

Data archiving strategies

in regard to CETAF facilities and planned DiSSCo
services – highlighted by COST Mobilise

Dagmar Triebel

- 26 active members
- confirmed their continuous interest in April 2020, see https://costmobilise.biowikifarm.net/wiki/Working_Group_WG4
- 13 members from 8 countries participated at the Warsaw 3-day meeting in February 2020



Establishing a **common** ground of **understanding** for long-term data preservation and archiving

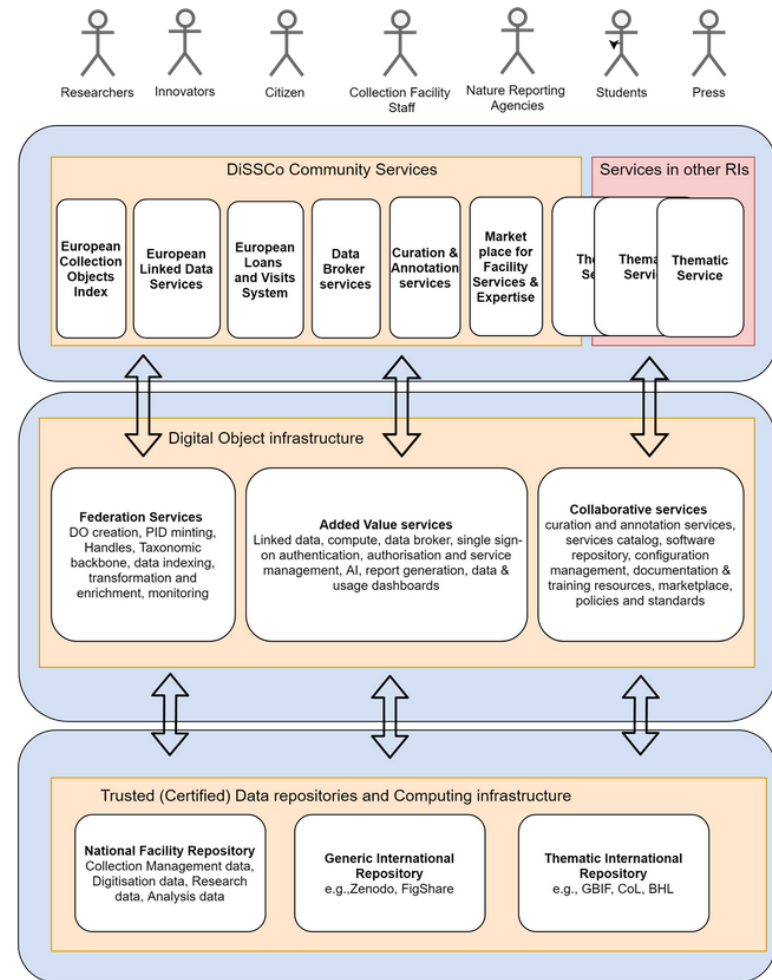
- especially in regard to CETAF taxonomic facilities and to the DiSSCo community services



3

2

1



<https://www.dissco.eu/what-is-dissco/technical-infrastructure/>

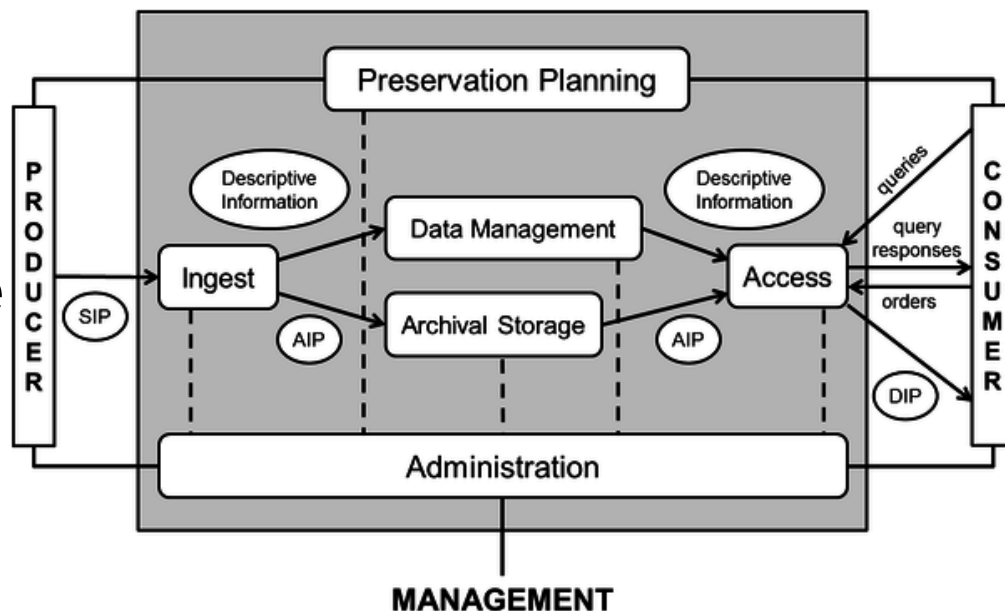
- Long-term data preservation and archiving is an essential part of the **FAIR** data life cycle



