



Long-Term Preservation and archiving in NFDI consortia

Long Term Data Issues: Workshop about Preservation, Archiving and
Access in the NFDI

06.11.2024

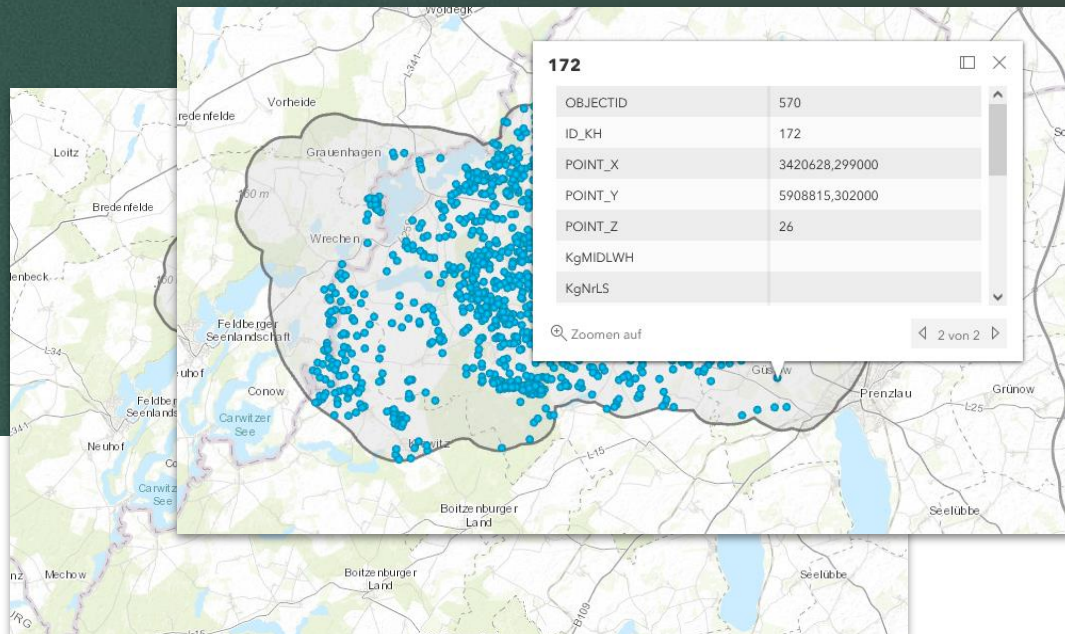


This work is licensed under a [Creative Commons Attribution-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nd/4.0/).

The BonaRes Repository

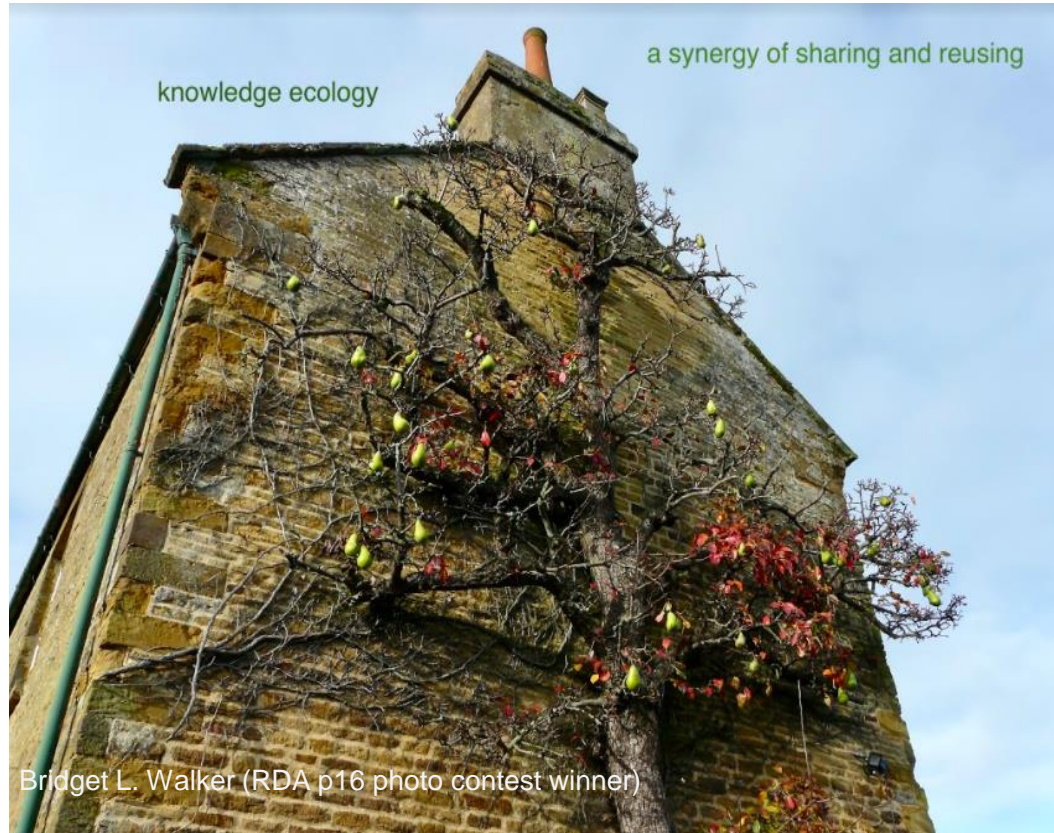
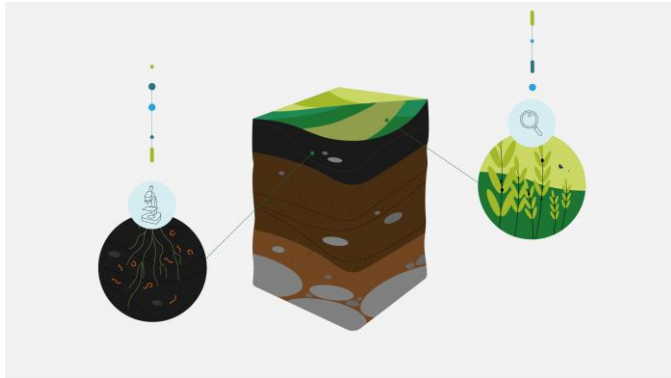
FAIR and sustainable data publication

