agency for social and economic data (contractual) (3): National Centre for registration of research data from social sciences and economics, its assigns and manages the DOI for the FDZ at IQB in cooperation with DataCite and jointly with the Leibniz Information Centre for Economics (ZBW).
- IZA – Forschungsinstitut zur Zukunft der Arbeit / Institute of Labour Economics (contractual) (4): IZA developed JoSuA (Job Submission Application), a controlled remote computing system that allows users to analyse sensitive data that cannot be provided in scientific use files. The IZA provides its JoSuA service for the FDZ at IQB where it is used for remote data access as well as data access for applications from countries outside of the EU/EFTA. Like the FDZ at IQB, IZA is a member of the German Data Forum (RatSWD), fulfilling the German Data Forum's quality standards for data repositories.
- Verbund Forschungsdaten Bildung (VerbundFDB) / German Network for Educational Research Data (GNERD) (5) (organizational): The FDZ at IQB is one of the project partners forming the German Network for Educational Research Data, which aims to archive, document and make visible research data from the whole field of educational research in Germany. It can be understood as a network of research data centres that aim to standardize and synergize their services. More specifically, GNERD provides a platform for the data upload and the creation of comprehensive metadata

that is used by the FDZ at IQB, which also helped to develop it. The FDZ at the IQB, particularly aims to archive, document and make visible data and studies concerned with academic competencies and educational achievement.

(1) see https://www.hu-berlin.de/en?set_language=en

(2) see <https://www.cms.hu-berlin.de/en/dl-en>

(3) see <https://www.da-ra.de/dara/mydara/index?lang=en>

(4) see <https://josua.iza.org/>

(5) see <https://www.forschungsdaten-bildung.de/>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

Other Relevant Information.

The Institute for Educational Quality Improvement (IQB), founded in 2004, is an academic institute that supports the 16 states (Länder) of the Federal Republic of Germany—who collectively fund the institute—in improving and assuring the quality of the educational system. The IQB is responsible for regularly monitoring the extent to which Germany's schools are achieving the national educational standards through sample-based, large scale school assessment studies (educational monitoring). It also helps the states put the standards, which are intended to serve as the main framework guiding all actors in the education system, into practice (implementation). The IQB is considered one of the leading institutes actively engaged in empirical educational research in Germany. The research data centre at the IQB (FDZ at IQB) was founded in 2007 with the aim of providing access to the large scale studies of the educational monitoring program.

The FDZ at IQB also provides comprehensive information material on its datasets and offers consultation and training for researchers. FDZ at IQB has a great expertise in data curation and aims to offer a selection of well curated, high-quality datasets.

The FDZ at IQB offers its target group support in all phases of the research process (or the data lifecycle). The FDZ at IQB thus not only consults secondary data users, but also supports data providers by offering consultation on data management, data documentation, data preparation, and data provision.

The FDZ at IQB has been accredited by the German Data Forum. It follows the German Data Forum's criteria (=at least one data access path, providing sufficient data documentation as well as ensuring the long-term availability of data) (6) for

research data centres in the fulfilment of its tasks. The FDZ at IQB is one of the project partners forming the German Network for Educational Research Data (GNERD), which also provides funding for two positions at the FDZ at IQB.

The FDZ at IQB holds data of 51 studies (September 2019) and receives about 60 data applications per year (each data application project can include several researchers). Overall, 122 publications have resulted from data applications at the FDZ at IQB. This includes bachelor and master theses, dissertations, book chapters as well as articles (44 in peer-reviewed journals). From 2018 to 2019, 26 publications (14 in peer-reviewed journals) were based on data provided by the FDZ at IQB.

The FDZ at IQB can be found at [re3data.org](https://www.re3data.org) (7).

(6) see attachment no. 23, p. 6 (1.2) on the topic German Data Forum's criteria: 23_RatSWD_FDZKriterien.pdf (This document is written in German.)

(7) see <https://www.re3data.org/repository/r3d100010427>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The research data centre at the IQB (FDZ at IQB) was founded in 2007 with the aim of archiving and providing access to the large-scale studies of the educational monitoring program.(8) After initial funding by the Ministry for Research and Education (BMBF), it has been funded since 2011 by both the federal states and the BMBF as part of the Centre for International Student Assessment (ZIB). It has two major goals: (1) archiving datasets in the field of education and (2) promotion of young scientific talents in the field of educational research through providing data access and training. The FDZ at IQB aligns its yearly work programs and reports (which are approved by the governing boards of the Centre for International Student Assessment (9) and the IQB (10)) with these two goals. The most recent yearly work program and yearly report are attached (11).

The FDZ at IQB archives and documents datasets from large national and international school assessment studies in Germany that are part of the national strategy for educational monitoring (e.g., IGLU, PISA, TIMSS, IQB National Assessment Study) as well as other national educational studies that include performance tests (e.g., ELEMENT, BiKS, ASCOT, BiLieF, BiWiss, PHONO, BiSpra) and makes them available to researchers for secondary analyses. Its collection policy (12) describes in more detail the scope of datasets that the FDZ aims to archive. The FDZ at IQB provides information on and access to data. It also offers consultation and professional training for educational researchers, e.g. by holding workshops on methods of analysing large datasets.

The FDZ at IQB's services are targeted primarily at researchers who work in the field of empirical educational research inter alia with focus on school performance studies. FDZ at IQB is actively communicating its services to this target group in numerous ways (exhibition stands and talks at conferences, publications, brochures, newsletter mailings, etc.).

(8) see attachments no. 21, p. 1, Section 1 (1) on the topic Tasks and Procedures of the FDZ at IQB: 21_FDZatIQB_TasksAndProcedures.pdf (This document is written in English.), and no. 22 on the topic IQB's Charter: 22_IQB-Satzung_idFv12032015.pdf (This document is written in German.)

(9) see <https://zib.education/das-zib.html>

(10) see <https://www.iqb.hu-berlin.de/institut/impressum>

(11) see attachments no. 1 on the topic Yearly Work Program: 1_FDZ-Arbeitsprogramm_2019-2018.pdf (This document is written in German, and for internal use only.) and no. 18 on the topic Yearly Report: 18_ZIB_BerichtFDZArbeit_2019.pdf (This document is written in German, and for internal use only.)

(12) see attachment no. 2 on the topic Collection Policy: 2_Collection Policy_Sammlungsrichtlinie_final.pdf (This document is written in German and English.)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Licenses

All metadata are available free of restriction under the Creative Commons CCO 1.0 Universal Public Domain Dedication (13).

Submission of data to the FDZ at IQB: Contract with data depositors (=Data Provision Contract) (14)

When researchers deposit their data at the FDZ, the FDZ team and the researchers agree on a concept of access which governs which material will be supplied to third parties in which manner (openly on the FDZ's website, together with data by default, only on demand, or never). This concept of access is included in a data provision contract that also treats further legal issues. For example, the data depositor assures that the rights of third parties are not affected by archiving the data at the FDZ, that all other copyrights of the data depositor remain unaffected and that all parties involved adhere to generally recognised rules and regulations for academia and technology such as laid out, for example, in the DFG's White Paper "Safeguarding Good Scientific Practice".

Data use by researchers: Data Research Contract (=Contract with data user/s)/Conditions of use (15)

Datasets are provided to data users for specific research projects and not for commercial use. They are made available upon application. Applications may be submitted by researchers with an academic affiliation, typically in the field of empirical educational research, educational sciences, psychology as well as other social sciences. Academic staff and students are equally welcome to apply; eligibility is normally determined by institutional affiliation to a university or a publicly-funded research institution.

A template of a data research contract can be found online: https://www.iqb.hu-berlin.de/fdz/Muster_FDZDatenn.pdf

Applications for data access are submitted online and include a brief project outline with a summary of the theoretical underpinning, hypotheses, and planned analyses. The application should be approximately 2-3 pages long. Application guidelines, a sample project proposal and the application form are available online (16).

