Long-Term Archival of operational databases at collection data centers in SIARD format - a case study from NFDI

(alternativ: Long-Term Archival of business databases from operational RDBMS in SIARD format - a case study from NFDI)

Authors:

- Journal-Vorschlag 1 CfP: Datenbank-Spektrum (Deadline 1. Februar 2024) with Instructions for Authors
- Journal-Vorschlag 2: Database: The Journal of Biological Databases and Curation
- Journal-Vorschlag 3: Biodiversity Data Journal

siehe auch

https://kb.gfbio.org/display/NFDI/2023-11-10+TA3+GDA-SNSB+SIARD+Besprechungsnotizen

https://docs.google.com/document/d/14kycACb7h6UeHznfK5Bdytgw8oQrduQstK9HxkxQiq0/edit?invite=CLyIvu8J

Sync+Share Link:
https://syncandshare.lrz.de/getlink/fiUCgapQFaUQVoXxA4F92T/BSMeryscoll

Distributed Archive als möglicher Punkt für den Artikel 2 (GDA)

Fragen: Warum Langzeitarchivierung, Herausforderungen der Archivierung bei komplexen RDBMS. Warum SIARD; was benötigt, damit die Daten archivierbar sind; Vorteil des Datenmanagements in DWB: Stellenwert der Archivierung, die begrenzt ist auf DWB als Installation bei den SNSB (Fallstudie); Etablierung einer Lösung, die für alle Installationen von DWB möglich ist; warum diese Kooperation gewählt wird; relationale Datenbank, die in Benutzung ist (anders als bei anderen SIARD-Projekten). Archivierung einer relationalen DB zu einem gewissen Zeitpunkt). Ausblick: auch für andere Datenbanksysteme möglich (SQL)

Outline

Long-Term Archival of operational databases at collection data centers in SIARD format

a case study from NFDI

Title page + author information

Authors suggested

Seifert, Weibulat, Schwartz, Weiss, Schmalzl, Triebel

Abstract (150-250 words)

xxx

1 Introduction (1.5 pages)

In natural and biodiversity sciences the number of data repositories as well as national and international data networks with long-term tasks in data management, data storage and publication was rapidly growing over the last two decades (#9). In parallel, the recognized museums and natural science collections (independently from their sponsorship and operating organizational units) extended their traditional core tasks in the direction of digitisation of their belongings of physical objects (#10). The new tasks include and will include the data publications of scientific and collection data as part of data inventories and research data infrastructure projects. Example in Germany are the contributions of the collection data centers in GFBio and NFDI4Biodiversity (see #11, #12). . In Germany . … example here: collection data centers at NFDI

- scientific data: research and long-term data management in biodiversity and geoscience research
- data publication and storage in archives
    - Why archiving is necessary: Important for reproducible science
    - ❖ challenges of archiving databases (binary data,...)

In the "Leipzig-Berlin Declaration" on cross-cutting topics of the NFDI consortia, long-term archiving and sustainable access were formulated as part of the data infrastructure to be developed. The "safeguarding of technical access and, in particular, productive re-use options for the original research and creation context"[1] are crucial [#7].

- partners of case study: GDA and SNSB

The Directorate General of the Bavarian State Archives (GDA) is an authority of the Free State of Bavaria that performs various cross-sectional tasks. It is responsible for the

[1] #7, p.4.

Bavarian state archives, their administration, organization, records management, IT services, long-term archiving and archival training. As a participant in the NFDI4Biodiverstiy consortium, the GDA is dedicated to the mobilization of historical biodiversity data and contributes its experience in the field of long-term archiving. This includes the enrichment of data with metadata, the design of archival interfaces and rights management [#1].
In addition to the close cooperation in the NFDI4Biodiversity and the common goal of keeping relevant research data available in a way that is findable, accessible, interoperable and reusable as long as possible, the two actors also work together on a purely institutional level in the long-term archiving of data of lasting value. Based on the Bavarian Archives Act, the Bavarian State Archives are responsible for archiving the data from the State Natural Science Collections that are worthy of long-term preservation. This includes research and collection data as well as data from administrative processes of lasting value. In contrast to the publication of data for science and research in the sense of first use, long-term archiving at the Bavarian State Archives also specifically archives metadata on the collection data, which serves internal administrative purposes but is sometimes crucial for the provenance of the objects and data may be relevant for secondary research purposes in new contexts. Since the data, as in many specialist procedures, is continually updated and is still required by the SNSB, time slices are archived; a now common procedure for archiving data from non-historizing systems, which is already included in the archive laws of some federal states in Germany.

The period of archiving is unlimited and includes making the data accessible for research and also completely different scientific and usage contexts.