Nikolai Svoboda
and the ZALF RDM team



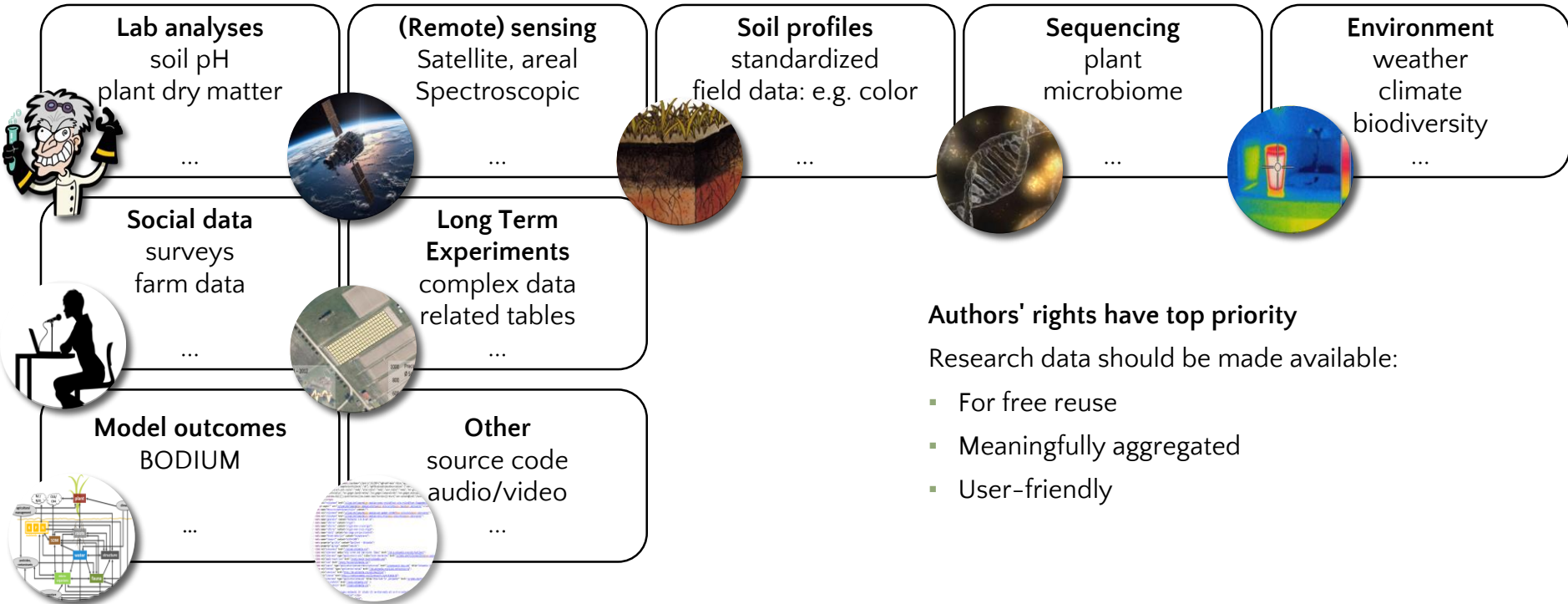
Date: 06.11.2024

- Managing diverse agricultural data
- BonaRes Repository for publishing
- Archiving published data
- CTS



We provide the infrastructure and services to support you in your scientific work so that you can harvest the fruits.

The challenge: Management of diverse agricultural data



Authors' rights have top priority

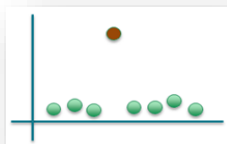
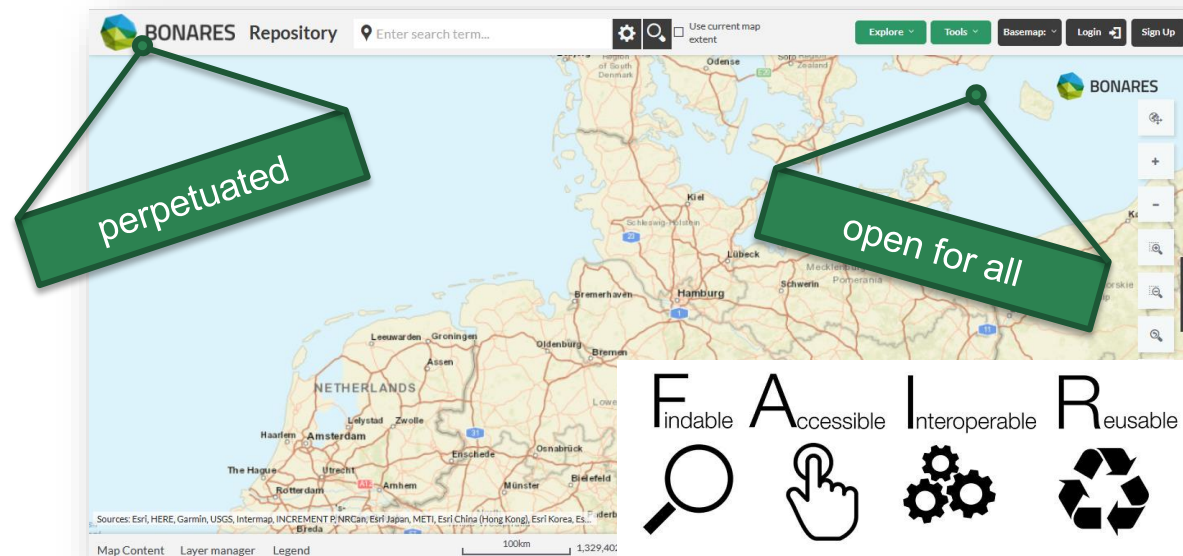
Research data should be made available:

- For free reuse
- Meaningfully aggregated
- User-friendly

The BonaRes Repository a networked geodata infrastructure



- Publication of research data
- Soil and agricultural sciences
- Description with metadata
- Quality-checked & legally sound
- Citable
- Interoperable
- Part of the NFDI - FAIRragro



By SangyaPundir - Own work, CC BY-SA 4.0

BonaRes Repository – tools

Use tools – save nerves



Long term archiving

clock image by Rantrum Diva: CC BY-NC 4.0

Upload tool

BonaRes Repository

LTE Map

DQ Kit

Bodenprofile

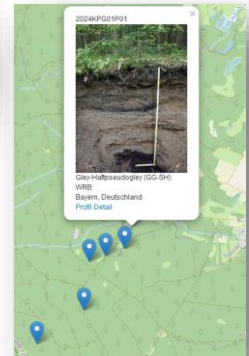
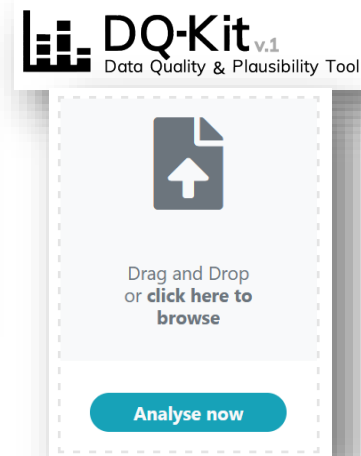
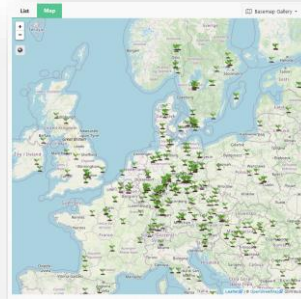
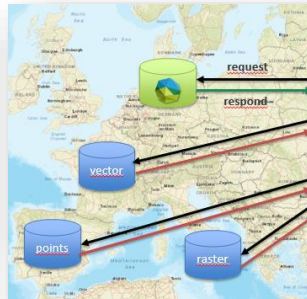
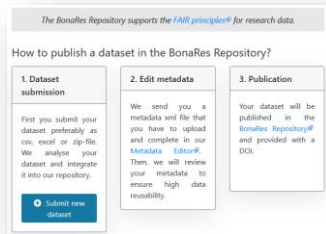
<https://upload.bonares.de>