Staff at the FDZ at IQB first checks applications to ensure they meet the formal criteria for approval. These criteria primarily cover four areas:

- 1) Are the data to be used for purely non-commercial and scientific purposes?
- 2) Is data privacy protection respected?
- 3) Do the intended analyses correspond to contractual agreements with the data depositor?
- 4) Is it ensured that the planned analyses do not endanger ongoing qualification and publication work at the data depositors, i. e. that they do not concern research questions that have been temporarily blocked by the data depositors?

If the formal criteria are met, the applicant will be sent a data research contract (17) which must be returned signed and in duplicate to the FDZ at the IQB before access to data can be granted.

To ensure data and privacy protection, data users agree to the following in their data research contracts with the FDZ at IQB:

- Use is only allowed for scientific purposes in research and teaching.
- A transfer of rights of use of the data to third parties by the data user is inadmissible.
- No attempts may be made to re-identify people from the dataset. If an accidental re-identification happens, the FDZ has to be notified.
- No data of individual persons or groups of fewer than five persons may be reported.
- Data must be deleted on completion of the project (or on expiration of the data research contract at the latest).

Furthermore, data users undertake to comply with relevant provisions under privacy law, including the General Data Protection Regulation (GDPR), the German Federal Data Protection Act (BDSG), as well as any applicable privacy laws at state level. In case of a violation of contract terms, the right of use becomes void immediately and the data user incurs a penalty amounting to €10,000 (18).

Data received from the FDZ at IQB have to be destroyed after completion of the analyses for which they were provided.

Any person who is to be the author or co-author of any publication derived wholly or in part from the data and/or materials supplied, and who requires access to the data for this purpose, has to conclude his/her own data research contract with the FDZ at IQB.

Conditions of data use include that users may store the data received from the FDZ at IQB only on password-protected storage media or in the form of a password-protected ZIP file. Data may only be brought to countries that have an appropriate level of data protection. Data users have to comply with relevant provisions under privacy law, including the General Data Protection Regulation (GDPR), the German Federal Data Protection Act (BDSG), as well as any applicable privacy laws at state level.

Any publication that relies on the supplied materials in whole or in part must cite the FDZ at IQB and the dataset (using its DOI). After completion of the research project and/or expiration of the data research contract, all materials provided must be destroyed. For documentation purposes, data users are asked to submit to the FDZ at IQB copies of every publication derived from the data. The FDZ at IQB checks if data depositors are cited correctly in publications by data users. It is not permitted for the data user to continue using the materials, including any modified records, after expiration of the data research contract. For each occasion that a provision of the data agreement is infringed, data users shall pay to the FDZ at IQB a penalty of €10,000 (19). Moreover, the right of use immediately terminates if the terms of use are violated. Further sanctions (such as a) reprimand, b) information to the community (German Data Forum, other research data centres), c) temporary or permanent suspension of the services of the data centre for affected persons) are not planned at present.

Sensitive data

Depending on the sensitivity of the data, there are various possibilities for accessing the data:

- In the case of less sensitive data—i.e. data in which more fine-grained demographical information has been removed—the applicant is provided the requested data in the form of scientific use files, usually in SPSS format. To enhance protection, scientific use files may only be brought to countries that have an adequate level of data protection as defined by the European Commission (20). In the data research contracts is contractually ruled out that data users take any records provided by the FDZ at IQB with them if they change location to a country that does not provide this level of data protection.
- In the case of more sensitive data—i.e. data containing fine-grained demographic information— data users are provided with scientific use files in which sensitive variables have been emptied. Data users can use these scientific use files to create command files for their analyses, which are processed using the controlled remote computing system JoSuA (Job Submission Application) developed by Institute of Labour Economics (IZA). This indirect way of access ensures that data users obtain no direct hold of sensitive data, as they cannot copy the data onto their own computer. The results outputs of their calculations are checked by FDZ staff for disclosure risk and only sent on to the data user if they are considered unproblematic.

(13) see <https://creativecommons.org/publicdomain/zero/1.0/deed.de>

(14) see attachments no. 15 on the topic Data Provision Contract: 15_Data Provision Contract.pdf (This document is

written in English, and for internal use only.), and no. 4 on the topic Data Provision Contract:

4_Datenbereitstellungsvertrag_Muster.pdf (This document is written in German.)

(15) see attachment no. 3 on the topic Data research contract:

3_FDZatIQB_DataResearchContract_Datennutzungsvertrag.pdf (This document is written in German.)

(16) see <https://www.iqb.hu-berlin.de/fdz/Datenzugang/SUF-Antrag/FormularSUFs>

(17) see attachment no. 3 on the topic Data research contract:

3_FDZatIQB_DataResearchContract_Datennutzungsvertrag.pdf (This document is written in German)

(18) see attachment no. 3, p. 5 , Section 7 on the topic Data Research Contract:

3_FDZatIQB_DataResearchContract_Datennutzungsvertrag.pdf (This document is written in German.)

(19) see attachment no. 3, p. 5, Section 7 on the topic Data research contract:

3_FDZatIQB_DataResearchContract_Datennutzungsvertrag.pdf (This document is written in German.)

(20) see [https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-d
ata-non-eu-countries_de#dataprotectionincountriesoutsidetheeu](https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-data-non-eu-countries_de#dataprotectionincountriesoutsidetheeu)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The FDZ at IQB is funded by the German federal states and the BMBF until the end of 2022 as part of the ZIB. The ZIB including the FDZ at IQB is going to apply for permanent funding from 2023 onwards. The FDZ at IQB aims to continue to be the major point of access to large-scale school assessment studies in Germany. The educational monitoring strategy of the Kultusministerkonferenz (KMK, Standing Conference) will be continued indefinitely—thus datasets from the regular large-scale educational monitoring studies such as the IQB National Assessment studies and PISA will continually be produced and need to be archived.

The FDZ at IQB assures the data depositor that its research data will be archived and made available for scientific (re-) use by the scientific community. This is stated in the Data Provision Contract between the data depositor and the FDZ at IQB (21).

Data storage and archiving is guaranteed as long as the FDZ at IQB exists and for at least 10 years (in accordance to good scientific practice as defined in the rules of the German Research Foundation (22)).

To ensure the continued availability and accessibility of the data it is contractually agreed between the FDZ at IQB and the data depositors that in the event of cessation of funding resp. in case of the FDZ at IQB's dissolution, the FDZ at IQB is entitled to transfer the data to another suitable data centre under the same conditions (23). Should such a case arise, no separate consent of the data depositor shall be necessary for the purpose.

The FDZ at IQB also supports the open access policy of the Humboldt-Universität zu Berlin (24) to make publicly funded research data accessible free of cost for scientific purposes. Furthermore, researchers using data made available by the FDZ at IQB are contractually obligated to conduct their research in accordance to good scientific practice as defined in the rules of the IQB, the German Research Foundation (25) and the Humboldt-Universität zu Berlin (26).

Furthermore, the FDZ at IQB is preparing a cooperation agreement with other data centres from the GNERD network on the issue of maintaining availability and accessibility in the event of cessation of any individual data centre belonging to the GNERD network. This agreement is to regulate how the data available at the FDZ at the IQB will be handled in case the data centre is dissolved.

In addition, the FDZ at IQB provides measures for technical preservation of its datasets (see sections R15, and R16).

(21) see attachment no. 4, p. 12, § 1 (1) on the topic Data Provision Contract: 4_Datenbereitstellungsvertrag_Muster.pdf (This document is written in German, for translation see attachment no. 15, p. 9, Section 1 (1) on the topic Data Provision Contract: 15_Data Provision Contract.pdf (This document is written in English, and therefore for internal use only.)

(22) see https://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html

(23) see attachment no. 4, p. 6, § 8 (2) on the topic Data Provision Contract: 4_Datenbereitstellungsvertrag_Muster.pdf (This document is written in German, for translation see attachment no. 15, p. 6, § 8 (2) on the topic Data Provision Contract: 15_Data Provision Contract.pdf (This document is written in English, and therefore for internal use only.)