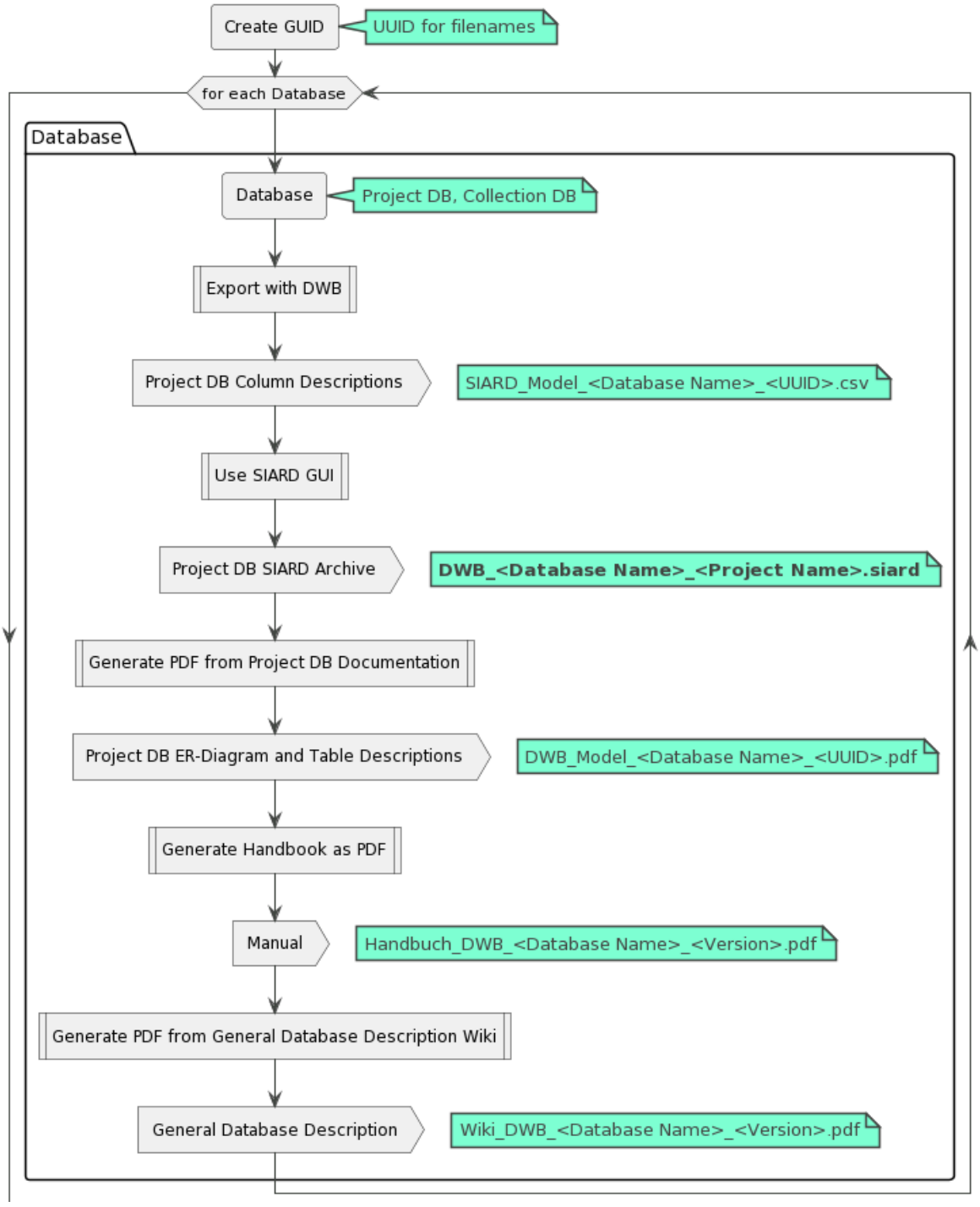
1.1. Challenges of long-term preservation of data/information (GDA)

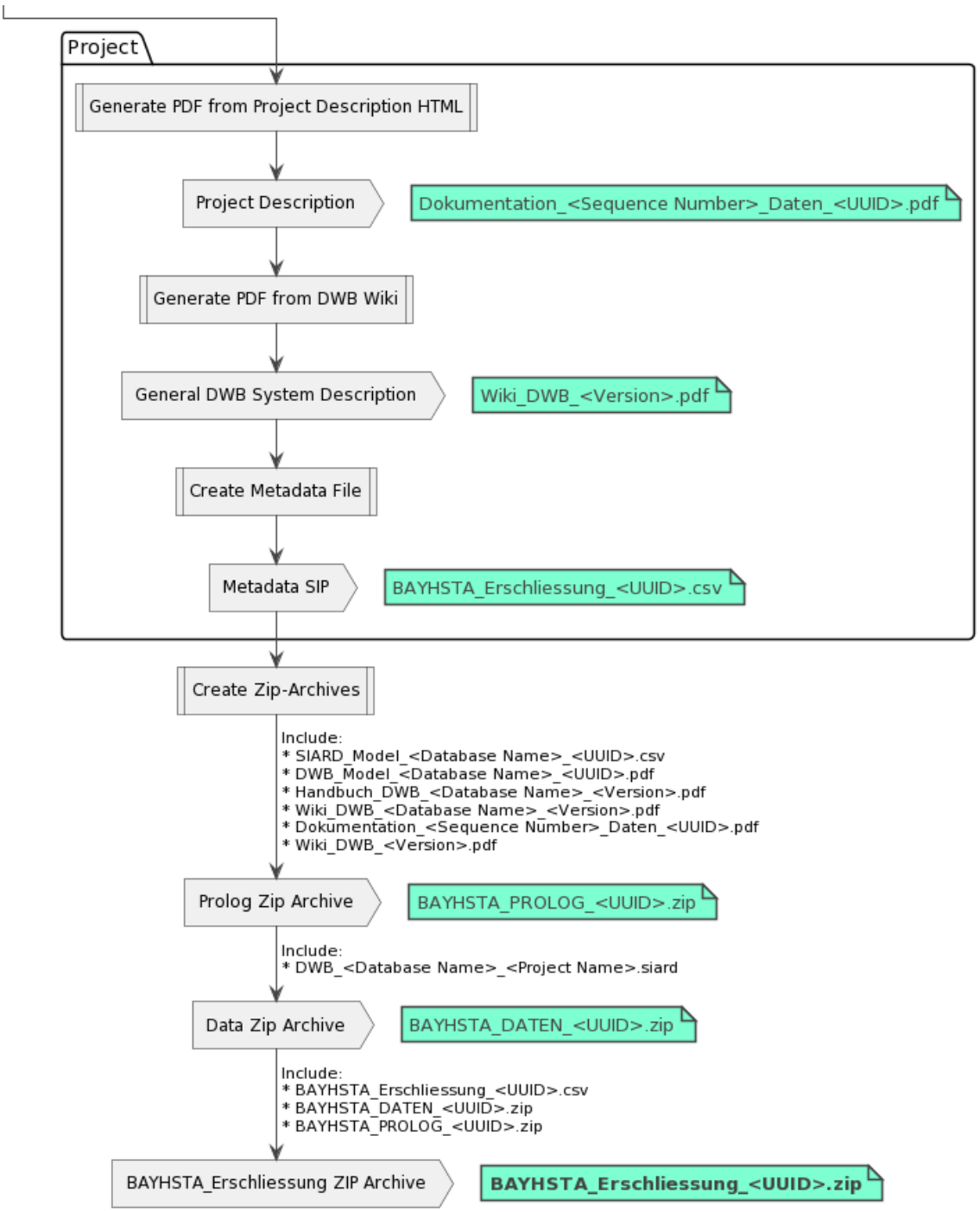
Preserving the bitstream is quite simple nowadays provided that professionally maintained servers with a geographically separated backup solution or even a professional data center are available.
The context in which the data was created and the infrastructures available at that time serve the current research purpose. The significantly more complex challenge for long-term preservation is to ensure data interpretability over several decades and longer, despite their subsequent use outside their context of production [#6]. Therefore, archives worldwide rely on the migration strategy.
This means that the data itself, regardless of its software or hardware context, is archived in formats that ensure the longest possible interpretability. These formats should, if possible, be non-proprietary, open, sufficiently documented and widely used. At certain intervals, a further migration stage into another archive format will be necessary accordingly. The same data will therefore be stored in the long-term archives in multiple representations over the decades and centuries [#8].[2] This means that in order to be kept interpretable in the long term, data must be exported from its original software and a series of specifications must be

