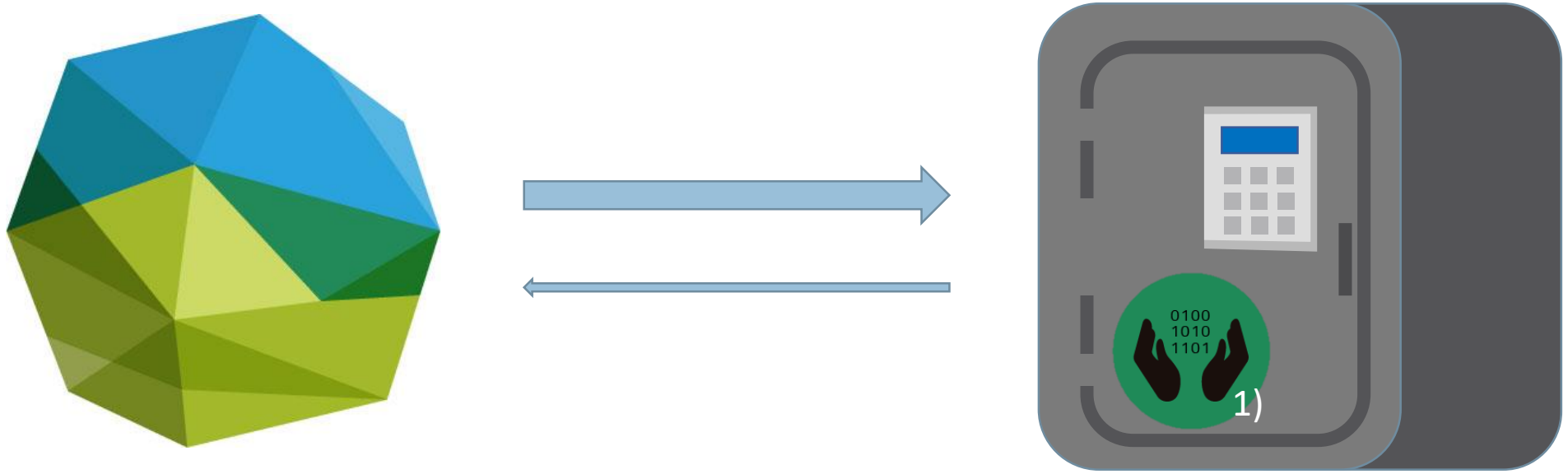


Long-term digital preservation of research data as a community-specific project

Katharina Markus,

ZB MED Digital Preservation

LIBER conference, Paving the way: Digital Access & Preservation

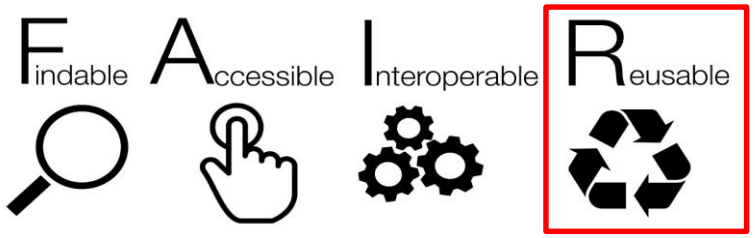
Information Centre for Life Sciences

ZB MED

1) Hands by ZALF

Digital Preservation: what does it entail?

- Can have many names (long term archiving, preservation etc.)
- Can have many meanings (institutional aims, intended level of preservation, etc.)



(1)