(24) see <https://www.cms.hu-berlin.de/de/dl/dataman/policy> and attachment no. 19 on the topic Research Data

Management Policy: 19_HU-Berlin-RDM-Guidelines_v1-3.pdf

(25) see https://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html

(26) see attachments no. 14 and 20 on the topic Good Scientific Practice: 14_IQB_GoodScientificPractice_GuteWissenschaftlichePraxis.pdf (This document is written in German and English, and for internal use only.), and 20_HU_SatzungWissenschaftlichesFehlverhalten.pdf (This document is written in German.)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

For the future: While you state "in the event of cessation of funding resp. in case of the FDZ at IQB's dissolution, the FDZ at IQB is entitled to transfer the data to another suitable data centre under the same conditions," do you have an existing repository or existing partnership in mind?

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Disciplinary and ethical norms

The work of the FDZ at IQB is based on a code of conduct (27) and is bound to the rules of good scientific practice by the German Research Foundation (28). The FDZ at IQB makes sure that study participants' data privacy is respected. All data deposited at and offered by the FDZ at IQB are anonymized.

The FDZ at IQB complies with the relevant data protection provisions, in particular the General Data Protection Regulation (GDPR) and the Federal Data Protection Act (BDSG).

The FDZ at IQB supports data depositors in data protection and anonymization issues by referring users to information material provided within GNERD (29) or in personal counselling. Trained staff are available and give advice to data depositors on the handling and processing of sensitive data as well as to data users on accessing and handling sensitive data and on ensuring non-disclosure of sensitive data in the presentation of their results.

More specifically, a standardized anonymization strategy for data preparation, comprehensive data research contracts (also see R2), and access limitations for sensitive data (see below) serve as organizational and technical measures to ensure data and privacy protection.

Data preparation: Anonymization

The FDZ at IQB checks incoming datasets from data depositors whether they contain any direct identifiers (such as names, postal or email addresses or telephone numbers) or other variables regarded as sensitive from a data protection standpoint (including, for example, information on income, country of origin, birth place, place of residence or other regional information, information on diseases or disabilities and string variables including information from interviews). Direct identifiers must not be included in datasets made available to users and are, therefore, removed from these datasets. Concerning the other sensitive variables, FDZ staff weighs the risk of disclosure against the potential for secondary analysis on a case-by-case basis. These appraisals are regularly reviewed by colleagues to ensure shared standards. Further strategies of anonymization then applied to the datasets include, for example, removal, aggregation, suppression, and/or recoding of variables. These strategies of anonymization decrease the risk of identifying persons in the dataset.

After consultation with data depositors, scientific use files for direct access as well as files where more fine-grained information has been retained (protected access only) are created. Prior to publication of data, the datasets are again screened by trained staff at the FDZ at IQB to ensure that all necessary steps of anonymization have been conducted. A double verification principle is followed in each step.

Data Research Contract

To ensure data and privacy protection, data users agree to conditions of use (see section R2).

Data access

The way in which data users may access the data is determined by the data's sensitivity (see also section R2).

The most common datasets provided at the FDZ at IQB are Scientific Use Files (SUFs), usually in SPSS format. SUFs are datasets that have been anonymized in such a way that no personal reference can be established without a disproportionate effort. By default the applicant is provided the requested data in the form of scientific use files.

Data users receive a list with cleared variables if applicable and can apply for remote access to those variables. Data users also receive a list with the file names and descriptions to facilitate syntax-based analyses.

As scientific use files may not be brought to countries that do not have an adequate level of data protection, data users located in such countries are not provided with scientific use files even for less sensitive data but are only given indirect access to the data via the controlled remote computing system JoSuA described above. In this case, data users can use the dummy datasets (i.e., datasets containing only the variable names and variable and value labels but no data) available for download from the FDZ at IQB's website in order to prepare the command files for their analyses.

In the case of more sensitive data datasets that are similar to the SUFs but also include more confidential information can be analysed via remote access. Each individual output is then checked by FDZ staff for disclosure risk before being sent to the data user.

Alternatively, the FDZ provides on-site access for guest researchers that want to work with more sensitive data. Similar to remote access, each output is manually checked by FDZ staff before being handed over to the data user.

Furthermore, the FDZ at IQB provide a data format for selected studies called Campus Files (CFs), designed for use in university teaching. Campus Files are completely anonymous, synthetic datasets derived from the scientific use files of the corresponding studies but containing significantly reduced numbers of cases and variables. They are available to any person with a university or research institution affiliation (evidenced by an institutional email address) who undertakes to comply with the FDZ at IQB's usage regulations for Campus Files by ticking the appropriate box on the Campus Files application form on the FDZ's website; no research proposal is required.

All datasets (SUFs, CFs and remote-access datasets) are based on data templates and are generally delivered unaltered to data users.

Complaints Office

Furthermore, the FDZ at IQB refers on its website to the Complaints Office of the German Data Forum which was opened in July 2016. If confronted with problems concerning the usage of FDZ data which cannot be solved together with the person in charge at the FDZ at IQB, users can refer these problems to the German Data Forum through an electronic contact form. The contact form as well as further information on fields of responsibility and details of the complaint procedure is to be found on the website of the German Data Forum (30).

(27) see https://www.cms.hu-berlin.de/de/dl/dataman/hu-rdm-guidelines/at_download/file,
https://fakultaeten.hu-berlin.de/en/lewi-en/research/Bestpractice?set_language=en

(28) see

https://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf

(29) see https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-3.pdf

(30) see <https://www.ratswd.de/en/info/complaints-office>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Funding/Sustainability

As mentioned in more detail above (see sections R1, and R2), the FDZ at IQB is funded by the German federal states and the BMBF until the end of 2022 as part of the ZIB; that means the funding is project-based. The ZIB including the FDZ at IQB is going to apply for permanent funding from 2023 onwards. The FDZ at IQB is governed by the IQB governing board (31) and the ZIB governing board (32). As Germany continues its educational monitoring strategy on a permanent basis, producing a continuous output of school assessment data, and as the IQB is the institution chosen by the German federal states to implement this educational monitoring strategy, it is highly unlikely that the FDZ at IQB will lose its funding in the near future.

The FDZ at IQB is affiliated to the Humboldt-Universität zu Berlin (which is permanently funded by the federal state of Berlin). Two staff positions are further funded by the GNERD network that has been described above. GNERD will also apply for permanent funding from 2022 onwards.

Qualified staff

The FDZ at IQB holds academic employees with skills in the fields of data management, research data, research data management, metadata, data protection as well as data preparation according to data protection and individual rights in addition to their disciplinary expertise in educational sciences or psychology. In total, the FDZ at IQB employs 5 researchers (including the academic head), 1 administrative staff member, 1 staff member for IT support and 4 student assistants. 2 staff members are employed on a permanent basis, 2 more for the duration of the GNERD project, the other 3 on a temporary basis with a view to doctoral or postdoctoral qualification.

Academic staff members of the FDZ at IQB hold university degrees in library and information sciences, educational science and psychology. Staff members take part in internal and external trainings on data management, metadata and other relevant fields. Furthermore, academic staff members take up doctoral studies as well as postdoctoral qualification (e. g., habilitation) in the field of empirical educational research working with datasets available at the FDZ at IQB. This facilitates a deeper understanding of the content of the datasets as well as of statistical methods necessary for analysing them.

Expertise/Exchange in the community

The FDZ at IQB is an active member of an expert community in the field of data storage, preparation and provision. Through membership or other forms of affiliation it is engaged in organizations like the German Data Forum (33), GNERD (34), RDA (35), and ZIB (36). Furthermore, the FDZ at IQB has been involved in committees and working groups developing metadata standards for educational science (“AG Metadaten” of the extended GNERD network) and recommendations for data management and data sharing in the educational sciences (“AG Forschungsdaten” of the GEBF [Gesellschaft für Empirische Bildungsforschung/Society for Empirical Educational Research], the largest German association for empirical educational research).