[2] Vgl. Repräsentationenmodell nach PREMIS

made for this in the form of logical archiving interfaces. In particular, it must be determined which data must be transferred because it is worth of archiving, in which formats, in which structure and how it must be explained in order to be able to interpret it permanently. In addition, documentation material for the original system must be archived in order to give later users the opportunity to trace the context of origin, the provenance, of the data. And finally, the data must be explained with metadata that allows semantic interpretation of the data even many years from now. It must also be determined which information is significant and must be retained over future migration cycles. The formats and the transfer structure must then be determined for this data as well as for the documentation material. Finally, it is important to decide how the data should be transported to the archive.
In this case, too, the basis for long-term digital archiving is provided by the international standard OAIS, which fundamentally defines how a digital archive works and defines important terms. The Open Archival Information System (OAIS) is a reference model for the exchange, preservation and long-term availability and interpretability of information; regardless of its type of representation [#3]. The standard (ISO 14721) was developed in a public process and is therefore freely available for use. Based on the data submitted by producers for the archive repository, the submission information package (SIP), the data is stored in the archive repository as an archival information package (AIP). In addition to the submission, this package also contains information on the context of origin and documents that ensure the interpretability of the information (e.g. manuals, data models, etc.). If there is a demand for the archive material, a dissemination information package (DIP) is delivered, which may also contain information on access and usage rights [#4]. Various public archives and libraries have been active in building OAIS-compliant digital long-term archives for many years, and the Nestor network is an expert network for the exchange of knowledge and the development of important fundamentals

Workflow-Grafik/Gesamtgrafik am Anfang, davon ausgehend den Text:


Create GUID
UUID for filenames
for each Database
Database
Database
Project DB, Collection DB
Export with DWB
Project DB Column Descriptions
SIARD_Model_<Database Name>_<UUID>.csv
Use SIARD GUI
Project DB SIARD Archive
DWB_<Database Name>_<Project Name>.siard
Generate PDF from Project DB Documentation
Project DB ER-Diagram and Table Descriptions
DWB_Model_<Database Name>_<UUID>.pdf
Generate Handbook as PDF
Manual
Handbuch_DWB_<Database Name>_<Version>.pdf
Generate PDF from General Database Description Wiki
General Database Description
Wiki_DWB_<Database Name>_<Version>.pdf


Project
Generate PDF from Project Description HTML
Project Description
Dokumentation_<Sequence Number>_Daten_<UUID>.pdf
Generate PDF from DWB Wiki
General DWB System Description
Wiki_DWB_<Version>.pdf
Create Metadata File
Metadata SIP
BAYHSTA_Erschliessung_<UUID>.csv
Create Zip-Archives
Include:
* SIARD_Model_<Database Name>_<UUID>.csv
* DWB_Model_<Database Name>_<UUID>.pdf
* Handbuch_DWB_<Database Name>_<Version>.pdf
* Wiki_DWB_<Database Name>_<Version>.pdf
* Dokumentation_<Sequence Number>_Daten_<UUID>.pdf
* Wiki_DWB_<Version>.pdf
Prolog Zip Archive
BAYHSTA_PROLOG_<UUID>.zip
Include:
* DWB_<Database Name>_<Project Name>.siard
Data Zip Archive
BAYHSTA_DATEN_<UUID>.zip
Include:
* BAYHSTA_Erschliessung_<UUID>.csv
* BAYHSTA_DATEN_<UUID>.zip
* BAYHSTA_PROLOG_<UUID>.zip
BAYHSTA_Erschliessung ZIP Archive
BAYHSTA_Erschliessung_<UUID>.zip

The Bavarian State Archives operate an OAIS-compliant digital archive. In cooperation with the respective state data producers, logical interfaces are designed and technically implemented that enable data archiving under these conditions for the heterogeneous original systems.There are now also some automated solutions for this, such as for e-files and specialist procedures. The Swiss governmental standard SIARD is also suitable for data from collection databases.

2 Operational databases of the Diversity Workbench and data management at SNSB (0.25 pages)

xxTextTextxx ..- documentations of operational database content and system design at SNSB (Wiki, Abouts, manual, ER diagrams …, XML archives, CSV, pdfs)

The modularized Diversity Workbench (= DWB) represents a virtual research environment for multiple scientific purposes with regard to management and analysis of life and environmental sciences data.

The databases at the SNSB are actively used, so the datasets are dynamic and data is changing constantly. (see also 3.1)

2.1 Description of the system design (0.5 pages)

It consists of 14 modules (abb.x) for different types of data and is a suite of independent, but linked software applications. These can be used individually or in combination with each other, concerning the own needs „linked data objects“
.Net Clients, webservices, Web-Apps
Optimised for different data domains