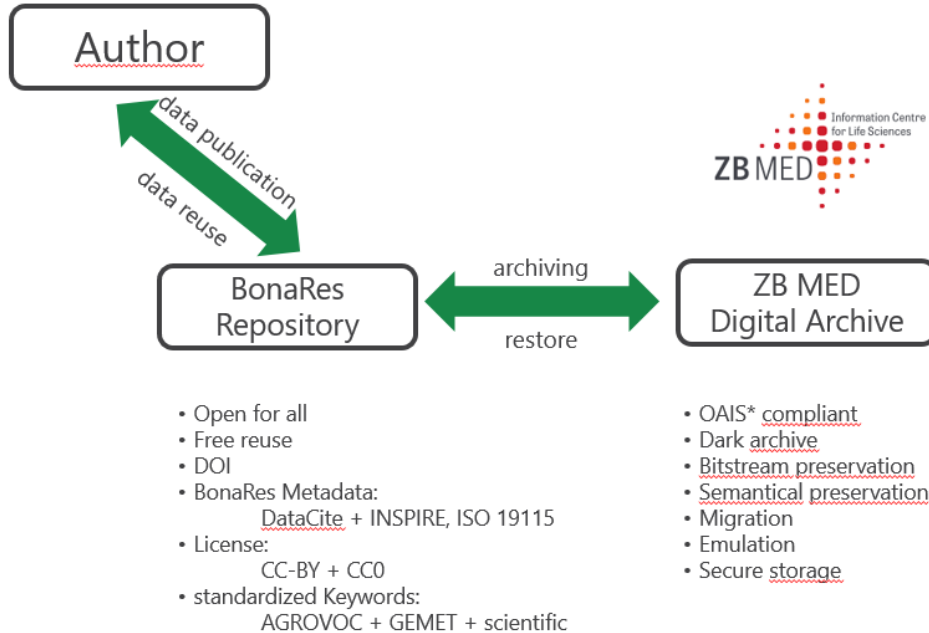
<https://lte.bonares.de/>

https://tools.bonares.de/bp_db/

<https://maps.bonares.de/mapapps>

https://dqkit.bonares.de/dq_kit/





Status

- Data (tabular) and metadata to archive
- Restoring from the archive

General workflow of archiving

*OAIS: Open Archival Information System

BonaRes

- DOI Vergabe: trigger
- Zusammenstellen von Informationen
 - Veröffentlichte Daten mit DOI
 - Metadaten
 - Wiederherstellungsinformationen
- Metadata mapping
 - BonaRes → Dublin Core
- Übergabe an die Archivinfrastruktur: Rosetta (ExLibris)

Rosetta

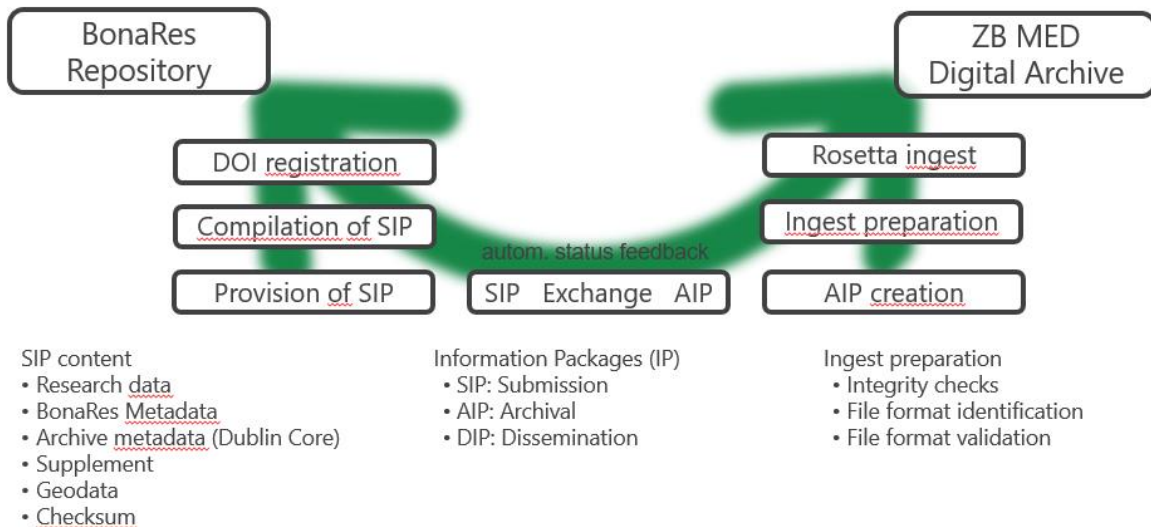
- Erzeugt und speichert Archivmetadaten (PREMIS standard)
- Strukturiert Metadaten Typen (METS standard)

BonaRes

- Erfolgreiche OAIS-konforme digitale Langzeitarchivierung von Argrardaten → BonaRes Ticketsystem
- Updates von Daten und Metadaten analog

Wiederherstellung

- Metadaten, Daten, Supplement



Technical implementation: there and back gain
clock image by Rantrum Diva: CC BY-NC 4.0

Why?

- Building trust in the community
- Basis for European level approval (HORIZON)

Who?

- Non-profit foundation under Dutch law

For whom?

- Repositories taking the responsibility for the curation and the long-term preservation of the usability into account
- Repositories built up on the requirements and knowledge base of the designated community

How?

- Self-assessment based on a catalog of requirements with 16 criteria
- Review of the self-assessment
- Final discussion on the CoreTrustSeal Board

Conditions

- Seal is limited to three years, thereafter renewal
- 3000 €



BonaRes Repository

- Under Review

Thank you for your attention.



Leibniz Centre for
Agricultural Landscape Research
(ZALF)

Contact: Nikolai Svoboda (svoboda@zalf.de); Xenia Specka (specka@zalf.de)

Long-term Archival of Biodiversity Data in



NFDI 4
BIODIVERSITY



Botanischer Garten
Berlin

Center for Biodiversity Informatics and Collection Data Integration

Botanic Garden Berlin


Jörg Holetschek

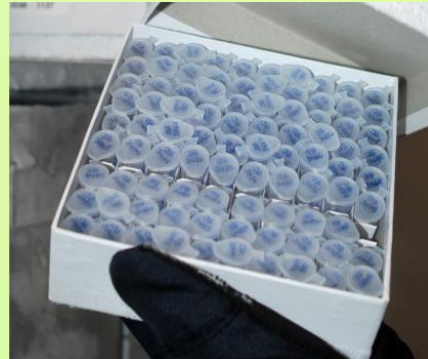


NFDI 4
BIODIVERSITY

Biodiversity Data

www.nfdi4biodiversity.org

 /NFDI4Biodiv
#NFDI4Biodiv

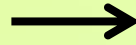


Complexity and Richness

Richness

- Very diverse collections
- ABCD data standard: 988 elements
- EFG extension: 1.828 elements