Particularly in the GNERD project and the German Data Forum mentioned above, there is a strong focus on harmonizing and standardizing processes across different research data centres with a view to developing and spreading “best practices”. For example, the research data centres in the GNERD network use common metadata standards and data ingest processes. The GNERD network includes large institutions such as GESIS [Leibniz-Institut für Sozialwissenschaften/Leibniz-Institute for the Social Sciences], which is the largest infrastructure-focused institute for the social sciences in Germany and already certified with the Core Trust Seal. The continuous working relationships to GESIS and other institutes have tremendously helped to further professionalize processes at the FDZ at IQB and will continue to do so.

(31) see <https://www.iqb.hu-berlin.de/institut/impressum>

(32) see <https://zib.education/das-zib.html>

(33) see <https://www.ratswd.de/en>

(34) see <https://www.forschungsdaten-bildung.de>

(35) see <https://www.rd-alliance.org>

(36) see <https://zib.education/en/home.html>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Internal advisers

Both the IQB and the FDZ at IQB have governing boards including experts from educational policy and research (IQB-Kuratorium and ZIB-Kuratorium) that govern its activities. The governing boards evaluate the FDZ's annual strategic aims and activity reports.

Both the IQB and the ZIB are evaluated at regular intervals by external experts. The FDZ at IQB is in constant exchange with the ZIB. For example, method experts at the ZIB provide the FDZ at IQB with information and workshops about new data types in the field of educational science (e.g., log files, social network data) and appropriate statistical tools to

analyse such data types.

The IQB has a Data Protection Officer that can be approached for any questions regarding data protection and privacy. When needed, the IQB also has access to consultations by lawyers of which the FDZ at IQB has previously made use (e.g., to check the data research contracts).

The members of the IQB governing board can be found here: <https://www.iqb.hu-berlin.de/institut/about>. Regarding the ZIB governing board: The ZIB's Chairperson is Prof. Dr. Kristina Reiss, Dean of the TUM School of Education. Further board members are Prof. Dr. Eckhard Klieme (DIPF) and Prof. Dr. Olaf Köller (IPN) (<http://zib.education/das-zib.html>).

External advisers

In the sections above (see section R0, and R5), it was also referred to the GNERD and the German Data Forum which provide tremendously helpful networks of research data centres in the social sciences that aim to standardize processes and develop "best practices" (see above). Through these networks, the FDZ at IQB is in regular discussion with other research data centres and has easy access to their expertise regarding organizational, technical and legal questions. Furthermore, the FDZ at IQB staff participate in conferences like Research Data Alliance (RDA) (37) and Open Science Conferences (38) to be well informed in the field of research data, metadata and open science as well as networked and up to date.

For a really high level of expertise, the GNERD advisory board could be named additionally:

<https://www.forschungsdaten-bildung.de/projekt#beirat>. The evaluation panel is always selected separately.

Feedback and input from the research community

In order to receive feedback from expert users, the FDZ at IQB conducts regular surveys of data depositors and data users. Moreover, staff members participate frequently in scientific conferences in the field of empirical educational research, the designated community of data users. Here, the FDZ at IQB engages in scientific exchange and acquires up-to date knowledge about current trends in data generation and data demands of the scientific community.

(37) see <https://www.rd-alliance.org/>

(38) see <https://www.open-science-conference.eu/>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Authenticity

Regarding the identity of data depositors, there is no formal proof of identity. However, the FDZ at IQB concludes data provision contracts with the data depositors. Data depositors must have an institutional affiliation, and every data provision contract must bear the stamp of the depositor's institution. Moreover, data provision contracts are negotiated in personal communication (by email, telephone or face to face) between staff of the FDZ and of the data depositors' institutions, which involves at least some degree of personal acquaintance.

Metadata include relevant information about authorship, members of research group and the license terms for every data package.

Integrity

Each data transfer is encrypted. Data at the FDZ at IQB is stored on a secured drive with permissions assigned so that they cannot be accessed by unauthorized persons. Backup copies of each dataset are regularly made and equally stored on a separate secured drive with permissions assigned as a safeguard against loss or destruction of data. This backup drive is stored in a safe.

The data of each study forms a closed unit and contains a definite number of research objects (dataset, accompanying material).

The integrity of the original data and the data given to data users is monitored by checksums uses the secure hash algorithm (sha1).

All checksums are checked regularly once per year. If necessary, additional checks are carried out ad hoc.

The FDZ at IQB operates a metadata management system (FDZ-StDB) suitable for documenting all processes of ingest, archival storage, and data access. This is done by checklists, metadata enrichment and information about data access. The metadata management system is accessible to authorized staff only. Regular backups are made.

The integrity of data is monitored as follows: Changes to data are documented and can be traced to the rationale and originator of the change (39). The scripts for all steps of data preparation are saved, so every change to the data is traceable and can be repeated at any time. All transformations made to data are documented and fed back with the data depositors. During the preparation process at the FDZ at IQB, the FDZ at IQB stays in contact with data depositors and prepares a data preparation form detailing all changes made from the data that were originally submitted by the depositors (e.g., further anonymization steps). This document is agreed with the data depositors and continuously adjusted.

In the context of quality management, after data preparation has been completed, the FDZ at IQB also sends an email to data depositors informing them that the data has been made available and asking them to check the metadata published on the FDZ and GNERD websites.

Ingest check

All data is transferred to the FDZ at IQB in encrypted form. Any incoming data is first decrypted (via the certificate manager Kleopatra) and then checked by FDZ staff. Ingest check follows standardized evaluation criteria (=Procedure of ingest and evaluation sheet of minimum requirements (40)); for further information see section R8.

For further information about the FDZ at IQB's workflows and processes please see section R12.

Version control

Every data package receives a DOI. This system enables transparency, visibility, citation and traceability. Through the allocation of a DOI the data can be easily cited in the long term. The FDZ at IQB publishes the metadata of research data, making them findable on the internet. Once a DOI is assigned, it can be found online and is also indexed by other services (such as GNERD, da|ra, GESIS data search, Google Dataset Search, the Web of Science, DataCite).

Furthermore, each data package receives a version number. Versioning of data is governed by a versioning guideline (41). The version history is documented in the metadata management system. All dataset versions from the original to the latest version are kept in the archive. The version control for all digital objects is ensured: after every change, the data package (according to the FDZ at IQB's guideline (42)) is assigned with a new DOI. The older version together with its DOI is transferred to the archive but remains valid. Syntax/setup files documenting the changes between the different versions are retained in addition. Data processing and data changes follow internal guidelines ensuring both data protection as well as improving data documentation. Changes in the original datasets are made in consultation with data depositors and are documented by creating a document with all suggested changes/modifications/adjustments.

(39) see attachment no. 5 on the topic Steps of Data Preparation: 5_DataPreparation_Documentation_template.pdf (This

document is written in German, and for internal use only.)

(40) see attachments no. 7 on the topic Evaluation Sheet of Minimum Requirements: 7_EvaluationSheetOfMinimumRequirements_BewertungsbogenMindestkriterien.pdf and no. 9 on the topic Procedure of Ingest: 9_ChecklistTechnicalFormal_ChecklisteTechnischFormal.pdf (These documents are written in German and English.)

(41) see attachment no. 6 on the topic Version Control: 6_VersionConcept.pdf (This document is written in English, and for internal use only.)

(42) see attachment no. 12 on the topic Guideline: 12_FDZ_Leitfaden.pdf (This document is written in German, for internal use only, and may be forwarded on request.)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Every single digital resource published on the FDZ at IQB is subject to quality control.

Collection policy

Based on its mission and funding (see sections R1, and R3), the FDZ at IQB has developed a collection policy in which

the criteria for data selection are described in detail (43).

Metadata creation

Research data are transmitted via the central file transfer portal (44) of the Research Data Education Network (GNERD) (45), a joint project of GESIS, the DIPF and the IQB. As part of the secure upload of the data, this tool mandates researchers to create comprehensive metadata to provide additional information about their project and the associated study/-ies.

Hereafter, staff at GNERD institutes determine which repository's scope the dataset falls into. Metadata of studies allocated to the FDZ at IQB are reviewed and enriched by trained FDZ staff.