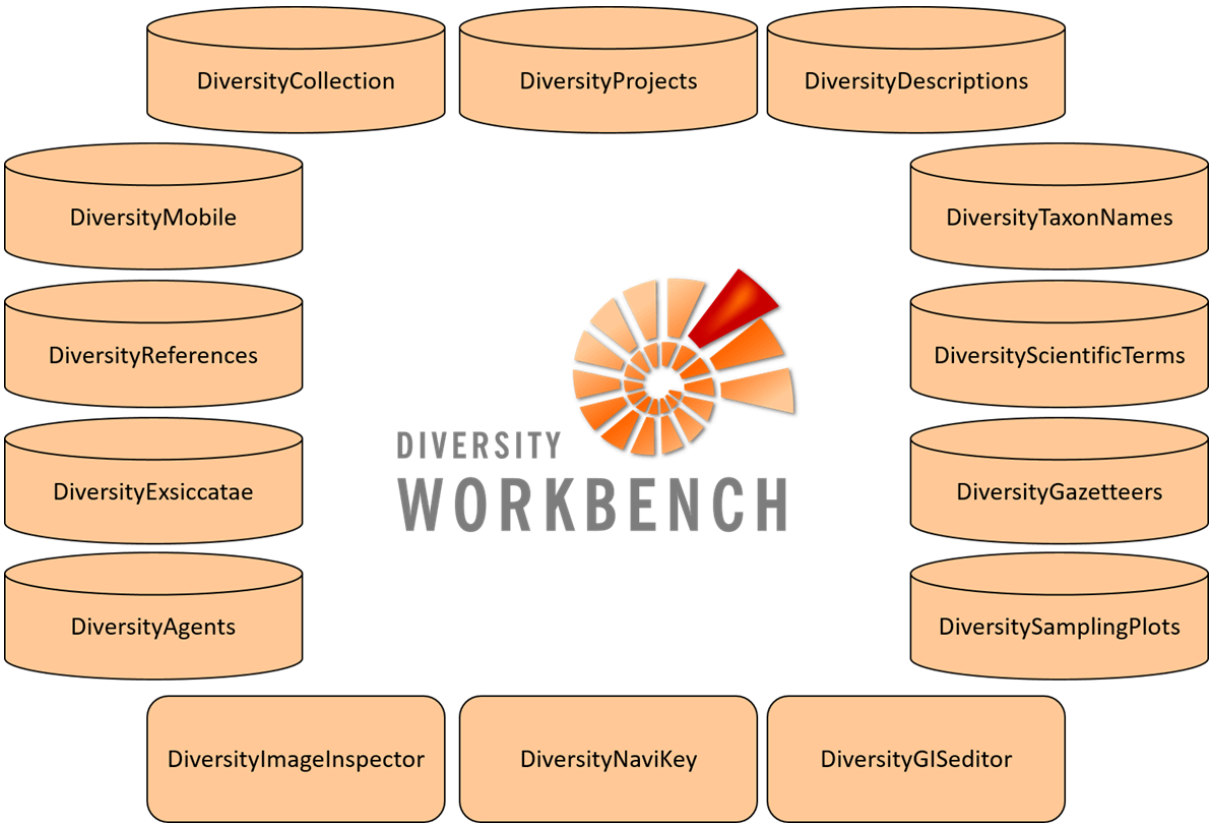


Abb.x: 14 modules of the DWB

DiversityProjects is the module for the management of metadata of research projects and data sets defined and used by modules of the Diversity Workbench. It is designed to document information about the data sets, their relations and hierarchy. All metadata on data set level are processed here, e.g. data set provenience, descriptions, responsible agents, identifiers including DOIs, etc. Each project can have different licenses on how the data can be used.

DiversityCollection is the module for the management of occurrence data, thus observation and collection data (GFBio Data Type 1a and b, https://kb.gfbio.org/display/KB/Major+Types+of+Biological+Data). To document data about collection events, collection objects/observations, experiments and analyses, object transfer/exchange, preparation and storage of specimens or multimedia files. Labels can be printed and distribution maps can be created.
It is suitable for Data from various disciplines like Botany, Mycology, Zoology, Palaeontology, Mineralogy, Anthropology, etc.

To avoid too complex database structures each module has its own relational database with an own database model to store the necessary information. This helps to keep the table relations clear and understandable. The linking between the modules is done by identifiers as linked data objects. These identifiers allow a coupling between items in the databases with link resolution on the client side. The business logic is represented in the table

relations, the table relations, the linking between the modules but relies also on informal agreements on how to store data in the database.

Case study: Mycological Data, mildews, Erysiphe

Noch auszuformulieren:

- ER diagrams
- conceptual and textual standards
- code and software application repositories
- manuals
- vocabularies of the DWB …
- Projects can have different licenses
- Dataset history explicitly stored (reproducible science) -> 3.2
- Pseudonymized editing history -> 3.2?
- Internal management details (transaction, loan,) -> 3.2?
- semantic links to other database modules (no foreign keys or direct db linking)
- business logic stored in table relations but also in informal agreements

2.2 Data pipelines for published and unpublished data (0.5 pages)

xxTextTextxx see DWB Specimen and Occurrence Data, PostgreSQL Cache, BPS API, BMS API, GBIF and NFDI - Serviceprofil; involving services at GDA, LRZ, SNSB and NFDI and documentations for data management; inventory data and research data
Data Flow image.

2.3 SNSB owned collection data and related research data (0.5 pages)

xxTextTextxx concept of extended digital specimens, data policies and core services

Beschreibung der Daten zu den physischen Objekten (Digital Specimen) Daten zu einem Physischen Sammlungsobjekt, Vortrag vom März mit Zitaten

Weibulat, Tanja, Schwartz, Anna Lisa, Seifert, Stefan, Schmalzl, Markus, Triebel, Dagmar (2023) Was haben Mehltaupilze mit digitaler Langzeitarchivierung zu tun? https://zenodo.org/records/8268840

Triebel/Puchta: Vortrag AG Long Term Archival (LTA) innerhalb der Sektion Infrastructure des NFDI e.V. am 17./18. August 2022 in der DNB Frankfurt

Ohnesorge, Vortrag "Archivierung von Daten aus relationalen DB im Schweizerischen BA", Tagung AK "Archivierung von Unterlagen aus digitalen Systemen" April 2009 im Staatsarchiv St. Gallen