Complex relationships between objects to be captured





10 Data Centers

3 on nucleotide, plant & environmental data




7 for Natural History Collections (specimens, tissues, culture collections, DNA samples)





NFDI 4
BIODIVERSITY

www.nfdi4biodiversity.org

 /NFDI4Biodiv
#NFDI4Biodiv

Data Submission Service

<https://submissions.gfbio.org/>



Data Submissions

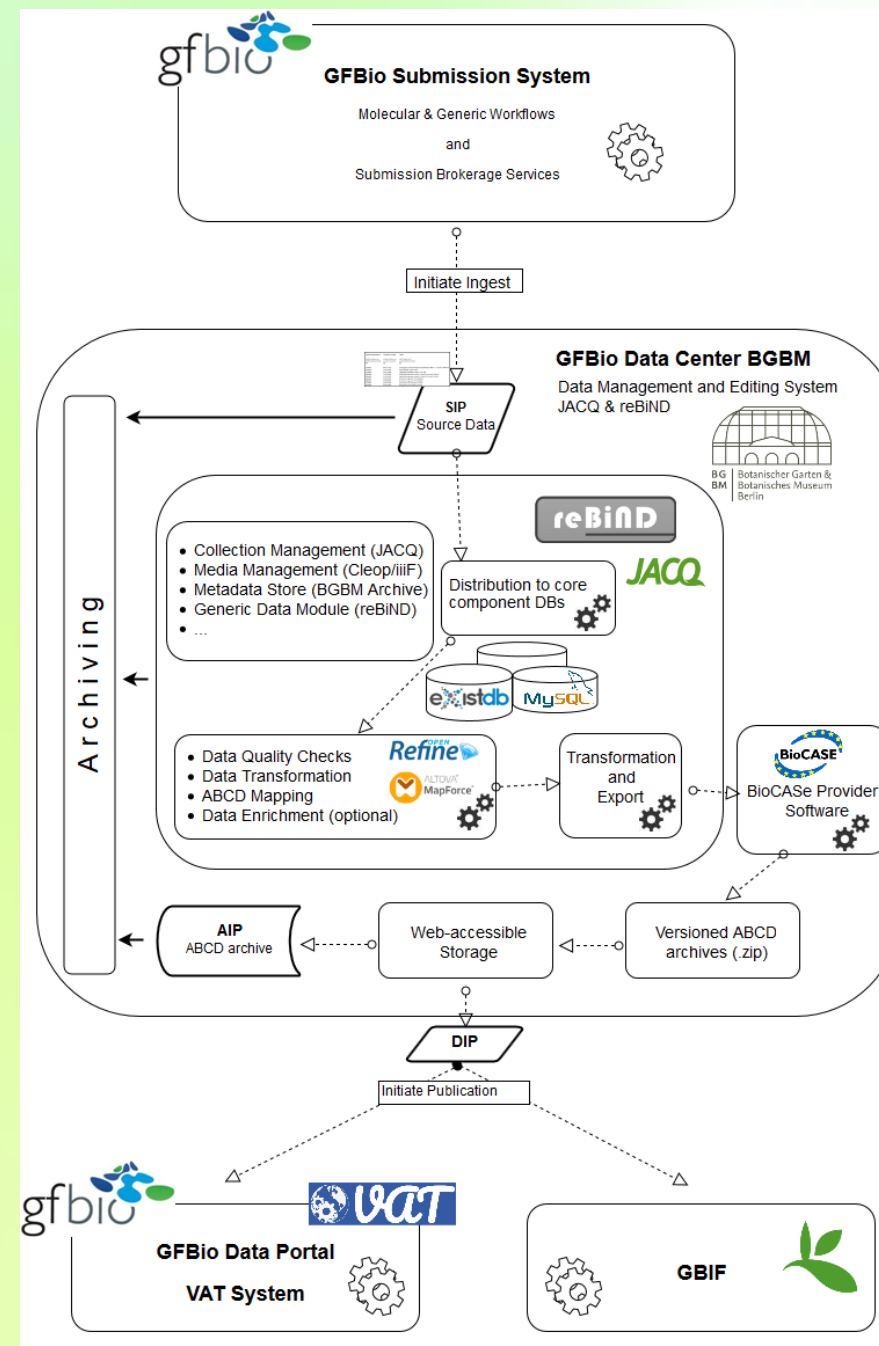
Long-term data archival & publication services for
Biodiversity, Ecology & Environmental Science

SUBMIT YOUR DATA !

Already using our service ? [Sign In](#)

Workflow at BGBM

1. Validation, Correction and Enhancement of data
2. Conversion into community data standards
3. Publication to community data portals
4. LTA of all data packages in accordance with OAIS (Open Archival Information System)





CnFdI

nationale
Forschungsdaten
Infrastruktur
for CULTURE

NFDI LTA Workshop | 2024-11-06

NFDI4Culture

Konsortium für Forschungsdaten zu materiellen
und immateriellen Kulturgütern

Matthias Arnold
UB Heidelberg

Gefördert durch die
Deutsche Forschungsgemeinschaft (DFG) - 441958017

DFG Deutsche
Forschungsgemeinschaft

Guideline



- Basics – What is digital preservation, FAIR-CARE-TRUST, Communities, and What archives relate to NFDI4Culture
 - How research data gets into the archive
 - What happens „inside“ the archive
 - Preservation strategies
-
- open community review phase incl. forum participants
 - expert reviews (esp. SLUB, TIB, HD, legal experts)


Grundlagen der Digitalen Langzeitarchivierung. Eine Handreichung zur Digitalen Langzeitarchivierung aus Perspektive der NFDI4Culture Community. (2024 v.1.03)

<https://zenodo.org/doi/10.5281/zenodo.10281971>.

Digital preservation in NFDI4Culture context

- **Subject specific archives**
(related to Specialised Information Services - FID's)

	SLUBArchiv.digital
URL	https://slubarchiv.slub-dresden.de/
NFDI4Culture ID	https://nfdi4culture.de/id/E3090
Anbieter	SLUB Dresden
Produktiv seit	2015
Zweck	Langzeitarchivierung
Schwerpunkt	Architekturgeschichte, Kunstgeschichte und Musikwissenschaft; Voraussetzung ist eine individuelle fachliche Bewertung durch die SLUB in Bezug auf die Passfähigkeit zu deren Sammlungsschwerpunkten
Zielgruppe	Öffentliche Kultur- und Wissenschaftseinrichtungen
Kosten	kostenpflichtig, volumenbasiert (Preise auf Anfrage ¹⁰¹)
Haltefrist	unbegrenzt
Dateiformate	nur nach Archivvorgaben ¹⁰²
Erhaltungsstrategien	Content Preservation (Formatmigration)
Zugriff	geschlossen (Dark Archive)