<https://www.gfbio.org/training/materials/data-lifecycle>



- Archiving should be done according the Open Archival Information System (**OAIS**) Reference Model (ISO norm 14721)
- Archival Information packages (**AIPs**) at all major steps of data processing
- Data archiving is **migration**

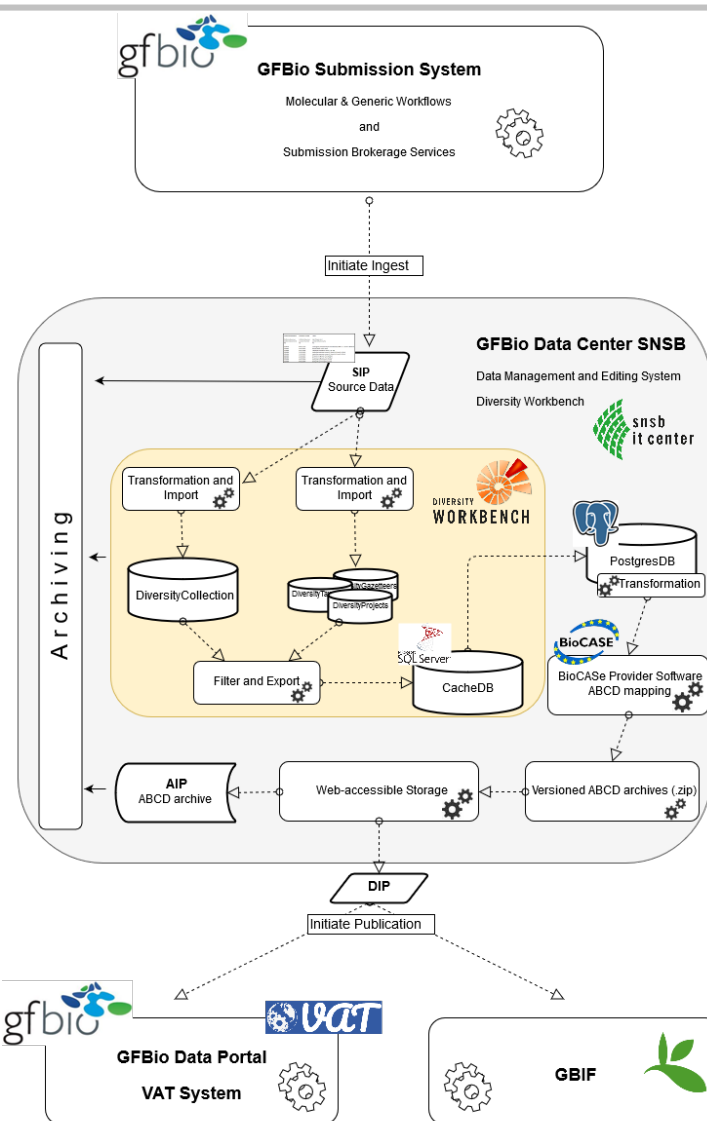


adapted from: Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-M-2, Magenta Book, June 2012, Figure 4-1.