Wilkinson et al. (2016) *Scientific Data, 3, 1-9*

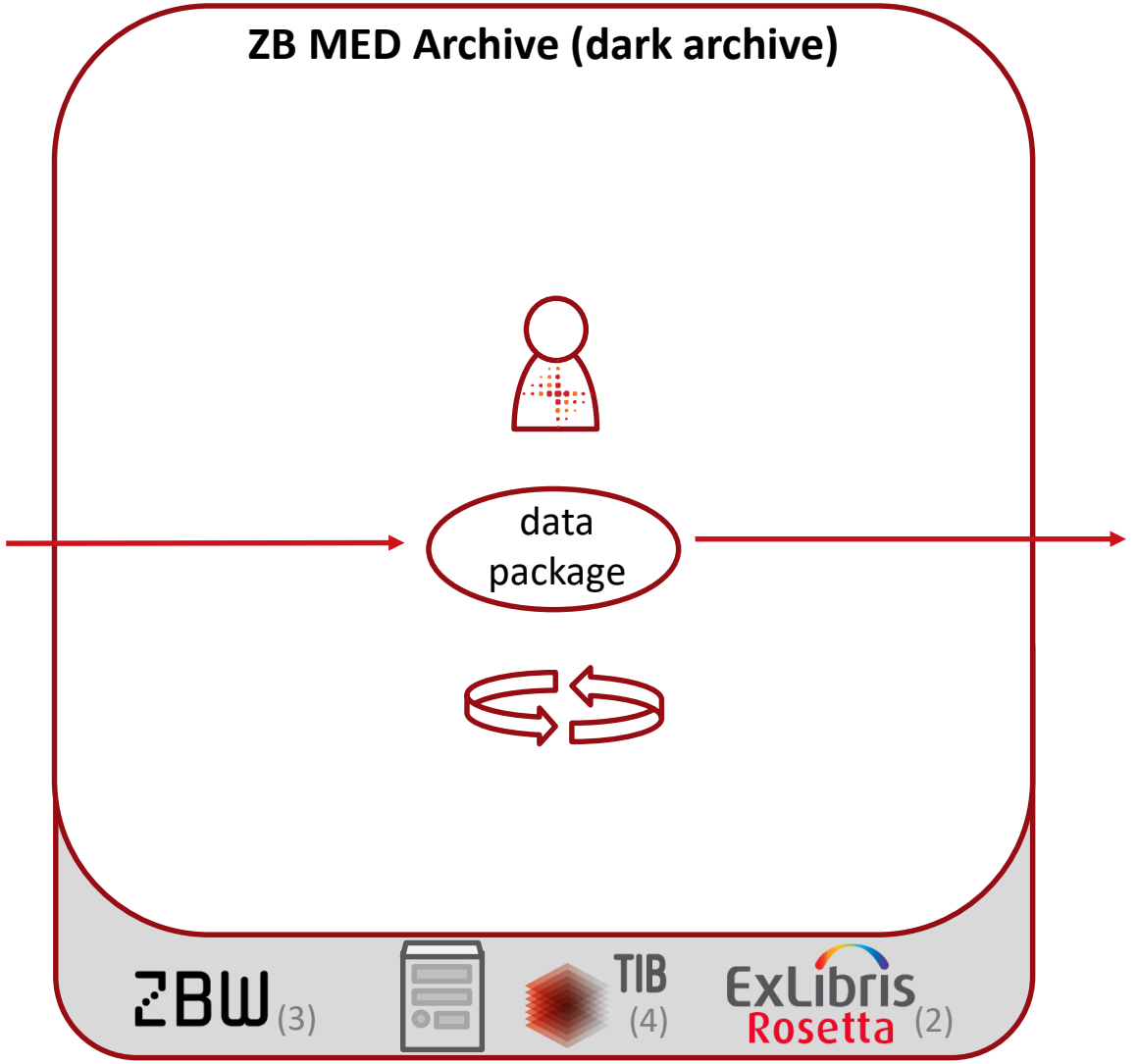
- Is not a guarantee of reusability but definitions of responsibilities and development of strategies
- Based on risks to reusability, long-term

„Object Levels of Preservation“ (p. 6 – 9 in Lindlar et al. (2020) *International Journal of Digital Curation, 15, 1, 1-26*)