	 heiARCHIVE
URL	https://heiarchive.uni-heidelberg.de/
NFDI4Culture ID	https://nfdi4culture.de/id/E3788
Anbieter	Universität Heidelberg, Kompetenzzentrum Forschungsdaten (KFD) ¹⁰⁴ , eine gemeinsame Serviceeinrichtung des Universitätsrechenzentrums ¹⁰⁵ und der Universitätsbibliothek Heidelberg ¹⁰⁶
Produktiv seit	28.04.2023 (Pilotbetrieb)
Zweck	Langzeitarchivierung
Schwerpunkt	fachunabhängiges Archiv für Forschungsdaten von Mitgliedern der Universität Heidelberg; Kunstgeschichte
Zielgruppe	Mitglieder der Universität Heidelberg; Kunsthistoriker:innen weltweit in Verbindung mit einer Open-Access-Publikation in arthistoricum.net
Kosten	Langzeitarchivierung in Verbindung mit Online-Publikationen auf arthistoricum.net im Rahmen des Fachinformationsdienst Kunst kostenfrei, ggf. anfallende Publikationsgebühren
Haltefrist	Archivierung von Publikationen und kulturellem Erbe: unbegrenzt, sonst: in Stufen, mind. 10 Jahre
Dateiformate	Bitstream Preservation: keine Beschränkung, Content Preservation: TXT, PDF, XML und TIFF (Stand: 08/2023)
Erhaltungsstrategien	Bitstream Preservation und Content Preservation (Formatmigration)
Zugriff	geschlossen (Dark Archive)
Zertifizierung	in Planung
Weiterführende Lite-	Universitätsbibliothek Heidelberg (2023): heiARCHIVE - der


Digital preservation in NFDI4Culture context

- Subject specific archives (related to Specialised Information Services - FID's)
- **Archives handling special formats (3D, A/V)**

	 TIB PRESERVATION-AS-A-SERVICE
URL	https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung
NFDI4Culture ID	https://nfdi4culture.de/id/E2939
Anbieter	TIB Hannover
Produktiv seit	2013
Zweck	Langzeitarchivierung
Schwerpunkt	Dokumente, Retro-Digitalisate, A/V-Daten, 3D-Daten und Architektur
Zielgruppe	Wissenschaftler:innen, Lehrende, Studierende an Universitäten, Hochschulen und außeruniversitären Forschungseinrichtungen, Mitarbeitende in der außerakademischen und kommerziellen Forschung, Bibliotheken, Archive, Museen und Infrastruktureinrichtungen, Projektträger wie die Deutsche Forschungsgemeinschaft (DFG) oder das Bundesministerium für Bildung und Forschung (BMBF)
Kosten	kostenpflichtig (Preise auf Anfrage ¹⁰³)
Haltefrist	unbegrenzt
Dateiformate	keine Beschränkung
Erhaltungsstrategien	Content Preservation (Formatmigration) und Bitstream Preservation
Zugriff	geschlossen (Dark Archive)
Zertifizierung	Data Seal of Approval (2015); CoreTrustSeal (2020); nestor-Siegel (2017, 2022)

Digital preservation in NFDI4Culture context

- Subject specific archives (related to Specialised Information Services - FID's)
- Archives handling special formats (3D, A/V)
- **Generic archive** with focus on NFDI4Culture community

	
URL	https://radar4culture.radar-service.eu/
NFDI4Culture ID	https://nfdi4culture.de/id/E2853
Anbieter	FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur
Produktiv seit	2022
Zweck	Publizieren mit Langzeitarchivierung
Schwerpunkt	generisch (NFDI4Culture-Disziplinen)
Zielgruppe	Forschende an öffentlich geförderten Forschungseinrichtungen und (Kunst-)Hochschulen sowie nicht-kommerziellen Akademien, Galerien, Bibliotheken, Archiven und Museen in Deutschland
Kosten	kostenfreie Publikationen bis 10 GB (Preise für größere Volumina ¹⁰⁷)
Haltefrist	unbegrenzt (mind. 25 Jahre)
Dateiformate	keine Beschränkung
Erhaltungsstrategien	dreifach redundante Bitstream Preservation mit kontextuellen Metadaten
Zugriff	öffentlich (Datensätze sind jeweils per DOI identifizierbar)
Zertifizierung	CoreTrustSeal-Zertifizierungen sind bisher für generische Repositorien ohne fachliche Kuratierung noch nicht möglich
Weiterführende Literatur	Soltau, K., & Goeller, S. (2023). <i>RADAR4Culture: Quickstart-Guide für Datengeberinnen und Datengeber</i> . https://doi.org/10.5281/ZENODO.8221341

Digital preservation in NFDI4Culture context

- Subject specific archives (related to Specialised Information Services - FID's)
- Archives handling special formats (3D, A/V)
- Generic archive with focus on NFDI4Culture community

	SLUBArchiv.digital
URL	https://slubarchiv.slub-dresden.de/

	 heiARCHIVE
URL	https://heiarchive.uni-heidelberg.de/

	 TIB PRESERVATION-AS-A-SERVICE
URL	https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung

	 RADAR <i>for CULTURE</i>
URL	https://radar4culture.radar-service.eu/

Some things we've learned

- Digital preservation **terminology** is still evolving
 - We established a working terminology for our guideline (mapped to GND) and added missing entries in Wikidata
 - We compiled a list of synonyms (in German) together with their English equivalent(s) and referenced the sources of the terms
 - We compiled a list of glossaries covering digital preservation
- There are many **ways** into the archive
 - We provide one simplified example, but do not imply a generic sequence
- Digital preservation archives use different **approaches**
 - Preservation strategies matter most: Bitstream – Content (logical/semantic)

Some things we've learned

- Special cases
 - Web archiving (separate chapter)
 - Social media preservation (evolving, not in-depth)
 - Software preservation (not yet covered)
 - Dynamic content (not yet solved)
- Remaining questions
 - Good practices for migration of data from the repo to the archive?
 - How to find items in dark archives?
 - How to know if data is (also) in an archive (e.g. MARC21 #583)?

Plans and Status

- What individuals can do
 - Expand the guideline checklists for researchers
- Certification
 - SLUB archive (done)
 - TIB archive (done)
 - heiARCHIVE (in progress)
- Federated services
 - for interoperable publication and archiving infrastructure (in planning)
- Internationalisation (I18n)
 - English translation of the guideline (in preparation)
- Digital preservation workflows
 - (Publications with) 3D objects
 - „Enhanced publications“
 - Business models