The metadata scheme used at FDZ at IQB is compliant with datacite and DataCite metadata schemes. Further structural and administrative metadata are created for internal use, i. e. technical and provenance information. The FDZ at IQB's metadata scheme can be viewed on the so-called MDR – the metadata registry (46).

Generating metadata is based on standards, thesauri and other controlled vocabularies from the designated community.

Requirements for data publication

Requirements for data publication, like preferred formats and documentation standards, are published on the FDZ at IQB's website (47).

Data and accompanying material are usually delivered in well-established formats like SPSS (.sav) and Stata (.stat), as well as .pdf, .doc(x) and .xls(x). Incoming objects in non-compliant formats are converted by the FDZ at IQB staff into suitable formats.

A detailed paper jointly developed within GNERD on how to prepare quantitative data for publication is available and cross-referred on the FDZ at IQB website (48).

Data depositors can also ask for personal advice by experienced staff of the FDZ at IQB.

Data and accompanying material quality control

For all incoming data, FDZ at IQB checks whether the delivered material is complete, correct and in a suitable condition (readable, virus free, etc.). Further checks concerning the plausibility, consistency, and data protection are carried out. Inconsistencies and errors are reported to the data depositors, any necessary corrections are made and missing information is added in close cooperation with the data depositors. This ensures that the data is complete, usable and interpretable, which is an important prerequisite for all following steps.

All work steps are documented in a database. Regular meetings of the FDZ staff take place in order to optimize and harmonize and standardize procedures.

In order to standardize procedures, the whole ingest process is guided by FDZ-internal documents, checklists and record sheets which are regularly updated if necessary. The FDZ at IQB's complete workflow is attached (49).

In detail this means:

a) The ingest procedure includes:

- decrypting encoded data
- executing virus scan
- creating checksums
- checking file name, especially deleting special characters (=These are merely formal corrections without renaming the content.)
- creating working copies of the data packages (=Further data preparation will be based on those working copies.)
- checking files for functionality (=All files can be opened and (are) checked for readability.)
- checking that all datasets meet technical and documentation requirements (=clarity of dataset and documentation) as well as legal requirements (=compliance with data protection, copyright(s) (law)) (50)
- checking completeness of the data by trying to replicate published results and statistics with the datasets

b) Minimum requirements that have to be met are technical, documentation (=clarity of dataset and documentation) and legal requirements (=data protection, copyright(s) (law)) (51). On this basis, decision is made whether the dataset will be accepted and made available for secondary analysis, or rejected. Furthermore, the decision is documented in regular meetings.

c) The FDZ at IQB evaluates the data's potential for secondary analysis. An internal checklist (52) contains clearly formulated, operationalized and highly reliable criteria. However, the FDZ at IQB consistently is in communication with researchers in order to define and operationalize criteria that are accepted and deemed relevant by our community. The results of this step are used to allocate resources to those studies first that are most urgently awaited by the FDZ at IQB's designated community.

d) Checking whether data meet the requirements for anonymization and data cleaning and examining the consistency of data and documentation forms the basis on which FDZ staff create a document containing all suggested changes/modifications/adjustments to be sent to the data depositor (four-eye-principle) (53).

In addition, the FDZ at IQB evaluates the data's potential for secondary analysis and, if applicable, checks non-disclosure notices made by the data depositors. On this basis, decision is made whether the dataset will be accepted and made available for secondary analysis or rejected. The decision is documented in the minutes of the meetings held regularly by FDZ staff. All work steps are documented in a database.

Communication with data depositor

Data depositors are encouraged to provide a list of research questions currently being investigated by members of their research team. These questions are protected from secondary analyses for up to two years following submission of the data.

After all checks described above have been completed, the data depositor is contacted by email to resolve any issues that

arose. If the documentation (metadata, accompanying material) is insufficient or there are questions pertaining to the consistency of data, depositors are asked for further documentation and information that has to be supplied in order to make the data re-usable. Another document contains a list of modifications that the FDZ at IQB recommend in order to assure sufficient anonymization. In some cases, data depositors supply a new dataset, in others, the FDZ at IQB is entrusted with fixing the data according to the depositor's feedback. Moreover, a concept of access is attuned with the depositor which governs which material will be supplied to third parties in which manner (openly on website, together with data by default upon successful application for data access, only on demand, or never).

(43) see attachment no. 2 on the topic Collection Policy: 2_Collection Policy_Sammlungsrichtlinie_final.pdf (This document is written in German and English.)

(44) see <https://www.forschungsdaten-bildung.de/daten-teilen>

(45) see <https://www.forschungsdaten-bildung.de>

(46) see <https://mdr.iqb.hu-berlin.de/#/catalog/6d5cedd5-6645-5f3a-570c-a6eed177a9de>

(47) see <https://www.iqb.hu-berlin.de/fdz/Datenuebergabe>

(48) see https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-4_v1_2.pdf

(49) see attachment no. 8 on the topic Workflow: 8_Workflow_gesamt_FDZ_am_IQB_engl_final.pdf (This document is written in German and English, and for internal use only.)

(50) see attachment no. 9 on the topic Procedure of Ingest: 9_ChecklistTechnicalFormal_ChecklisteTechnischFormal.pdf (This document is written in German and English.)

(51) see attachment no. 7 on the topic Procedure of Ingest: 7_EvaluationSheetOfMinimumRequirements_BewertungsbogenMindestkriterien.pdf (This document is written in German and English.)

(52) see attachment no. 16 on the topic Data's Potential for Secondary Analysis: 16_Nachnutzungspotenzial_Kriterien.pdf (This document is written in German, and for internal use only.)

(53) see attachments no. 10 and no. 11 on the topic Templates for Data Preparation: 10_Vorlage_Studie_Skalenhandbuch.pdf, and: 11_Vorlage_Studie_Pruefung.pdf (These documents are written in German, and for internal use only.)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The IQB and therefore also the FDZ at IQB is embedded in the infrastructure of the Humboldt-Universität zu Berlin. That means, the Humboldt-Universität zu Berlin and the IQB have concluded a cooperation agreement, from which it follows that the Humboldt-Universität's Computer and Media Service (CMS) will work within the framework of a service agreement for the IQB and thus also for the FDZ at IQB, please see attached list. (54)

The department CMS administers the servers – including those used by the IQB and the IQB's FDZ – and takes care of backups, media monitoring and refreshing.

The security and data recovery plans for all the services and servers operated by the CMS are based on the "IT-Grundschutz-Katalog" (catalogue of technical, organisational, personnel, and infrastructural safeguards to guarantee an appropriate security level for all types of information of an organisation) of the Bundesamt für Sicherheit in der Informationstechnik (Federal Office for Information Security, BSI) (55).

The drives of the FDZ at IQB are included in the backup strategy (56) of the CMS. The backup service of the CMS is realized by the software product Spectrum Protect from IBM, whose former name Tivoli Storage Manager (TSM) continues to be used in the documentation and description of the software. TSM is a software system that enables automated, policy-based backup and archiving strategies for data on servers (Windows, Mac, Unix, Linux) to be centrally deployed in a network. In a convenient way, users can initiate regular backups manually and automate them. Furthermore, TSM supports longer-term archiving of files or file complexes according to a management class, thus freeing up space on local machines, as well as the possibility of saving files in an extended file system (HSM - Hierarchical Storage Management). In contrast to the backup, there are no versions during archiving and the file remains on the server, regardless of whether it still exists on the client's side or not. Archiving is recommended when users want to delete files but do not want to "throw them away", when taking a "snapshot" of individual files or directories, or when data needs to be

stored longer for legal reasons. TSM is a client-server application. For more information about security, please see section R15.

There are data recovery provisions in place. If necessary, a formal application is made to the CMS and the required state is restored within an hour.

CMS ensures the technical monitoring of storage media and installs the necessary systems. In case of problems with storage media, these are reported via a server monitoring software (information via mail to administrators) and caught by the redundancy of the storage. If storage media are considered too risky, they are disposed of and replaced. All CMS staff are qualified, instructed and authorized. Only staff with administrator rights has access to storage media.

