

## *Mobilising Data, Policies and Experts in Scientific Collections - CA 17106*



---

### **WG 4 WORKSHOP Warsaw**

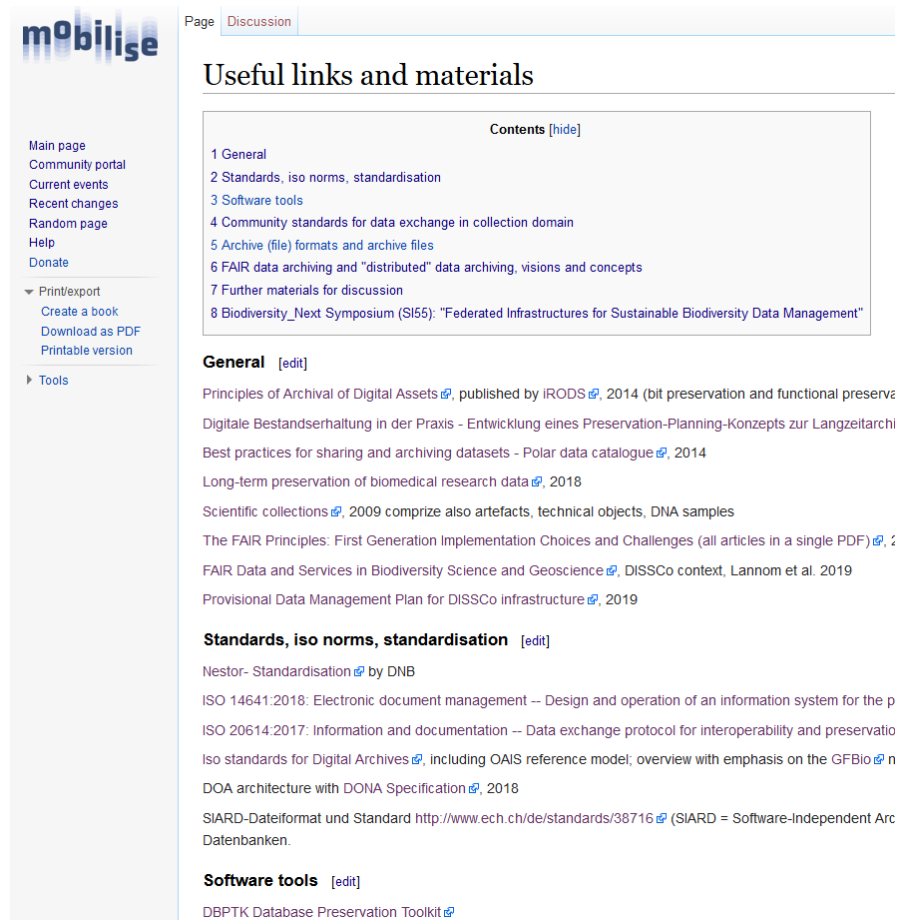
**Data storage and archiving strategies:  
Towards a documentation and guideline**

## Sofia Workshop:

- Defining of timeframes is essential
- Parameters and scope has to be scalable, reliable and reachable
- We preserve and archive to get access
- Archiving practices should result in FAIR data
- Archiving of data is migration of data

- Link collection on the topic “archiving” and “long-term preservation”
- Collection of WG 4 documents and presentations
- Glossary

[https://costmobilise.biowikifarm.net/wiki/Working\\_Group\\_WG4](https://costmobilise.biowikifarm.net/wiki/Working_Group_WG4)



The screenshot shows the mobilise website interface. On the left is a navigation menu with links like 'Main page', 'Community portal', and 'Tools'. The main content area is titled 'Useful links and materials' and features a 'Contents [hide]' table of contents with 8 numbered items. Below this, there are sections for 'General', 'Standards, iso norms, standardisation', and 'Software tools', each containing a list of relevant links and documents.

**Contents [hide]**

- 1 General
- 2 Standards, iso norms, standardisation
- 3 Software tools
- 4 Community standards for data exchange in collection domain
- 5 Archive (file) formats and archive files
- 6 FAIR data archiving and "distributed" data archiving, visions and concepts
- 7 Further materials for discussion
- 8 Biodiversity\_Next Symposium (SI55): "Federated Infrastructures for Sustainable Biodiversity Data Management"

**General [edit]**

- Principles of Archival of Digital Assets [↗](#), published by iRODS [↗](#), 2014 (bit preservation and functional preserve)
- Digitale Bestandserhaltung in der Praxis - Entwicklung eines Preservation-Planning-Konzepts zur Langzeitarchivierung [↗](#)
- Best practices for sharing and archiving datasets - Polar data catalogue [↗](#), 2014
- Long-term preservation of biomedical research data [↗](#), 2018
- Scientific collections [↗](#), 2009 comprise also artefacts, technical objects, DNA samples
- The FAIR Principles: First Generation Implementation Choices and Challenges (all articles in a single PDF) [↗](#), 2019
- FAIR Data and Services in Biodiversity Science and Geoscience [↗](#), DISSCo context, Lannom et al. 2019
- Provisional Data Management Plan for DISSCo infrastructure [↗](#), 2019