3 Specification of the Case Study with SIARD (0.25 pages)

The prerequisite for long-term archiving of data at the Bavarian State Archives is that the data is worth of archiving within the meaning of the Bavarian Archives Act. Since the collection data of the State Natural Science Collections is permanently important scientific information, the lasting value is given primarily for the purposes of science but also for those of the collecting authority itself. In addition to the actual data on the collection objects, other internal data from the database is also archived, such as information about processing and those processing it and data provenance.
Three SNSB collections with a high degree of indexing were selected for the pilot project, which is also represented in the database: The Erysiphales Collection (BSMeryscoll), the Fosill Fish Collection (BSPGpiscescoll) and the Arthropoda Varia Collection (ZSMarthrovariacoll) [#13].
The special feature of the case study is the archiving of relational databases, that has not yet been established in the German archiving sector. SIARD provides a standardized, efficient and easy-to-use possibility to archive relational databases and was therefore used for this study. SIARD [XXX delcis] is a standard which describes how to store relational databases in an archivable content format using XML and the SQL2008 dialect. The Swiss Archive provides an open source toolset (SIARD suit, [XXX github]) to migrate a databases into the SIARD format. Besides the transformation of the database it also applies to the use of SIARD that the data to be archived must be prepared for long-term archiving or explained with metadata. This concerns, on the one hand, the resolution of the keys and codes used that is as human-readable as possible, as well as the explanation of the attributes and values used. In addition, the indexing metadata to prove the archived data in the archive information system must be created and supplied. Finally, the significant properties must be defined that should guarantee interpretability of the information over several migration cycles.

3.1 Valuable scientific content description of archived collection data (0.75 p) (SNSB)

Natural History Collections accommodate specimens of various materials like dried plant or mycological objects, (parts of) animals or plants in alcohol, microscopic slides, DNA samples, insects, bones, furs, fossils, etc. These objects are often digitized by taking pictures of the objects and typing all information given about the objects in a database. These data can be published in different national and international portals, thereby other researchers can re-use the data for their research projects. They can also use the published information to decide which Natural History Collection they want to visit for having a closer look at the specimens of which specimens they want to loan.
The case study was started with three different occurrence datasets - one mycological, one paleontological and one zoological. They are all managed in the same data management system, Diversity Workbench, and therefore have the same structure, including the same DWB modules used.
Archival value, scientific value

The mycological dataset is called ‘The Erysiphales Collection at the Botanische Staatssammlung München’ (http://www.botanischestaatssammlung.de/DatabaseClients/BSMeryscoll/About.cfm, https://doi.org/10.15468/sookye). It consists of data about 6.200 herbarium specimens from 600 species, each including mildew and the host plant from all over the world.Erysiphales are well known by gardeners with their characteristical flour-like cover of different ornamental shrubs like roses. The specimens are a valuable historical collection with the oldest specimens about 200 years old.

The paleontological dataset is called ‘The Fossil Fish Collection at the Bayerische Staatssammlung für Paläontologie und Geologie’ (http://www.snsb.info/DatabaseClients/BSPGpiscescoll/About.html). It consists of data about 2.159 fossil specimens from around 20 countries, partly including

The zoological dataset is called ‘The Arthropoda Varia Collection at the Zoologische Staatssammlung München’ (http://www.snsb.info/DatabaseClients/ZSMarthrovariacoll/About.html).

3.2 Challenges of RDBMS in operation (0.75 p) (SNSB,GDA)

Difference between the Database and the RDMBS: It would be possible to archive only the data without the management information behind. This is done already in the case of disseminated data packages which are provided for external data aggregation services. Here only the relevant Metadata to understand the current state of the data is exposed and for documentation archived. The data in a database management system is, especially in a

database system for scientific data management and curation, always under modifications. It is not always possible to reconstruct which modifications on the data was done and why these were done, from the current state of the published dataset. In the in-house management system Diversity Workbench (DWB)l modifications are logged automatically in the separate tables. Those modifications, like updates, inserts or new withholds are logged with timestamps and pseudonymized agent information in the log-tables. These logs allow that all modifications of the dataset can be reproduced to assure data quality and reproducibility. This shows that not all data from the DBMS is suitable to be published, but important to understand the genesis of the dataset. Further examples are legal documents for the specimen collection which are referenced in the database, als well as loans that are documented. This

- in-house management systems … legacy systems, migrations, business model, documentation, schema
-

3.3 SIARD installations at SNSB and GDA (0.75 p) (SNSB,GDA)

xxTextTextxx

- SIARD Standard and ISO Norm
    - ❖ proprietary databases into a non-proprietary (SIARD) format
    - ❖ developed at Swiss Federal Archive (mention one of their use cases?)
    - ❖ rare use in the german archiving sector to the present day
- SIARD format and tools
- SIARD in the case study
    - ❖ SIARD files, use of the tool
- SIARD and related management of multimedia
    - ➔ Approximate time estimate from the first installation steps to the final compilation of the submission package

3.4 specific requirements for the ingest and content preservation (0.75 p) (GDA)

The Bavarian State Archives operate the "Digital Archive" as a long-term digital repository in which data and supplementary information can be stored indefinitely. The aim is not only bitstream preservation but the preservation of the interpretability of the content of the data. Therefore the data must be prepared and enriched with metadata and documentation material and several specifications must be made. The ingest is being prepared by a multi-stage approach based on a recently published specialist concept [#2, 14−17].

The first step before transferring data is a rough evaluation. It is particularly important to determine whether the data can be relevant for understanding other data in addition to its lasting value. In the next step, further information is obtained from the data providing institution, for example on the structure of the original system (in this case Diversity Workbench) and its documentation. After examining existing export interfaces, which were not suitable for the case study as they were limited to archiving the content, SIARD was chosen as a new approach. SIARD is a widely used, internationally used and well-documented standard. The additional archiving of the database documentation,