To ensure data integrity (see section R7 above), FDZ staff regularly back up all data stored in the FDZ at IQB on a monthly basis.

A risk analysis carried out at the FDZ at IQB has shown that the maximum technical security level is implemented and that the FDZ at IQB has a uniform high security level. It works state of the art. These factors also contribute to the high confidentiality level maintained at the FDZ.

(54) see attachment no. 17 on the topic Services of CMS used by IQB: 17_Services_CMS_used_by_IQB.pdf (This document is written in English.)

(55) see https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/itgrundschutzkataloge_node.html (in German); https://www.bsi.bund.de/EN/Topics/ITGrundschutz/itgrundschutz_node.html (in English)

(56) see https://www.cms.hu-berlin.de/de/dl/systemservice/fileservice/tsm/tsm?set_language=de

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

The FDZ aims to continue to be the major point of access to large-scale school assessment studies in Germany. It has already been decided that the educational monitoring strategy of the KMK (Standing Conference) will be continued and that the FDZ at IQB will continue to archive and make available the IQB Trends in Student Achievement as well as the PISA-studies. That means, ongoing funding beyond 2022 is expected (for a more detailed description, see sections R1, and R3). Furthermore, in case of cessation of funding resp. in case of the FDZ at IQB's dissolution, it is contractually agreed (57) that the FDZ at IQB is entitled to transfer the data to another suitable data centre under the same conditions.

Within this general organizational structure, the FDZ at IQB is responsible for preserving its datasets. These are individual level data from empirical studies in the field of education in typical table form (cases in rows, variables in columns).

Data is mostly archived in formats (SPSS and/or Stata) which are widely used by the designated community and that are continuously updated (e.g., for new SPSS versions). As a preservation measure, the FDZ at IQB has chosen the strategy of format migration i. e. that the FDZ at IQB is currently in the process of additionally converting 1. SPSS-files into .csv or/and .txt-files with accompanying input-syntaxes for different statistical packets and 2. .pdf- and .doc(x)-files (e. g., accompanying material) into PDF/A-files by default.

Furthermore, the FDZ at IQB also works with "R", a programming language and open source software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. The FDZ at IQB deals with development in this field to optimize its data preparation.

Contractual matters

The contractual agreements with data depositors allow for the FDZ at IQB to make further data preparation steps that help to implement the preservation steps. For example:

- The FDZ at IQB acquires the right to process the records and the accompanying materials for the purpose of archiving and recording them in suitable formats.
- With the takeover of the materials, the FDZ at IQB acquires the right to pass on the data and the accompanying materials to interested researchers (= data users).
- The data depositor assures that (s)he is/they are entitled to hand over the data records to the FDZ at IQB for the stated purposes and that rights of third parties do not conflict.

- The FDZ at IQB is entitled to transfer the data in the event of its dissolution to another suitable data centre under the same conditions.
- The FDZ at IQB acquires the right to document the data on the basis of metadata in a suitable form, preferably in an electronic database, and to make this metadata generally accessible to the public (integration in browser and search functions within and between studies on the FDZ at IQB's website). The data depositor expressly agrees to this form of publication.
- The metadata is generated by the FDZ at IQB with a free choice of its means.

Preservation plan

A preservation plan is in progress.

(57) see attachment no. 4, p. 6, § 8 (2) on the topic Data Provision Contract: 4_Datenbereitstellungsvertrag_Muster.pdf (This document is written in German, for translation see attachment no. 15, p. 6, § 8 (2) on the topic Data Provision Contract: 15_Data Provision Contract.pdf (This document is written in English, and therefore for internal use only.)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

In the future, one would expect to see an explicit preservation plan/policy.

XI. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Academic staff members of the FDZ at IQB hold university degrees in library and information sciences, educational science and psychology. Staff members take part in internal and external trainings on data management, metadata and other relevant fields. Furthermore, academic staff members take up doctoral studies as well as postdoctoral qualification (e.g., habilitation) in the field of empirical educational research working with datasets available at the FDZ at IQB. This facilitates a deeper understanding of the content of the datasets as well as of statistical methods necessary for analysing them. Scientifically working with datasets leads to better user support by the staff at FDZ at IQB.

To ensure data quality, the FDZ at IQB has defined minimum requirements that submitted data has to meet. Data depositors have to deliver all materials and information which are necessary for secondary analysis by other data users, these are at least:

Information:

- title of the study
- primary researcher(s)
- study description
- sample
- region
- publications or references to publications based on the respective data
- time method
- research design
- (ideally: information about (used) instruments (e. g. questionnaire(s), educational measurements and tests))

Materials:

- the data itself (prepared/edited)
- codebooks /scale documentations/manuals including detailed descriptions of the scales (i.e., reliabilities, descriptive statistics) used in the studies
- information on study sample (i.e., sampling procedures, application of sampling weights); ideally: reports (technical report, final project report, if applicable)

The FDZ at IQB checks if data documentation is sufficient for researchers not involved in data collection to analyse the datasets. The FDZ at IQB assists data depositors with the documentation of datasets. Data depositors are encouraged to provide further documentation material (e. g., handbook of scales, technical report, study design, and survey instruments). Furthermore, the FDZ at IQB checks if reported results can be replicated with the submitted data and if the data corresponds with technical reports and scale documentations. Finally, the FDZ at IQB applies measures of data cleaning (e. g., correct typos, add variable labels, value labels, and missing categories) to increase data quality. All proposed

changes in the data are discussed with the data depositors with the aim of balancing data protection issues while preserving a maximum of information of the original data.

Quality of metadata is ensured by a two-step process: Firstly, metadata are supplied by data depositors via an online form which ensures adherence to the metadata scheme, controlled vocabularies, and thus facilitates comparability between data. The FDZ at IQB's metadata scheme can be seen in the metadata registry (MDR) (58). The used metadata scheme is compliant with the *da|ra* and DataCite metadata schemes. In a second step, the data and metadata are checked by experienced staff who are able to assess quality and completeness of data and documentation. Metadata then are enriched by the FDZ at IQB's staff who – for example – add keywords and categorize measured constructs. Adherence to the metadata scheme and relevant thesauri is ensured by a local database application.

Accompanying material can be viewed by potential data users on the FDZ's website, i.e. the doi landing page, if those who share the data have consented. Otherwise, these materials are delivered on demand. Thus, a broad range of information for assessing data quality is available for and findable by end users.

In order to receive feedback from expert users, the FDZ at IQB conducts regular surveys of data depositors and data users. Further on, the FDZ at IQB plans a so-called comment function for the designated community to comment on, and/or rate data and metadata.

Publication lists

The FDZ at IQB records all publications on the data named by the data depositor and regularly searches for further, latest publications on/with the data. In addition, the FDZ at IQB receives copies of publications that have been produced with the data from the data users. In this way, the FDZ at IQB tries to ensure that as far as possible complete publication lists are kept on the studies and the work with the data. The PIDs, if available, are also always included. Publications with PIDs are also listed for the corresponding study in *da|ra*.

(58) see <https://mdr.iqb.hu-berlin.de/#/catalog/6d5cedd5-6645-5f3a-570c-a6eed177a9de>

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The FDZ at IQB works along defined workflows—from data selection (according to the FDZ at IQB's collection policy) and ingest (incl. check and validation) through contracting and data preparation (incl. fine check, deep inspection, metadata enrichment and communication with data depositor) to documentation and data availability (incl. publication and release).

Following the functional model of the Open Archival Information System (OAIS) the FDZ at IQB is the archive/repository, or connector, between the data depositor and the data user. Data depositors who hand their data (=submission information package; SIP) over to the FDZ at IQB are given advice by FDZ staff; the conversion of these so-called SIPs into so-called archival information packages (AIP) takes place at the FDZ at IQB. AIPs are then submitted to the data preparation processes described above (see section R8), and converted into research data (=dissemination information package; DIP), which can be used by the data user upon application.