ZB MED archive

Collaboration of 3 Germany National Subject Libraries, Information Centres

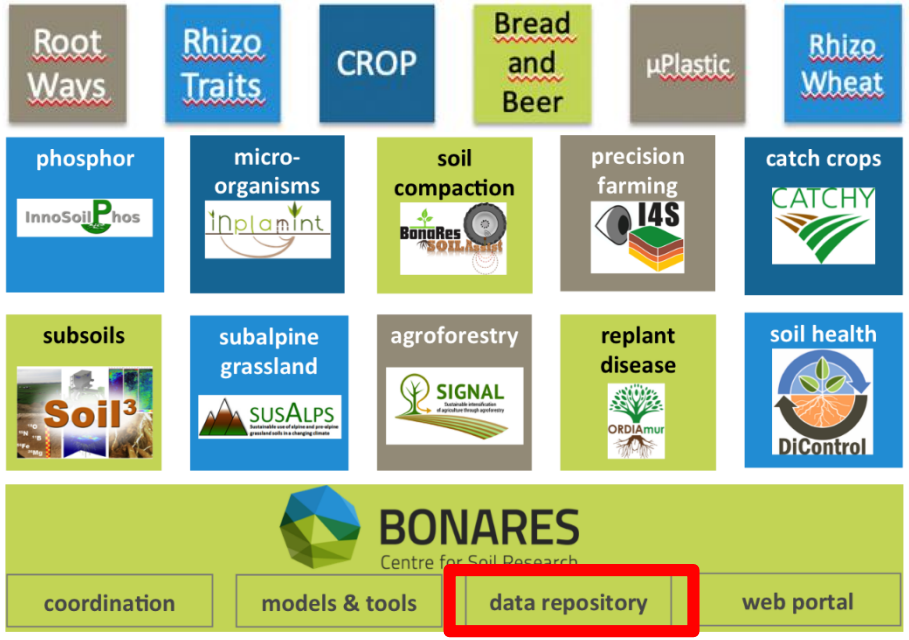
- Software: Rosetta (ExLibris) (2)
- System
  - Multi-tenancy
  - Hosting and administration: TIB
- Regular knowledge and experience exchange



Partner and data provider: ZALF, BonaRes Data Repository

Leibniz Centre for Agricultural Landscape Research (ZALF)

- BonaRes Project
  - Start: 2015
  - Duration: 9 years – perpetuation
  - 10+6 Collaborative Projects & BonaRes Centre

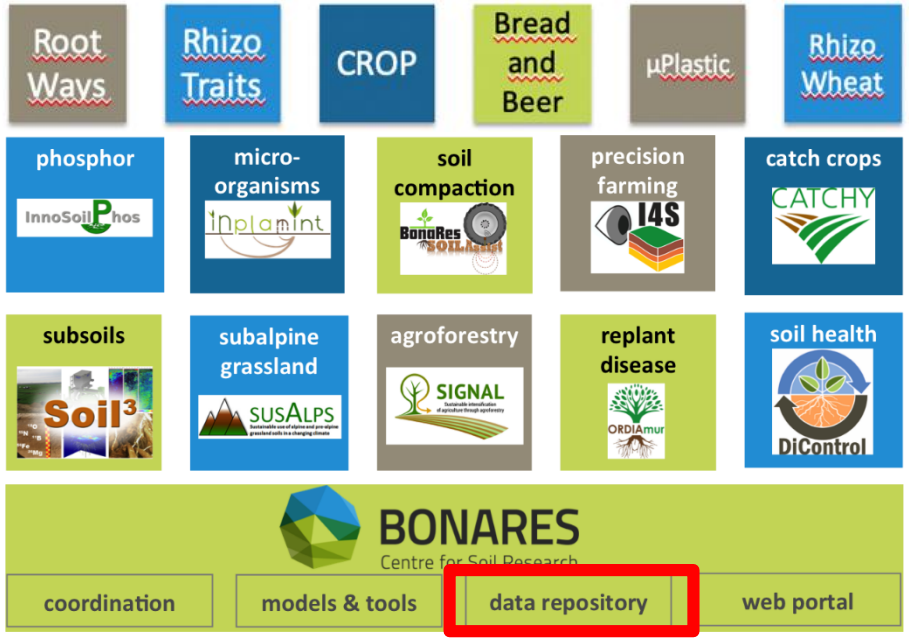


Slides:
Nikolai Svoboda,
BonaRes

Partner and data provider: ZALF, BonaRes Data Repository

Leibniz Centre for Agricultural Landscape Research (ZALF)

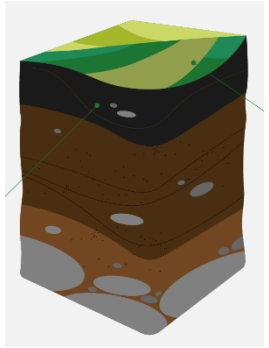
- BonaRes Project
  - Start: 2015
  - Duration: 9 years – perpetuation
  - 10+6 Collaborative Projects & BonaRes Centre



BonaRes Data Repository

Data: Soil & soil related data from:

- Long Term Experiments (LTE)
- (External) Research projects
- Publicly available sources
- Soil profiles
- …



Software

- con terra software

Slides: Nikolai Svoboda, BonaRes

Project Set-up

Collaboration of two institutions

- Taking advantage of two established, specialized systems and expert knowledge
- Tool development at both institutions
- Aims:
  - workflow set-up and tests
  - Balancing automation and human quality control
  - Knowledge exchange, including new developments (preservation watch, community watch)

DP of research data: some other initiatives

- EU Archiver project (EOSC)

https://archiver-project.eu/

- Generic data centers, discipline-specific data centres and databases

https://www.coretrustseal.org/

- Preservation of databases: SIARD standard

  Artefactual Systems and DPC (2021) *http://doi.org/10.7207/twgn21-06*

BonaRes Data analysis

Data set schema

Publication, data set landing page
| -- Research Data.csv 1-n
| -- Research Data.xlsx 1-n
| -- Research Data.gdb 1-n
| -- Research Data.txt 1-n
| -- Image Data.zip 1-n

| -- Meta data.pdf 1
| -- Meta data.xml 1

| -- Supplemental Material.* 0-n

1 data set = 1 data package

Selection of data

- Published datasets with DOI (atm.)
- All information necessary for re-creating the publication
- Format suitable for preservation (open simple formats)

BonaRes Data analysis

Data set schema

Publication, data set landing page
| -- Research Data.csv 1-n
| -- Research Data.xlsx 1-n
| -- Research Data.gdb 1-n
| -- Research Data.txt 1-n
| -- Image Data.zip 1-n