including the data model and a detailed explanation of the attributes of the state natural science collections, also guarantees future interpretability. The DELCIS Board [XXX Declis page] recommends the creation of an SIP using the Content Information Type Specification for Relational Databases using SIARD (CITS SIARD, [XXX]) but in this project the Archive has its own recommendations for the SIP structures.
In addition to the digital archive, an archival subject information system (ActaProDesk) operates in which the metadata is stored. Currently operating export interfaces serve the EAD and ABCD XML format. The data records in both systems are uniquely identifiable via a UUID [#15, 269−70].
Following the OAIS reference model, the data is transferred as an submission information package that correspondence to the archival information package (AIP) in this case study [#2, 30−31].
The structure and content must be agreed with the submitting authority, as must the indexing metadata to be included and transferred to ActaProDesk. As these are not analog but digital files, they follow a separate indexing guideline [#14, 64−76].
The submission portion consists of a ZIP container, which in turn comprises several individual ZIP files. A quasi-standard from the Bavarian State Archives is used for the delivery structure of the SIP, which is also used for the XML client of the Bavarian State Archives: two ZIP packages contain the primary data and the associated metadata, which are necessary for indexing the submission. A third file contains the indexing metadata as a csv-file. The SIP and the indexing metadata can be clearly associated through a UUID. In order to ensure the long-term interpretability and following the OAIS reference model, the first ZIP folder contains documentation material. For each DBMS module, in this project DWB-Projects and DWB-Collections, this includes wiki pages and data models. The delivery structure and file name conventions developed in this project are presented in Table XXX.
The indexing metadata in csv format can be used for import into ActaProDesk.
The indexing metadata include information on how and over what period of time the data collection was created, the licensing of the metadata, its origin and restrictions in use, taking into account the requirements of the data-holding body and the Bavarian Archives Act [#5].
References to multimedia files stored on external servers outside the primary system are also listed with a contact address.
Once the files have been compiled, the transport method and frequency still need to be defined. As the project involves a database dumb and no continuous transfer via automated archival interfaces, determining a transfer cycle is also part of the long-term preservation strategy.
As soon as they have been ingested, interested users can research the archival collections in the reading room of the Bavarian State Archive. With the forthcoming launch of the Virtual Reading Room, this task be possible online.

4 Results and Perspectives (3 p) (GDA, SNSB)

In the cooperation project of the Bavarian State Natural Science Collections and the Bavarian State Archives, an efficient and resource-saving generic solution was found as to how object-related collection data of lasting value can be transferred to a digital long-term archive. The use of the Swiss government standard SIARD guarantees the traceability of the

database structure and a reduced effort when exporting the data. By adopting sufficient documentation material about the original system and explaining attributes and values, the long-term interpretability of the information is guaranteed. SIARD initially developed for archiving at database retirement [xxx lindley] is in the used version 2.2 [XXX Delcis Siard] an efficient solution for time slice archiving of a running database system. It converts a relational database into an fileformat which is agnostic from the archival system. The indexing metadata coordinated in the project allows for rapid processing and guarantees data quality and findability. Just like the transfer structure and file naming convention, these specifications allow transfer to other object-oriented collection databases.
Some questions still remain unanswered, including how defined documents in the form of photographs and other files that are stored in external storage can be integrated into the transfer. Some of the provenance-related significant properties for future data migration and inventory preservation also still need to be defined. Nevertheless, one can speak of a success.
This also means an important milestone for the use case of the NFDI4Biodiversity with which the long-term archiving process for scientific collections is to be tested. Based on this successful example, an important building block of sustainable data management can be built for the entire consortium at a crucial point in the data life cycle in the long term.
In addition, the generic and easily transferable solution approach also makes a significant contribution to the topic of long-term archiving of research data for the National Research Data Infrastructure, in which the sustainable storage of relevant data sets has still largely been ignored. Finally, the SIARD archiving interface that has now been implemented also represents an important step for the archive community in Germany. After all, it is one of the first SIARD-based database archiving in public archives in Germany.
Based on the best practices now in place, further collection databases will be archived for the long term. The aim is also to preserve collection databases from other state collections and museums in Bavaria at the Bavarian State Archives in the long term for the future.

xxTextTextxx
Generic approach? Relation to RDC? Proxy and mediation layer, see https://kb.gfbio.org/pages/viewpage.action?pageId=113904360 ?

SIARD-Standard suitable and efficient for archiving of collection data from scientific data bases; One of the first uses of the Swiss standard SIARD in Germany for the long-term archiving of data

multimedia objects and big data? LRZ involved

distributed archives …and linked data?

OAIS compliant archival access via NFDI? Metadata in the DIP?

role of long-term state archives in NFDI?

- consulting services for long-term archiving
- long-term archiving und fairification of relevant data of governmental institutions

blueprint for archival services at the collection data centers at the GFBio e.V. ?

involvement of archives in German states? digital documents …structured digital data with documents

“Documents that are of archival value are documents of **legal** or **administrative importance** or which contain valuable information.”

xxTextTextxx

- *NHC-Objekte sind Referenzobjekte zur Artenschutz-Gesetzgebung —> Daten (= Inventar- und Forschungsdaten) dazu sind archivwürdig*
- *NHC-Objekte unterliegen Gesetz zum Schutz von Kulturgütern —> Daten (= Inventar- und Forschungsdaten) dazu sind archivwürdig*
- *NHC-Objekte sind Referenzobjekte der Forschung für taxonomische Namen von Organismen —> Daten (= Inventar- und Forschungsdaten) dazu sind archivwürdig*

Bundesnaturschutzgesetz (BNatSchG)
Schutz der Kulturgüter ”Blue Shield”, siehe
https://www.bbk.bund.de/DE/Themen/Schutz-Kulturgut/Was-ist-Kulturgut/Identifizierung-Kulturgut/identifizierung-kulturgut_node.html

SIARD has proven to be an efficient standard for the long-term archiving of collection data. Based on the best practices now in place, further collection databases will be archived for the long term. The aim is also to preserve collection databases from other state collections and museums in Bavaria at the Bavarian State Archives in the long term for the future.

5 Acknowledgement

xxTextTextxx
This work was supported by the German Research Foundation (DFG) within the project ‘Establishment of the National Research Data Infrastructure (NFDI)’ in the consortium NFDI4Biodiversity (project number 442032008). Further funding was provided by COST Action CA17106 (https://www.cost.eu/actions/CA17106/).