An overview about every single work step can be found in internal documents (59). There are templates for ingest checks, appraisal processes and data publishing procedures so it is guaranteed that different staff members arrive at substantially the same outcome.

Amongst others, the above-mentioned archiving workflows pertain to:

- selection of data (= Does the data lie within the FDZ at IQB's scope [i. e. data and studies concerned with academic competencies and educational achievement, see collection policy (60)]? If not: rejection of data and assistance with finding suitable data centres [e. g., research data centres that are organized at GNERD or at German Data Forum])
- ingest of data (=data appraisal and reuse criteria)
- contracting (incl. citation, DOI proposal)
- detailed data checking (=comparison of codebook and dataset; checking data-sensitive variables and data cleaning)

- metadata enrichment (=categorizing measured constructs)
- consultation and communication with data depositor (=creating 1., a document with all suggested changes/modifications/adjustments [steps of data preparation form], and 2., a concept of access)
- data preparation (= writing and implementing syntax; i. e., syntax for anonymization and data cleaning)
- data publication
- backup of data

For further information see also section R7.

All work steps are documented in inhouse-developed database applications such as FDZ-StDB (=FDZ study database) and FAME =FDZ database for the management of applications for data use) that are available to authorized staff only. The FDZ at IQB's metadata management system supports experienced employees in curating research data, i.e. in metadata enrichment and the categorizing of measured constructs.

Data and documentation are archived in well-defined, standardized file formats (see sections R8, and R10). Syntax and setup files documenting the changes between different versions are kept in addition. Existing internal documentation is available to all staff members.

To ensure consistency across archival copies, checksums are created for all incoming research data as well as for data given to data users. All checksums are checked regularly once per year. If necessary, additional checks are carried out ad hoc. All transformations made to data are documented. That means, the FDZ at IQB documents the data received from the data depositors as well as the data preparation form for the data depositor in order to explain which steps were made to the data during the data preparation at the FDZ at IQB (61).

All datasets (Scientific Use Files [SUFs], Campus Files [CFs] and remote-access datasets) are based on data templates and are generally not altered for data users.

Datasets are made available to interested researchers in different formats depending on the confidentiality of the datasets as well as the location of the researchers. Applications for data access are submitted online. Staff at the FDZ at IQB first checks applications to ensure they meet the formal criteria for approval. If the formal criteria are met, the applicant will be sent a data research contract (62) which must be returned signed and in duplicate to the FDZ at the IQB before access to data can be granted (see sections R2, and R4).

The workflow for preparing and sending different formats of the datasets to interested researchers (data users) is documented in the database FAME.

Each dataset contains an internal version number which facilitates individual support for data users. If variables or datasets are added to the datasets, a new version number is created and older data versions are archived. Each data version has its own DOI which allows researchers to replicate published results based on older data versions.

Regular meetings of the FDZ at IQB staff take place in order to optimize, harmonize and standardize procedures.

Regarding the change management of workflows, the FDZ at IQB has set up a task force which is in charge of implementing new workflows and documenting them in a guide for all staff members. This task force holds regular meetings to identify aspects of workflows which can be improved and redesigned. Furthermore, the task force organizes internal workshops as needed in order to explain and train the staff in following the new workflows.

(59) see attachments no. 8 on the topic Workflow: 8_Workflow_gesamt_FDZ_am_IQB_engl_final.pdf (This document is written in German and English, and for internal use only.) and no. 12 on the topic Guideline: 12_FDZ_Leitfaden.pdf (This document is written in German, for internal use only, and may be forwarded on request.)

(60) see attachment no. 2 on the topic Collection Policy: 2_Collection Policy_Sammlungsrichtlinie_final.pdf (This document is written in German and English.)

(61) see attachment no. 5 on the topic Steps of Data Preparation:

5_Dokument_Datenaufbereitung_Dokumentation_Vorlage.pdf (This document is written in German, and for internal use only.)

(62) see attachment no. 3 on the topic Data research contract:

3_FDZatIQB_DataResearchContract_Datennutzungsvertrag.pdf (This document is written in German.)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

In order to increase the visibility of the data, the FDZ at IQB generates metadata (incl. DOI and citation proposal) that describes the research data in more detail, and distributes these metadata, so they can also be fed into databases and found easily. The used metadata scheme (63) is compliant with the da|ra and DataCite metadata schemes.

Every data package receives a DOI for permanent, persistent identification, citability and indexing by other services (such as GNERD, da|ra, GESIS data search, Google Dataset Search, the Web of Science Data Citation Index, DataCite, Crossref, German Data Forum). Furthermore, the FDZ at IQB can be found in various repositories: re3data.org, DFG RIsources.

The FDZ at IQB provides information about available studies in form of metadata via website, da|ra, GNERD, reports, and newsletter.

The FDZ at IQB is member of the German Data Forum (64) as well as partner of GNERD (65) which both offer search engines on their website through which the data can be found.

In order to further publish the data, the FDZ at IQB informs data depositors about the possibility a) to make their authorship known (for example by registering with ORCID (66)) and b) to disseminate the data proof (=citation proposal)—for example via their project website(s), in publication list(s), or at Research Gate (67). This will ensure that research results are widely spread.

The FDZ at IQB's newsletter (68) regularly informs about new data publications.

Research data with DOI and citation proposal are shown on the FDZ at IQB's website (69). This is the central access point to studies that are available at the FDZ at IQB.

Metadata is also shown on the portal website of the GNERD (70) and can be searched. Via da|ra, the FDZ at IQB facilitates machine harvesting of the metadata.

In any publication derived wholly or in part from the data and/or materials supplied, data users agree always to name the FDZ at IQB and to quote the respective dataset(s), indicating the DOI(s) and the suggested citation.

(63) see <https://mdr.iqb.hu-berlin.de/#/catalog/6d5cedd5-6645-5f3a-570c-a6eed177a9de>

(64) see <https://www.ratswd.de/en>

(65) see <https://www.forschungsdaten-bildung.de/>

(66) see <https://orcid.org>

(67) see <https://www.researchgate.net>

(68) see <https://www.iqb.hu-berlin.de/fdz/news/anmeldung>

(69) see <https://www.iqb.hu-berlin.de/fdz/studies>

(70) see <https://www.forschungsdaten-bildung.de/studienliste.php>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Required metadata

All datasets provided by the FDZ at IQB are aimed at researchers conducting secondary analyses.

The central prerequisite for reuse of data is the documentation provided. Therefore, besides the data itself, additional information in the form of metadata (see sections R8, and R11 for details on the metadata) and accompanying material is necessary (e. g.: questionnaires, educational measurements and tests, codebooks, method reports, etc.). Thus, data depositors have to deliver at least the following materials to the FDZ at IQB:

- study description
- (used) instruments (e. g., questionnaire(s), educational measurements and tests)
- codebook
- reports
- publications or references to publication based on the respective data.

Understandability

After submission of the data, experienced staff checks whether data and documentation and accompanying material is complete so that the data are comprehensible and secondary analysis and reuse is possible.

Academic staff members of the FDZ at IQB hold university degrees in library and information sciences, educational science and psychology, and take up doctoral studies as well as postdoctoral qualification (e. g., habilitation) in the field of empirical educational research working with datasets available at the FDZ at IQB. This facilitates a deeper understanding of the content of the datasets as well as of statistical methods necessary for analysing them. Therefore, if needed, the FDZ at IQB supports data depositors wherever possible in preparing the necessary documentation. For further information about qualified staff, please see section R4.

Formats and migration

Data is mostly archived in formats (SPSS and/or Stata) which are widely used by the designated community and that are continuously updated (e.g., for new SPSS versions). As a preservation measure, the FDZ at IQB has chosen the strategy of format migration i. e. that the FDZ at IQB is currently in the process of additionally converting 1) SPSS-files into .csv or/and .txt-files with accompanying input-syntaxes for different statistical packets and 2) .pdf- and .doc(x)-files (e. g., accompanying material) into PDF/A-files by default.

Furthermore, the FDZ at IQB also works with “R”, a programming language and open source software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. The FDZ at IQB deals with development in this field to optimize its data preparation.