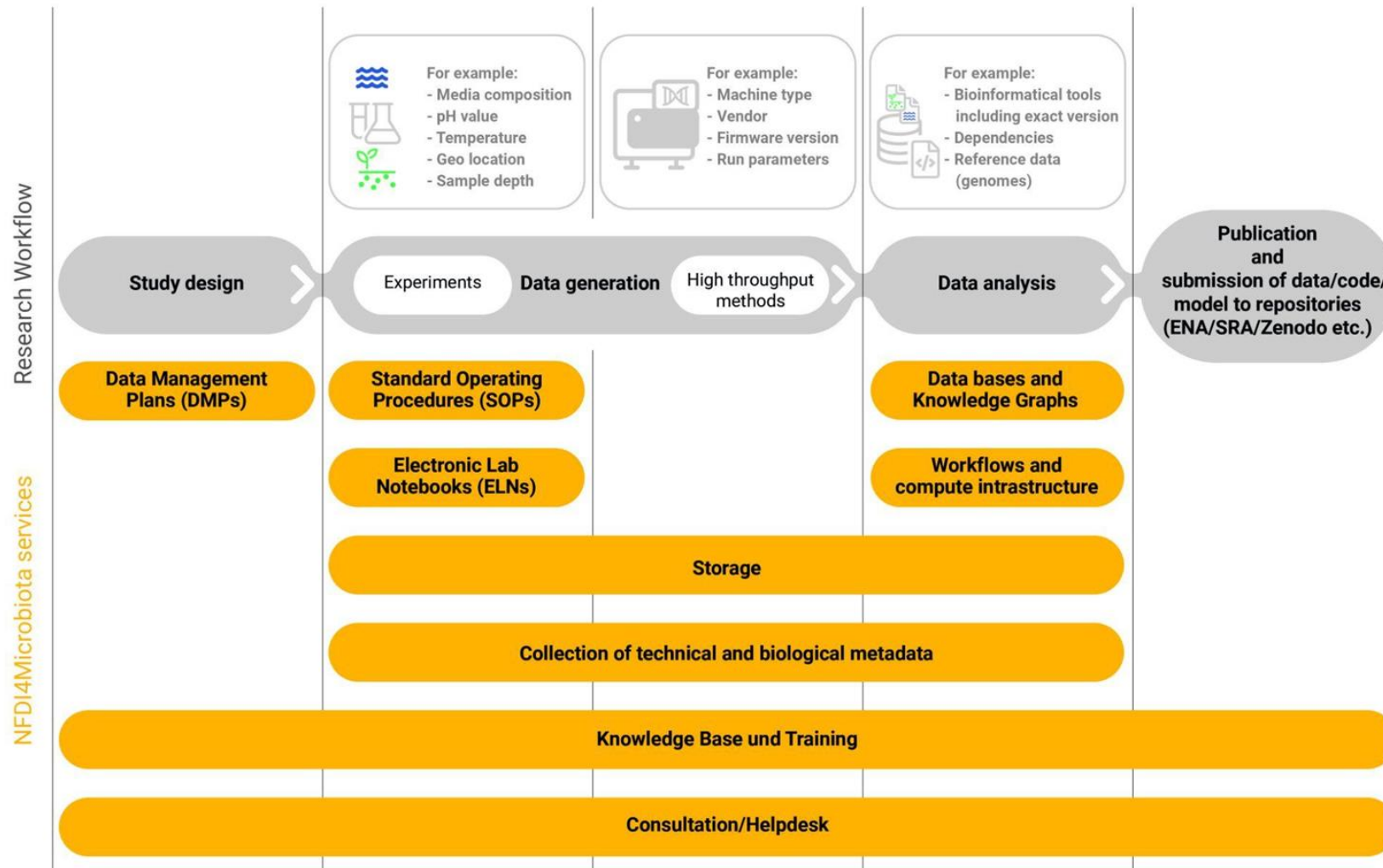
NFDI4Microbiota LTA activities

06.11.2024 NFDI WG LTA Workshop

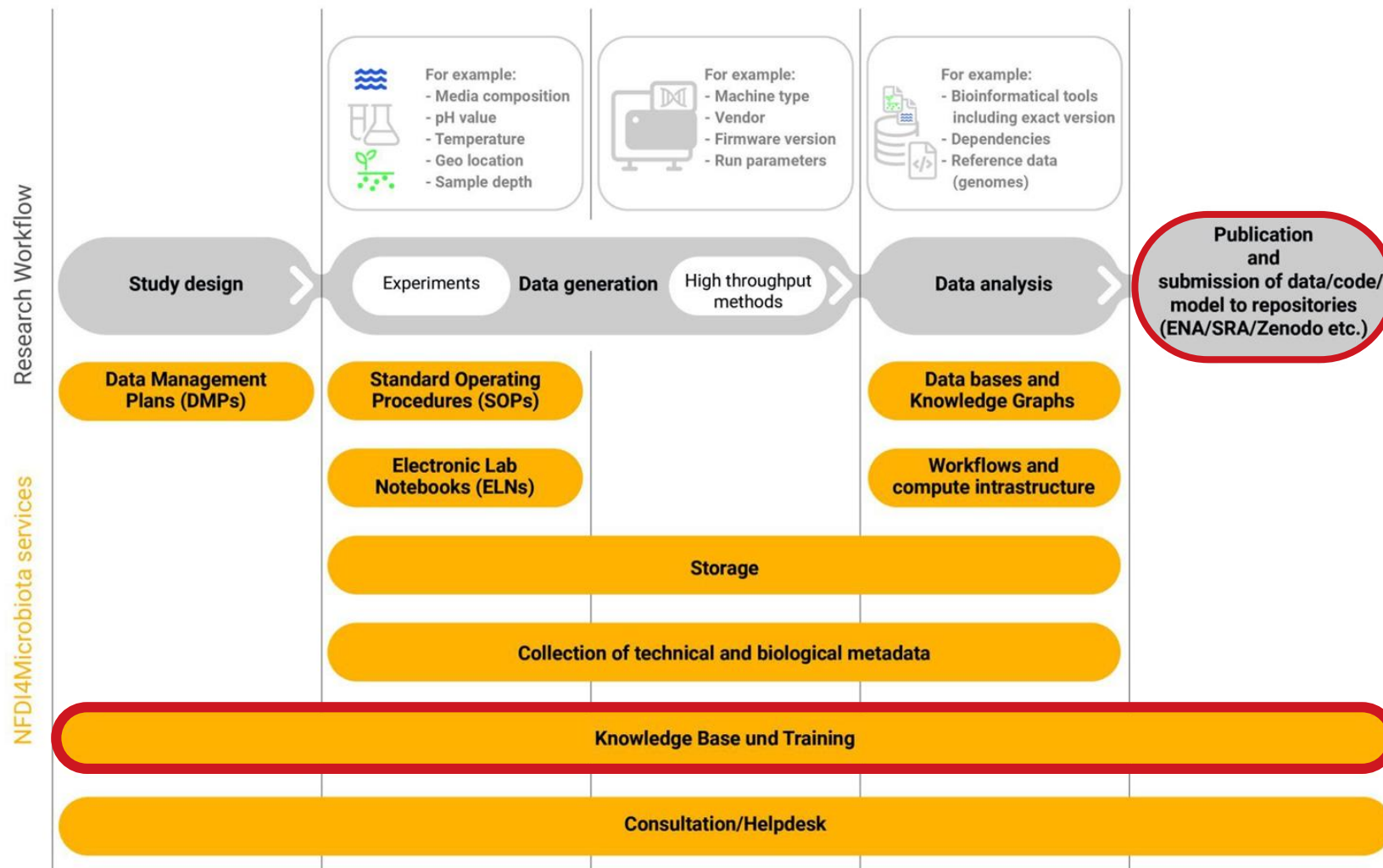
Katharina Markus



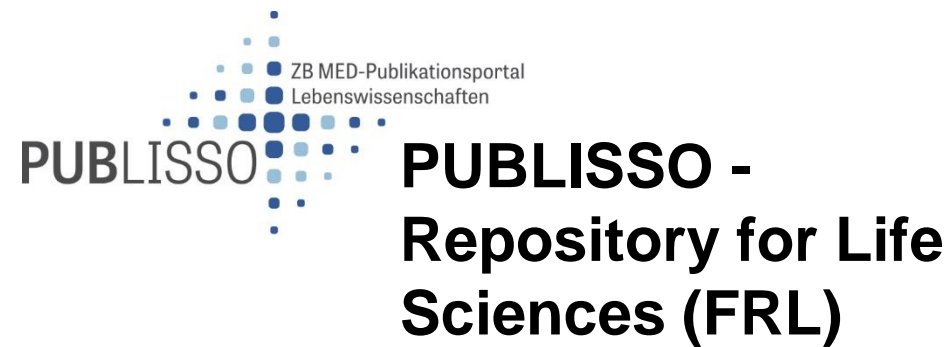
(1) Fig.: Barbara Sieman,n CC-BY-NC-ND



Source: NFDI4Microbiota 2024



Aruna Object Storage ARUNA
EMBL-EBI Repositories EMBL-EBI
Repository for Life Sciences (FRL)



- Intended: bitstream preservation via ARUNA
 - any (microbiota) data
- Content preservation via EMBL-EBI
 - omics data
- Content preservation via ZB MED Digital Archive
 - Long-tail data / long-tail formats



2024

- SOPs, training material regarding LTA
 - Proposal: M3.6 Long-Term Preservation
 - -> Webpage on the NFDI4Microbiota Knowledge Base (3)
- Developing LTA topics, discussions and exchange
 - -> Workshop on Preservation of Life Science Data hosted by NFDI4Microbiota in September 2024

- Preservation of (long-term reusability of) research **software**
- Long-term **responsibility** and preservation policies across microbiota research community

- ⦿ (1) <https://digitalpreservation.nl/seeds/icon-digital-preservation/> by Barbara Siermann CC-BY-NC-ND <https://creativecommons.org/licenses/by-nc-nd/4.0/>