| -- Meta data.pdf 1
| -- Meta data.xml 1

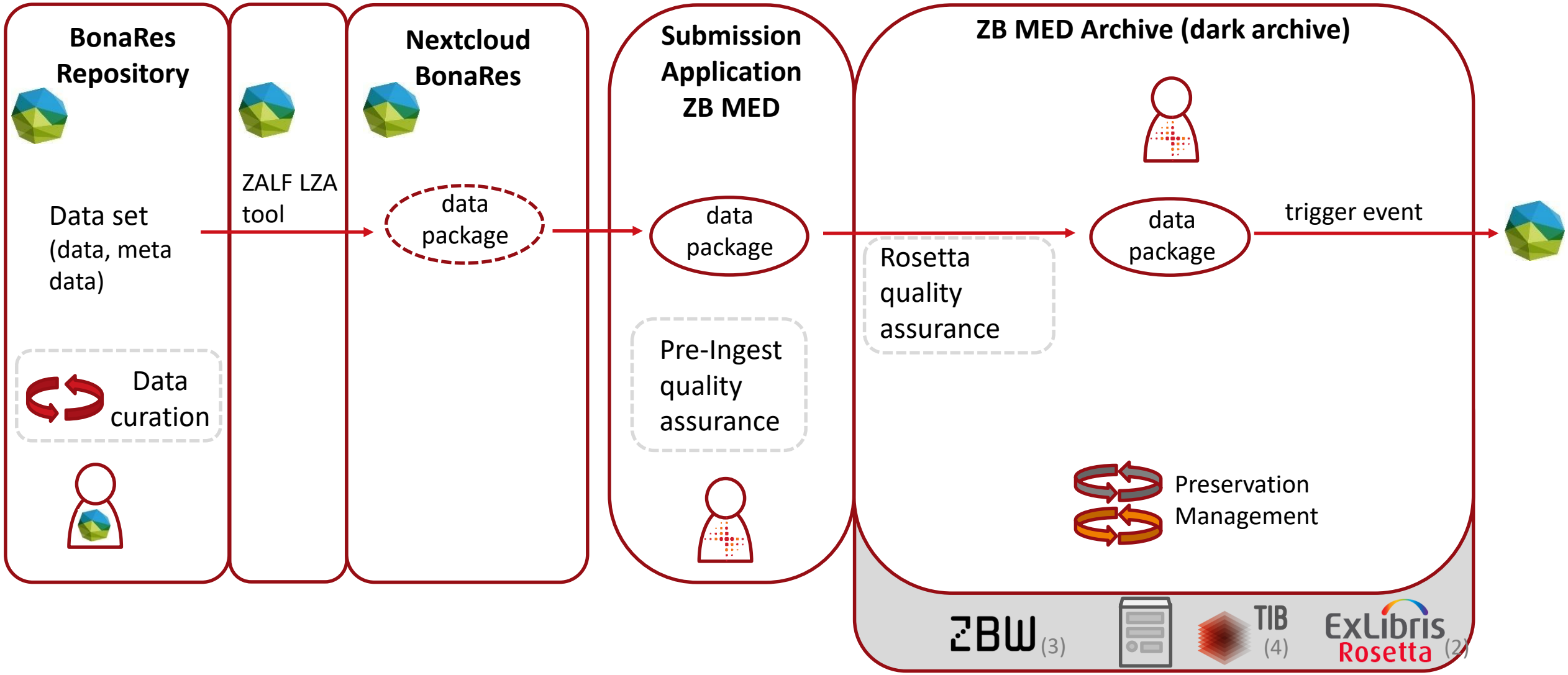
| -- Supplemental Material.* 0-n

1 data set = 1 data package

Selection of data

- Published datasets with DOI (atm.)
- All information necessary for re-creating the publication
- Format suitable for preservation (open simple formats)

BonaRes - ZB MED archive workflow setup



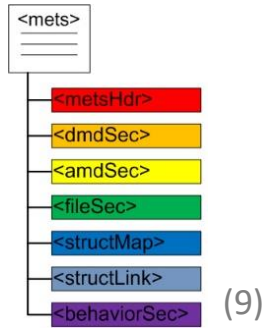
Documentation: metadata

Rosetta data model combines standards:

- Findable objects: descriptive meta data
  - Dublin Core (dc): e. g. creator, title
- Provenance, access rights, technical information
  - PREMIS: e. g. file format as file-md
- Structured MD and Rosetta data model components
  - METS: e. g. amdSec (administrative metadata section), sourceMD section

(7)

PREservation Metadata: Implementation Strategies (8)



(9)

ExLibris Rosetta (2)

ExLibris Rosetta (2)

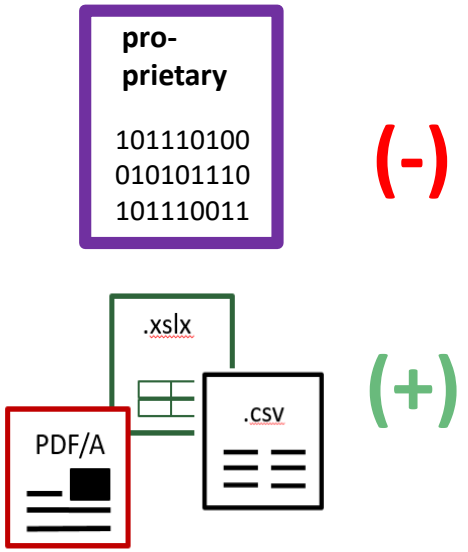
BonaRes Data-Workshop with researchers

Recommendations for BonaRes data publication

Aiming to decrease challenges specific to digital preservation of reseach data

- Recommendation of DP suitable file formats
  - Well known formats
  - Formats with open specifications, text-based formats/files

- Publications with sufficient rights, metadata



Data publication recommendations: https://zenodo.org/record/5747118#.YpcYIt9CRPY
DP recommentations: https://zenodo.org/record/5786303#.YpcX8d9CRPY

Lessons Learned

- Complex data structures are a challenge
- Workflow for updated data sets: versioned ID and/or date last modified needed
- Evaluation of degree of synchronization / automation and human quality control
  - resources and declaration of responsibility for active preservation
  - API development for RD transfer

Summary

- Challenge: resources and responsibilities
  - WF connecting data repository and archive
    - Automation at both institutions
    - Generation of Rosetta-compliant data packages via data provider
    - Quality control at various steps
  - Active preservation: responsibilities of ZB MED, exchange of specialist knowledge

- Challenge: multiple md standards -> combination of md of different sources

- Challenge: format diversity -> workshops with researchers

Outlook

- Workflow for complex data structures
- Testing of return workflow (ZB MED -> BonaRes repository)
- Development of data loss risks / use cases
- Establishment of preservation planning

Thank you for your attention!

Informationszentrum Lebenswissenschaften

ZB MED

www.zbmed.de

markus@zbmed.de