https://gfbio.biowikifarm.net/wiki/ISO_Standards_for_Digital_Archives



- Data archiving is essential for
 - open and distributed information infrastructures and services like DiSSco
 - national and regional research data repositories like GFBio data centers
 - single CETAF taxonomic facilities with data repositories



https://gfbio.biowikifarm.net/wiki/Publication_of_Type_1_Data_via_BioCASE_Data_Pipelines_at_SNSB_Data_Center



- Acting as an European expert and interest group
- **Working together for**
 - for collecting useful information (MediaWiki-based knowledge base in the Biowikifarm)
 - for setting up a draft glossary of terms

Definitions of core terms in the data archiving context

DRAFT GLOSSARY [\[edit\]](#)

Data archive or data library or data repository

- *Definition: a centre of expertise in data archiving; some of them are listed in the [Registry of Research Data Repositories](#) --> re3data.org; see https://en.wikipedia.org/wiki/Data_library and http://oad.simmons.edu/oadwiki/Data_repositories*

Data archiving

- *Definition: is the process of moving data that is no longer actively used to a separate storage device for long-term retention. Archive data consists of older data that remains important to the organization or must be retained for future reference or regulatory compliance reasons. Data archives are indexed and have search capabilities, so files can be located and retrieved.... (see <https://searchdatabackup.techtarget.com/definition/data-archiving>)*

Bit (level) or bitstream preservation

- *Definition: Bit Level Preservation: A term used to denote a very basic level of preservation of the digital object as it was submitted (literally preserving the bits forming a digital object) (see <https://libguides.bodleian.ox.ac.uk/digitalpreservation/whatisdp>, <https://irods.org/uploads/2014/07/Principles-of-Archival-of-Digital-Assets.pdf> published by iRODS, 2014)*

Bit (level) preservation processes and strategies

- *Definition: (see <http://www.paradigm.ac.uk/workbook/preservation-strategies/degree-bitstream.html>)*

File format

- *Definition: A file format is a standard way that information is encoded for storage in a computer file. It specifies how bits are used to encode information in a digital storage medium. File formats may be either proprietary or free and may be either unpublished or open. (see https://en.wikipedia.org/wiki/File_format)*

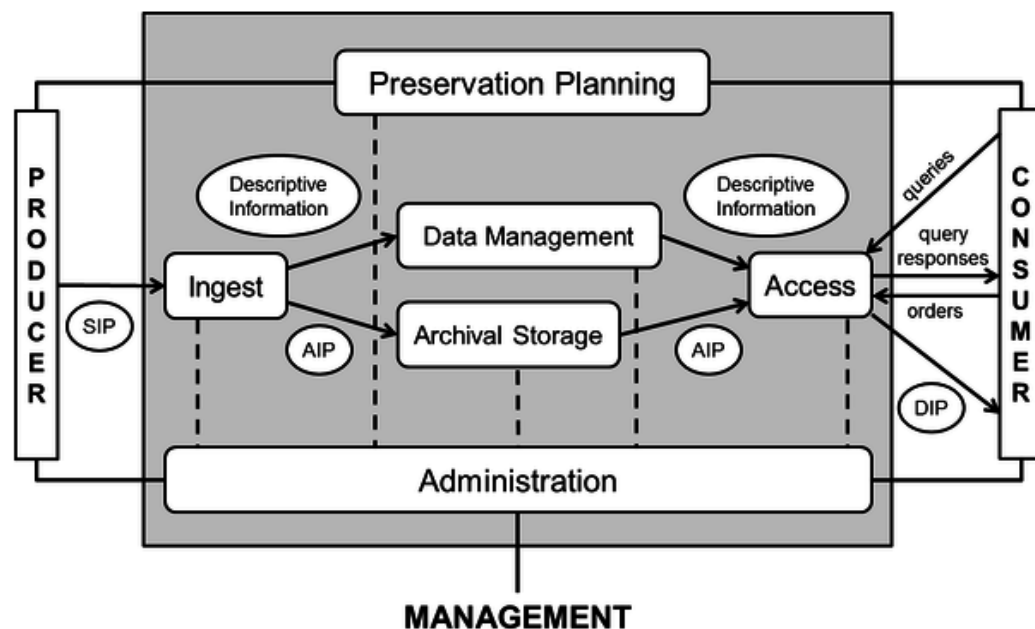
Archive formats and archive files



https://costmobilise.biowikifarm.net/wiki/Definitions_of_core_terms_in_the_data_archiving_context