- ⦿ (2) <https://knowledgebase.nfdi4microbiota.de/RDM-Preserve/24-aruna-object-storage.html>
- ⦿ (3) <https://knowledgebase.nfdi4microbiota.de/RDM-Preserve/25-digital-preservation.html>

Thank you!



markus@zbmed.de

www.nfdi4microbiota.de
[@nfdi4microbiota](https://twitter.com/nfdi4microbiota)



Funding information
DFG NFDI 52/1
Projekt no. 501930651

National Research Data Infrastructure for and with Computer Science

<https://nfdixcs.org/> as part of <https://nfdi.de>

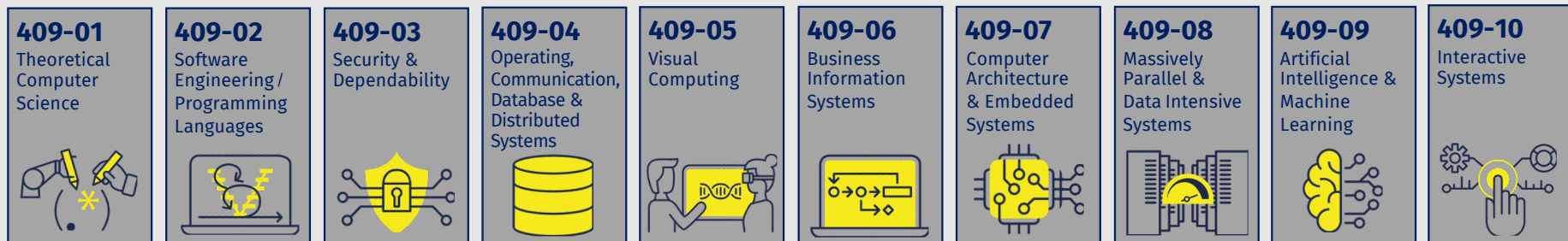
Long Term Archival presented by Michael Goedicke



Challenges of LTA in Research Data Management in Computer Science



- Context ... Execution, Experiment
- Software
- Diverse Types, Meta Data



Scientific Evidence

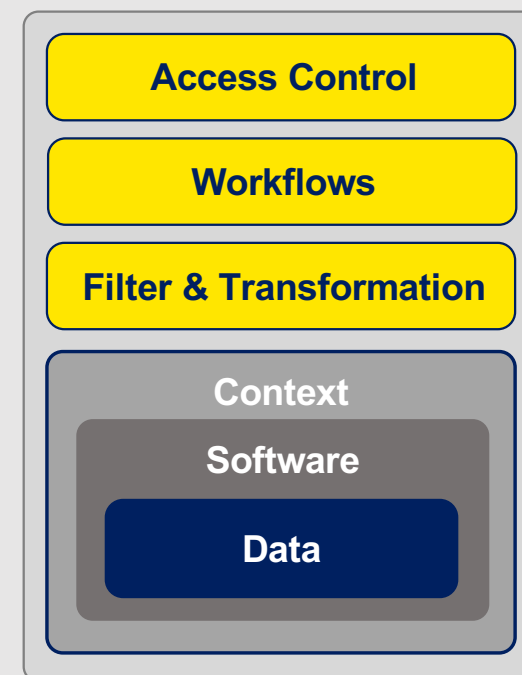
Research Data

Publication

No Data without Software
... what is about Hardware?

RDMC as Time Capsule c.f. FAIR Data Object

- Data → processed in different forms ... SW needed
- Bringing Data back to life → Execution of SW
- Operating system, frameworks, libraries ...
... need to be fixed and optionally referenced
... Software Heritage Foundation