6 References

https://link.springer.com/journal/13222/submission-guidelines#Instructions%20for%20Authors_References

#1
Grau, B, Schmalzl, M, Unger, M (2023) Die Beteiligung der Staatlichen Archive in der NFDI: Ein Werkstattbericht. Archiv. Theorie & Praxis 76,1: 40−42. https://doi.org/10.5281/zenodo.8131770

#2
Puchta, M, Ksoll-Marcon, M, Grau, B, Kirstein, M, Nestl, A, Schmalzl, M, Unger, M (2023) Fachkonzept für das Digitale Archiv der Staatlichen Archive Bayerns. Version 2.1 (29 August 2022). Generaldirektion der Staatlichen Archive Bayerns. https://doi.org/10.5281/zenodo.7743888. Accessed 12 December 2023

#3
The Consultative Committee for Space Data Systems (2012) Reference Model for an Open Archival Information System (OAIS). Magenta Book https://public.ccsds.org/Pubs/650x0m2.pdf. Accessed 12 December 2023

#4
Oßwald, A, Brübach, N (2010) Das Referenzmodell OAIS – Open Archival Information System. In: Neuroth, H, Oßwald, A, Scheffel, R, Strathmann, S, Huth, K (ed) nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, version 2.3. Nestor, Göttingen. https://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf. Accessed 18 December 2023

#5
Bayerisches Archivgesetz (BayArchivG) (1989) https://www.gesetze-bayern.de/Content/Document/BayArchivG/true. Accessed 18 December 2023

#6
Borgman CL (2019) The Lives and After Lives of Data. Harvard Data Science Review 1(1). https://doi.org/10.1162/99608f92.9a36bdb6

#7
Bierwirth M, Glöckner FO, Grimm C, Schimmler S, Boehm F, Busse C, Degkwitz A, Koepler O, Neuroth H (2020) Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung. https://doi.org/10.5281/zenodo.3895209 Accessed 7 January 2024

#8
Funk SE, Ullrich D (2010) Digitale Erhaltungsstrategien In: Neuroth, H, Oßwald, A, Scheffel, R, Strathmann, S, Huth, K (ed) nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, version 2.3. Nestor, Göttingen. https://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf Accessed 7 January 2024

#9

Triebel D, Hagedorn G, Rambold G (2012) An appraisal of megascience platforms for biodiversity information. MycoKeys 5: 4563. https://doi.org/10.3897/mycokeys.5.4302 Accessed 14 January 2024.

#10
Triebel D (2009) Pilzherbarien – Neue Aufgaben im Bereich Biodiversitätsinformatik und Datenmanagement. In: Bayerische Akademie der Wissenschaften (ed), Rundgespräche der Kommission für Ökologie, vol. 37 (Ökologische Rolle von Pilzen). Pfeil, München, 131–145.

#11
Diepenbroek M, Glöckner FO, Grobe P, Güntsch A, Huber R, König-Ries B, Kostadinov I, Nieschulze J, Seeger B, Tolksdorf R, Triebel D (2014) Towards an Integrated Biodiversity and Ecological Research Data Management and Archiving Platform: The German Federation for the Curation of Biological Data (GFBio). In: Plödereder E, Grunske L, Schneider E, Ull D (eds), Informatik 2014 – Big Data Komplexität meistern. GI-Edition: Lecture Notes in Informatics (LNI) – Proceedings 232. Köllen Verlag, Bonn, 1711–1724. https://cs.emis.de/LNI/Proceedings/Proceedings232/1711.pdf. Accessed 14 January 2024.

#12

Glöckner FO, Diepenbroek M, Felden J, Güntsch A, Stoye J, Overmann J, Wimmers K, Kostadinov I, Yahyapour R, Müller W, Scholz U, Triebel D, Frenzel M, Gemeinholzer B, Goesmann A, König-Ries B, Bonn A, Seeger B (2020) NFDI4BioDiversity − A Consortium for the National Research Data Infrastructure (NFDI). Zenodo, 1−128. https://doi.org/10.5281/zenodo.3943645. Accessed 14 January 2024.

#13
(2001–2024) The Erysiphales Collection at the Botanische Staatssammlung München. https://www.botanischestaatssammlung.de/DatabaseClients/BSMeryscoll/About.cfm. The Fossil Fish Collection at the Bayerische Staatssammlung für Paläontologie und Geologie. https://www.snsb.info/DatabaseClients/BSPGpiscescoll/About.html. The Arthropoda Varia Collection at the Zoologische Staatssammlung München. https://www.snsb.info/DatabaseClients/ZSMarthrovariacoll/About.html. All accessed 14 January 2024

#14
Bachmann C, Herget R, Stehr M, Unger M, Puchta M, Schmalzl M (2023) Richtlinien zur Verzeichnung von Archivgut der Staatlichen Archive Bayerns. Generaldirektion der Staatlichen Archive Bayerns, München. https://www.gda.bayern.de/fileadmin/user_upload/Medien_fuer_Unterseiten/Verzeichnungsrichtlinien_Stand_10-2023.pdf. Accessed 14 January 2024.

#15
Hartman A, Herold H, Schwartz AL, Schmalzl M (2023) FAIRe Geodaten aus dem Archiv. Ein Tool zur automatisierten Metadatenextraktion aus historischen Karten. ABI Technik 43(4), 260–271. https://doi.org/10.1515/abitech-2023-0045. Accessed 14 January 2024.

—---------- (darunter noch nicht im Text verwendete references)
xxTextTextxx

SIARD-Formatspezifikation eCH-0165 (2013) eCH Verein.
http://www.ech.ch/de/ech/ech-0165/1.0

Weber, A, Piesche, C (2021) Datenspeicherung, -kuratierung und Langzeitverfügbarkeit. In: Putnings, M, Neuroth, H, Neumann, J (ed) Praxishandbuch Forschungsdatenmanagement. De Gruyter Saur, Munich, pp 327-355.
https://doi.org/10.1515/9783110657807-019

Ohnesorge, KW (2011) Aufbau und Entwicklung der digitalen Archivierung im Schweizerischen Bundesarchiv. Archivalische Zeitschrift 92,1, xx-xx. https://doi.org/10.7788/az.2011.92.1.117

http://t.co/yZFvj4xxQv SIARD vs. Chronos
https://www.dpconline.org/docs/miscellaneous/events/2016-events/1488-transdata-presentation-lindley/file Siard vs. Chronos Presentation

Andrew Lindley (2013) Database Preservation Evaluation Report - SIARD vs. CHRONOS - Preserving complex structures as databases through a record centric approach? In Proc. iPRES 2013 – 10th International Conference on Preservation of Digital Objects. ISBN 978-972-565-493-4 Biblioteca Nacional de Portugal, Lisbon, pp. 29 - 38.

Preserving Databases (2021) Data Types Series, Artefactual Systems and the Digital Preservation Coalition, http://doi.org/10.7207/twgn21-06

Henry M. Gladney (2009) Long-Term Preservation of Digital Records: Trustworthy Digital Objects. The American Archivist 72 (2): 401–435. https://doi.org/10.17723/aarc.72.2.g513766100731832

A Study on SIARD Verification as a Preservation Format for Data Set Records /행정정보 데이터세트 보존포맷으로서 SIARD 검증에 관한 연구 (evtl. für weitere Literatur)

Julia Glöde 2009 DIPLOMARBEIT. Archivierung relationaler Datenbanken auf der Grundlage von XML Konzeption eines OAIS-konformen Archivierungsmodells und Entwicklung eines neuen Ansatzes zur Archivierung von Datenbankabfragen
Studiengang: Dokumentation
Matrikelnummer: 6763
E-Mail: julia.gloede@fh-potsdam.de

Dokument von NFDI4Culture (viele Literaturverweise)

https://docs.google.com/document/d/1GQ-Tdzx3V2idJ1ZHXG5fN37eWIk1cnE1TFxYvD8gT_g/edit (intern an NFDI Section Infra, AG LTA wegen Input versandt)

Open Preservation Foundation. (2022, Dezember). *International Comparison of Recommended File Formats*. Google Docs. https://docs.google.com/spreadsheets/d/1XjEjFBCGF3N1spNZc1y0DG8_Uyw18uG2j8V2bsQdYjk/edit

wegen SIARD format

https://www.dpconline.org/docs/technology-watch-reports/2481-preserving-gis/file wegen Preserving GIS

Comment, Jean-Marc. 'The Digital Repository of the Swiss Federal Archives: a Case Study'. 1 Jan. 2011 : 63 – 68. DOI: 10.3233/WDL-120075 (noch kein Zugang)

Aufbau und Entwicklung der digitalen Archivierung im Schweizerischen Bundesarchiv https://www.vr-elibrary.de/doi/epdf/10.7788/az.2011.92.1.117

Vorlage: Paper

Triebel, D., Grunz, A., Seifert, S., Link, A. & Rambold, G., 2021. DiversityNaviKey, a Progressive Web Application for interactive diagnosis and identification. In: , . (Hrsg.), INFORMATIK 2021. Gesellschaft für Informatik, Bonn. (S. 517–538). DOI: 10.18420/informatik2021-040

https://dl.gi.de/server/api/core/bitstreams/f97e624d-36a1-40d7-9bde-f1d58f850644/content

.

Heuscher, S., Jaermann, S., Keller-Marxer, P., & Moehle, F. (2004, October 5). Providing Authentic Long-term Archival Access to Complex Relational Data. European Space Agency. Symposium "Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data", Frascati. https://doi.org/10.48550/arXiv.cs/0408054

Naumann, K. Databases for 2080. Workshop proceedings https://d-nb.info/1265590877/34

zu zitieren:
Heseler, J., Büttner, A., & Arnold, M. (2023). Grundlagen der digitalen Langzeitarchivierung. Eine Handreichung zur digitalen Langzeitarchivierung aus Perspektive der NFDI4Culture Community. (1.0.1). NFDI4Culture. https://doi.org/10.5281/zenodo.10423858

Alex R Hardisty, Elizabeth R Ellwood, Gil Nelson, Breda Zimkus, Jutta Buschbom, Wouter Addink, Richard K Rabeler, John Bates, Andrew Bentley, José A B Fortes, Sara Hansen, James A Macklin, Austin R Mast, Joseph T Miller, Anna K Monfils, Deborah L Paul, Elycia Wallis, Michael Webster, Digital Extended Specimens: Enabling an Extensible Network of

Biodiversity Data Records as Integrated Digital Objects on the Internet, *BioScience*, Volume 72, Issue 10, October 2022, Pages 978–987, https://doi.org/10.1093/biosci/biac060

Triebel, D., Ivanovic, D., Kahila Bar-Gal, G., Bingert, S. & Weibulat, T. 2021. Towards a COST MOBILISE Guideline for Long Term Preservation and Archiving of Data Constructs from Scientific Collections Facilities. Biodiversity Information Science and Standards 5: e73901. https://doi.org/10.3897/biss.5.73901

Harjes et al***

und aus dem **Internet Archive**:

GFBio Wiki - Category:Data Archiving
https://web.archive.org/web/20220704162136/http://gfbio.biowikifarm.net/wiki/Category:Data_Archiving (last archived by Internet Archive at July 04, 2022)

mit

GFBio Wiki - Technical documentation of long-term archiving solutions at the GFBio collection data centers
https://web.archive.org/web/20220704162053/https://gfbio.biowikifarm.net/wiki/Technical_documentation_of_long-term_archiving_solutions_at_the_GFBio_collection_data_centers (last archived by Internet Archive at July 04, 2022)

und

GFBio Wiki - Publication of Type 1 Data via BioCASe Data Pipelines at SNSB Data Center
https://web.archive.org/web/20220925110353/https://gfbio.biowikifarm.net/wiki/Publication_of_Type_1_Data_via_BioCASe_Data_Pipelines_at_SNSB_Data_Center (last archived by Internet Archive at Sept 25, 2022)

und

GFBio Wiki - ISO Standards for Digital Archives
https://web.archive.org/web/20230625050849/https://gfbio.biowikifarm.net/wiki/ISO_Standards_for_Digital_Archives (last archived by Internet Archive at June 25, 2023)

und

https://web.archive.org/web/20230318102739/https://gfbio.biowikifarm.net/wiki/Data_exchange_standards,_protocols_and_formats_relevant_for_the_collection_data_domain_within_the_GFBio_network (last archived at March 18, 2023)

Zu zitieren NESTOR: 2019

http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_399.pdf

https://nbn-resolving.org/urn/resolver.pl?urn=urn:nbn:de:0008-2010030597

zu zitieren SIARD in Bayern, 2023: https://zenodo.org/records/10021391