- Drafting and publishing a **guideline** on data archiving
 - addressees: curatorial and research staff of CETAF facilities (that means, for a non-IT specialist audience)
 - five group members are building an **editorial team**
 - **main chapters** concern:
 - General topics (glossary included)
 - Functional topics
 - Best practices, costs and certification

- **Scope** of long term data preservation and archiving
- **Limitation** of long-term data preservation and archiving
- **Commitments** for data archiving tasks and long-term financial support
- Archival time span



adapted from: Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-M-2. Magenta Book, June 2012. Figure 4-1.

https://gfbio.biowikifarm.net/wiki/ISO_Standards_for_Digital_Archives



- Data archiving **strategies** (migration versus emulation)
- **Standards** for functional archiving + access of archival information (e.g., zip)
- Applicable binary and textual **archiving formats** (e.g., tiff, pdf, xml, SIARD)
- Applicable ISO norms
- Archiving **complexity** and relational content (e.g., databases)

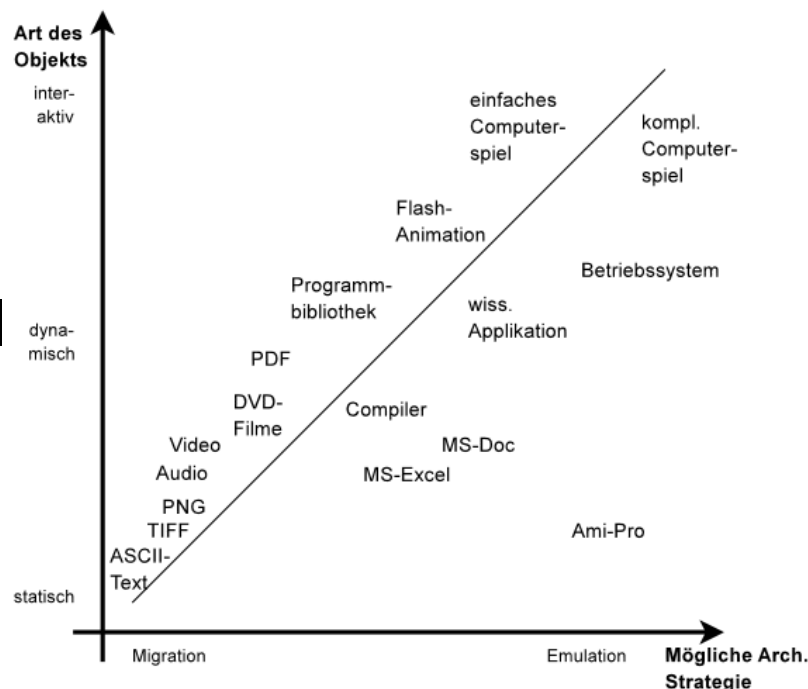


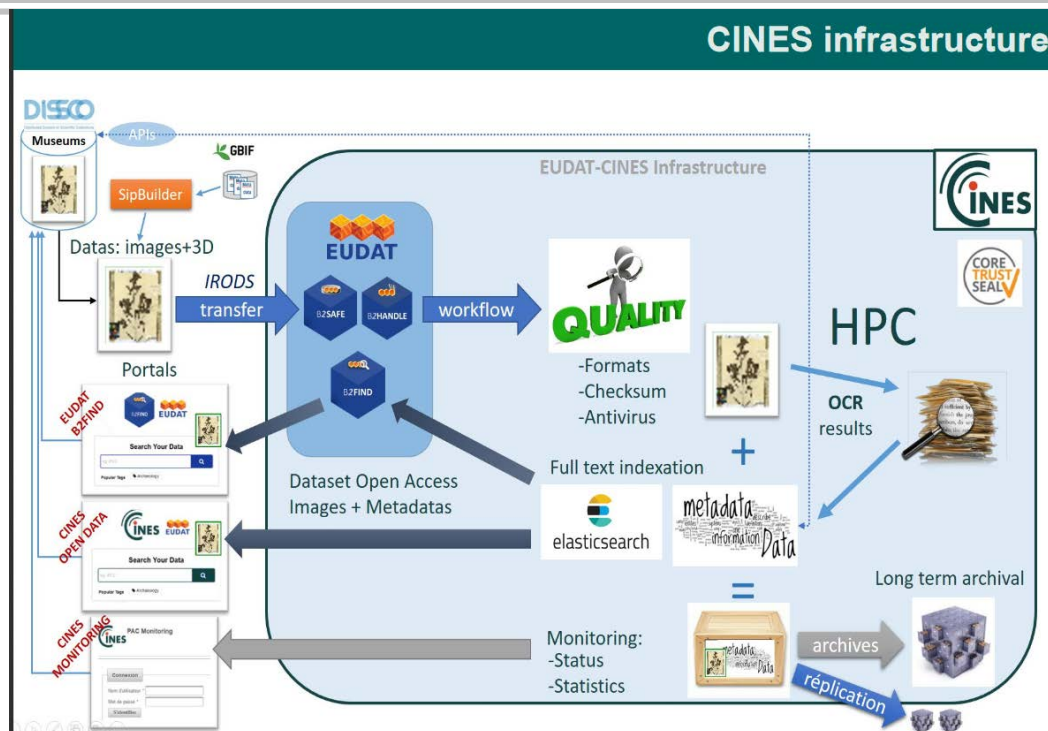
Abbildung 3.10: Vorschlag einer Einordnung verschiedener digitaler Objekte in Abhängigkeit ihres Typs in eine Langzeitarchivierungs-Strategiematrix.

See Suchodoletz 2009, <https://d-nb.info/1047826933/34>



- Documentation of best practices in selected **trusted** data repositories
 - taxonomic facility-specific
 - thematic
 - open network-specific
 - general
- Certification (e.g., Core Trust Seal)