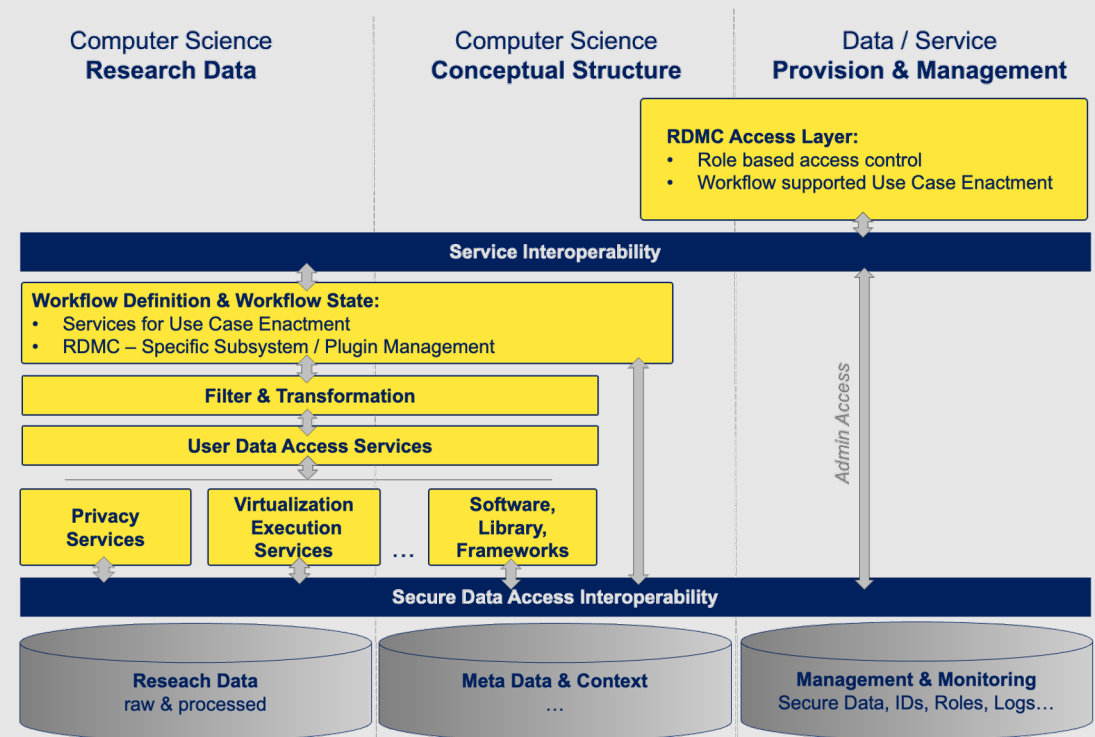


RDMC – Central NFDI_XCS Concept

Architecture – Implementation – Operation



- Architecture: sustainable result
- Implementation(s)
 - ... series of prototypes
 - ... reference implementation
 - ... purpose-built versions
- Reusable Execution Environment
- Technology change
 - ... Architecture!



Some Further Aspects

- Very large datasets
- Various levels of guarantees
 - **No** → completely open, minimal requirement regarding meta data, context etc.
 - **Privacy, IPR** involved → pseudonymized/open – partially or completely – encrypted using PKI, role based access control and possibly workflows needed
 - **Stronger version of privacy / IPR** needed → pseudonymized/trusted (see above) plus run in a trusted environment (provided by NFIDxCS)
 - **Strongest requirements for privacy** → Detached/Standalone no internet connection, computing is moved to the protected data
- Compatibility / Interoperability e.g. FAIR DO

NFDI_xCS Consortium



Thank you for your attention





NFDI4Objects

Research Data Infrastructure
for the Material Remains of
Human History

LTA in NFDI4Objects

TA 5: Storage, Access, Dissemination: the “Machine Room”
to Open and FAIR Access Points

Juliane Watson (DAI)

How does NFDI4Objects practice digital preservation or LTA?

Beyond Passive Data Storage: Long-Term Archive Services IANUS and baureka.online for the N4O-Community

Physical archiving of incoming N4O research data:

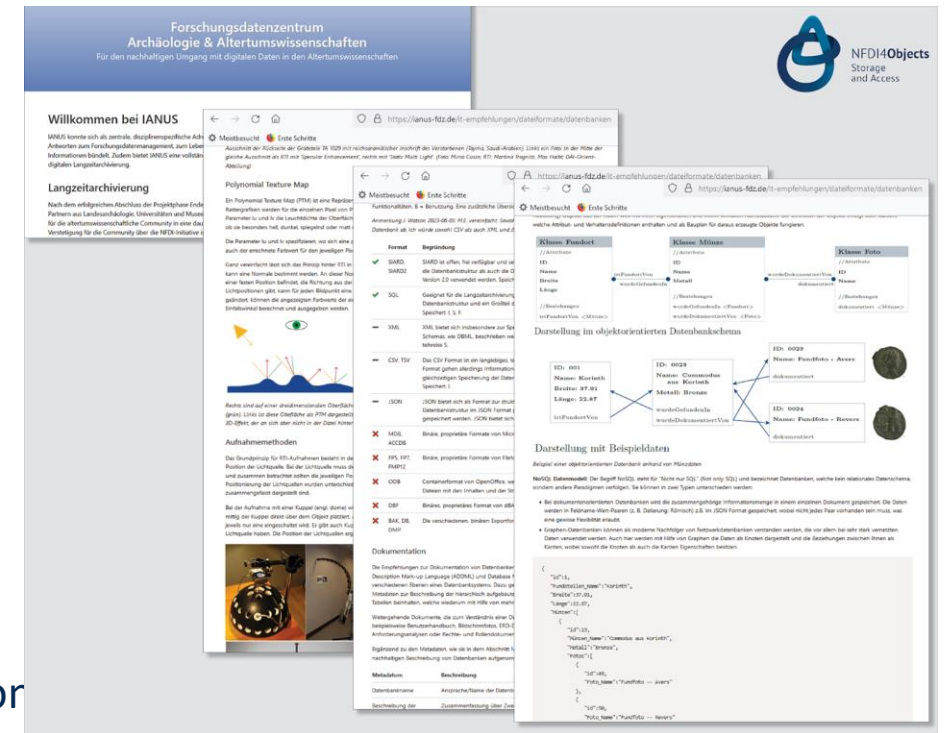
- **subject specific bitstream**
IANUS: logical or semantic preservation, based on the amount of capacity/resources available for a data package
- Process chain to **prepare incoming research data** and **interfaces** to N4O's main information resources **for a semi-automatic ingest**
- **IANUS features OAIS Reference Model** and modular architecture and additionally **FAIR dissemination of N4O research data and re-use via the IANUS Research Data Center and baureka.online portal**



Which LTA-related activities or initiatives are part of N4O's tasks?

LTA as a service for the community and Development and implementation of **Standards and Best Practices:**

- N4O supports researchers producing, archiving and re-using object-specific data through a large part of the research data life cycle. → bottom up + top down CC Onsite Documentation and TWG Sach- und Geodaten
- IANUS provides a **rich set of Research Data Recommendations** enabling the correct **preparation of research data**, the “**IANUS FDM-Empfehlungen**”. The existing “**IANUS FDM-Empfehlungen**” will be **curated and hosted as a community driven N4O service**
- Subject specific DMP template will be provided to include LTA from the beginning and create more awareness for the necessary steps (technology and data structure)
- Help desk, teaching materials



Which LTA-related milestones has N4O consortium planned? What's the status?

- **Provide LTA for the community:**
Core services are operational
100%
- In order to **scale up the service** to the N4O community needs, the existing technological basis is currently in an **upgrade and renewal process**
10%
 - Targeted **acquisition of project data** to test and strengthen the pipeline
- **Interfaces** to relevant N4O Systems (KuniWeb, Goobi, iDAI.field etc.) are being **implemented**
20%
- **Development of a business model** with **template contracts**
20%



What does N4O want to discuss in detail with regard to LTA?

- Community (Standards, Archivwürdigkeit (selection criteria)
- What is L-long? How long are or can contracts be made for? what happens after? Data is removed? Given back?
Responsibility for LTA beyond lifetimes of people, employees in the digital archive
- Common NFDI wide Technology watch? Or does every archive need its own accepted formats sheet?