**Standards, iso norms, standardisation [edit]**

- Nestor- Standardisation [↗](#) by DNB
- ISO 14641:2018: Electronic document management – Design and operation of an information system for the p
- ISO 20614:2017: Information and documentation – Data exchange protocol for interoperability and preservatio
- Iso standards for Digital Archives [↗](#), including OAIS reference model; overview with emphasis on the GFBio [↗](#)
- DOA architecture with DONA Specification [↗](#), 2018
- SIARD-Dateiformat und Standard <http://www.ech.ch/de/standards/38716> [↗](#) (SIARD = Software-Independent Arc Datenbanken).

**Software tools [edit]**

- DBPTK Database Preservation Toolkit [↗](#)

**Focusing** on core requests of scientific collection target group:

- **Natural Science Collections**
- Similar types of Collections, e.g. those with biomedical data

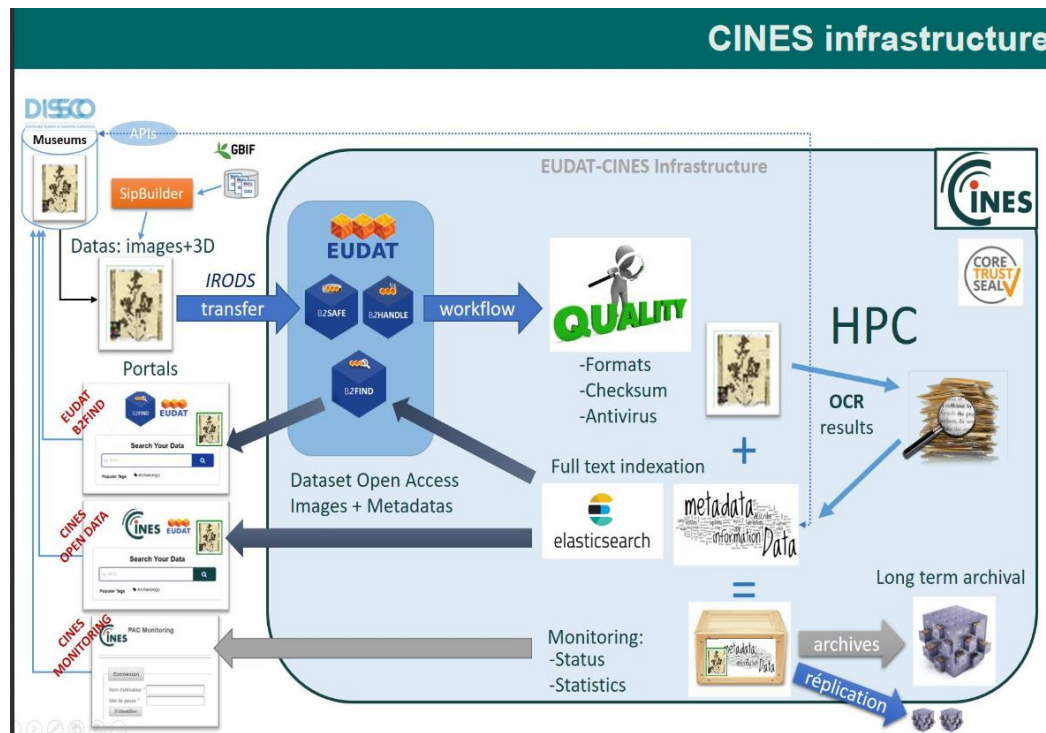
**Focusing** on long-term preservation and functional archiving of widely used archivable formats, e.g.,

- TXT UTF-8
- ODS 1.2
- XSD (XML based)
- SIARD (XML based)
- TIFF
- WAV
- PDF 1.7

## **Focusing** on archiving according (de-facto) standards and guiding principles

- Digital Object Architecture (DOA) with “digital collection” and “digital specimen”
- OAIS standard with AIPs, SIPs and DIPs
- E-ARK AIP pilot specification
- Archiving for FAIR usage

- (a) describing two or three best practice archiving solutions



- *(b) start to work on writing single chapters for a common guideline on archiving digital specimens and digital collections*



- *(c) start to work on certification issues of archiving processes and repositories with digital specimens and collections data*



We have 6 slots à 1,5 to 2 h

- Introduction, describing the focal points (e.g., stakeholders, technical formats, standards and architecture)
- Some presentations
- Working in 3 groups (= 3 slots) on single topics
- Summarizing results