See Cazenave et al. 2019,
https://icedig.eu/sites/default/files/deliverable_d6.2_icedig_digitisation_infrastructure_design_for_eudat.pdf



Participants [\[edit\]](#)

Convenors: Gila Kahila Bar-Gal and Dagmar Triebel

last name	first name	email address	country, affiliation	notes
Alvarez Dorda	Beatriz	balvarez@mncn.csic.es	Spain, CSIC ↗ and MNCN ↗	group "Functional topics"
Ben-Hur	Eyal	eyal.ben-hur@mail.huji.ac.il	Israel, HUJI ↗	group "Best practices"
Bingert	Sven	sven.bingert@gwdg.de	Germany, GWDG ↗	group "Best practices", editorial team
Caviere	Fabien	fabien.caviere@mnhn.fr	France, MNHN-UMS ↗	group "General topics"
Chikalanov	Alexander	ctmdevelopment@yahoo.com	Bulgaria, ULSIT ↗	group "Best practices"
Gagnier	Pierre-Yves	pierre-yves.gagnier@mnhn.fr	France, MNHN ↗	group "Functional topics"
Grobe	Peter	p.grobe@leibniz-zfmk.de	Germany, ZFMK ↗	group "Best practices"
Ivanovic	Dragan	dragan.ivanovic@uns.ac.rs	Serbia, UNS ↗	group "Functional topics"; editorial team
Jakimovski	Boro	boro.jakimovski@finki.ukim.mk	Macedonia, UKIM ↗	group "General topics"
Kahila Bar-Gal	Gila	gila.kahila@mail.huji.ac.il	Israel, NNHC ↗	group "General topics", editorial team
Neuman	Shoshana	Shoshana.Neuman@biu.ac.il	Israel, BIU ↗	group "Best practices"
Struck	Torsten	t.h.struck@nhm.uio.no	Norway, UiO-NHM ↗	group "General topics", rapporteur
Triebel	Dagmar	triebel@snsb.de	Germany, SNSB ↗	group "Functional topics", editorial team

https://costmobilise.biowikifarm.net/wiki/WG4_Workshop_%22Data_storage_and_archiving_strategies:_Towards_a_documentation_and_guideline%22_in_Warsaw