In collaboration with GNERD, the FDZ at IQB has published a list of preferred formats and required accompanying material (71).

(71) see https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-4_v1_2.pdf

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

XV. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

Technical infrastructure

The Computer and Media Service (CMS) operates the central network infrastructure of the Humboldt-Universität zu Berlin as well as the connection to the world wide web. This is put into effect through the junction to the German National Research and Education Network, DFN, the communications network for Science and research in Germany as well as the Berlin scientific network BRAIN (72). The bandwidth to the internet is at the moment 1 Gbps. The backbone net of the Humboldt-Universität zu Berlin is based on 10-Gigabit-Ethernet technology (73).

The IQB does not operate its own servers, but accesses the servers (74) operated by the CMS via desktop PCs. The desktop PCs are supported by qualified IQB staff (hardware, system software).

In order to ensure the best available performance of these systems, a four year plan is set up to project all necessary purchases of desktop hardware and system software.

Standard software for data processing is installed on all desktop PCs: STATA, SPSS, MPIus.

Installations on all systems are documented to improve technical assistance and to restore broken systems when necessary.

Metadata management system

The FDZ at IQB operates a metadata management system (so-called FDZ-StDB) suitable for documenting all processes of ingest, archival storage, and data access. The FDZ-StDB is an in-house-developed database that supports experienced employees in curating research data. This is done by checklists, metadata enrichment, categorizing of measured constructs, and information about data access. The metadata management system is accessible to authorized staff only. Regular backups are made.

The metadata management system at FDZ is based on the metadata management system at IQB, developed and maintained since 2005. Beyond the specific requirements of the FDZ, the metadata management system is at the heart of several self-developed applications and databases at the IQB that mostly relate to the management of a large pool of test-items and test booklets (since the main focus of the IQB are educational assessment studies). Therefore, the systems' performance, reliability and sustainability are highly prioritized and constantly monitoring by the IQB IT staff. Since 2019, the system has been based on XML-catalogues via a DOI referencing system, hosted at GitHub with strict version control procedures. This architecture ensures that metadata references as part of data objects can always be resolved.

All programming code (vb.net, angular/typescript) is published under MIT license at GitHub (https://doi.org/10.5159/IQB_MDR_FDZ_v1) . Beginning in 2020, the IQB is adopting modern software engineering principles like continuous integration.

(72) see <https://www.brain.de/home/view>

(73) see attachment no. 13, p. 44 on the topic Yearly Report: 13_Jahresbericht_CMS_2016_17.pdf

(74) No further information available.

Reviewer Entry

Reviewer 1

Comments:

The repository is encouraged to explore how the architecture of their systems compare to standard reference models so this can be factored in to any future CoreTrustSeal application.

Reviewer 2

Comments:

In the future, the standards that the repository uses for reference should be added.

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Security

In case of storage services, the CMS is responsible for IQB services, its servers and the underlying infrastructure (SAN, Backup, Network) (75).

The storage area network of the Humboldt-Universität zu Berlin is redundant on all levels (76). The CMS service SAN provides failsafe, freely configurable hard drive capacity for HU servers administered by local managers. Technical security is ensured through simple redundancy and the distribution of servers and storage systems to several HU buildings. Secure, locked rooms with uninterruptable power supply and air conditioning are provided at the respective locations for this purpose. Every server uses two FC-HBAs to connect to the SAN and the data is written to two mirrored (RAID1) storage devices (RAID5). These storage devices reside in different buildings of Humboldt-Universität zu Berlin. With that configuration, two HBA or more than one hard drive in the two RAID5 storage devices have to fail simultaneously for potential loss of data. Thus loss of data from hardware malfunction is very unlikely. Data loss from incorrect handling from users or even attacks from inside or outside the Humboldt network can be repaired with the backup system, which keeps deleted files for at least 60 days.

The CMS server rooms and the whole computer infrastructure are all equipped with uninterruptible power supplies, redundant air conditioning and controlled access (only CMS staff members can enter them).

All of the servers reside behind the centrally-managed firewall of the Humboldt-Universität zu Berlin (77). The network of the Humboldt-Universität zu Berlin itself is redundantly connected to the German Research Network (DFN) (78).

All of the backup data is mirrored every night on two large tape storage libraries in two separate buildings of the Humboldt-Universität zu Berlin with about 30km distance to each other. Every data loss or change in the last 60 days can be repaired from the backup service. (79)

Community-supported software is in use only for communication between project participants.

Disaster Plan

The Computer and Media Service (CMS) of the Humboldt-Universität zu Berlin administers the server and takes care of backups, media monitoring and refreshing. To protect the data, the following backup and access control procedures are in place to guarantee the (physical) safety of the digital archive holdings:

- Computing centre and server rooms are secured against unauthorized access.
- Smoke and water detectors are in place.
- Temperatures in server rooms are monitored.
- Redundant data storage in different locations
- Frequent incremental and complete backups
- Diversity of storage media and frequent media refreshment

The Humboldt-Universität Storage Area Network (SAN) provides failsafe, freely configurable hard drive capacity for HU servers administered by local managers. Technical security is ensured through simple redundancy and the distribution of servers and storage systems to several HU buildings (2 separate locations in Berlin at Grimm-Zentrum in Mitte and at Erwin-Schrödinger-Zentrum in Adlershof). Secure, locked rooms with uninterruptable power supply and air conditioning are provided at the respective locations for this purpose. Also, in the case of technical failure, malicious action or human error, CMS has mirrors of the drives at CMS. They store the last 2 weeks.

IT Security System

All CMS servers reside behind the centrally-managed firewall of the Humboldt-Universität zu Berlin (80).

Access to desktop PCs is via authentication and authorization. Authentication is done via user name and password on a specific device. Access to network resources (files, printers, components of the computer infrastructure) is only possible with prior authentication and authorization.

Sophisticated access rights and rules management limits access to resources on shared drives down to a very granular level. By policy, nothing is stored on desktop machines. Instead employees access files on network drives directly and maintain their files there.

Access to (FDZ at) IQB premises is only possible via a keycard lock system; visitors must always be accompanied by an authorised person.

In addition, regular training courses are held for temporary staff in order to avoid possible risks with regard to data handling, data protection and such like.

Business Continuity Plan

The risk analysis of the FDZ showed that a failure of the FDZ activities of up to 2 weeks is to be assessed uncritically. The services are guaranteed. With IQB, it is prepared that other premises within IQB could be used by the employees if FDZ at

IQB rooms should become unusable. In addition, there is also the option of a home office with VPN access.

Real estate is protected by a security guard (24h). Standard anti-burglary devices are used to secure the premises.

Dissemination of data

Each data transfer is encrypted.

Data packages are delivered via the secure HU file exchange portal (81).

Conditions of data use (82) include that users may store the data received from the FDZ at IQB only on password-protected storage media or in the form of a password-protected ZIP file. Data may only be brought to countries that have an appropriate level of data protection. Data users have to comply with relevant provisions under privacy law, including the General Data Protection Regulation (GDPR), the German Federal Data Protection Act (BDSG), as well as any applicable privacy laws at state level.

After completion of the research project and/or expiration of the data research contract, all materials provided must be destroyed.

(75) see https://www.cms.hu-berlin.de/en/dl-en/storage-services?set_language=en

(76) see <https://www.cms.hu-berlin.de/de/dl/systemservice/fileservice/san/standardseite>

(77) see <https://www.cms.hu-berlin.de/de/dl/netze/firewall>

(78) see <https://www.dfn.de/en/>

(79) see <https://www.cms.hu-berlin.de/de/dl/systemservice/fileservice/tsm>

(80) see <https://www.cms.hu-berlin.de/de/dl/netze/firewall>

(81) <https://www3.hu-berlin.de/dateiaustausch> (in German)

(82) see attachment no.3, § 2 (2-5) on the topic of Conditions of Data Use:

3_FDZatIQB_DataResearchContract_Datennutzungsvertrag.pdf (This document is written in German.)

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: