TIB

The relevance of FAIR data and digital preservation by implementing a data repository at a research institution

Oleg Nekhayenko,
Vienna, 16. November 2018
Focus on Open Science

Leibniz
Association

Agenda

1. Project LaZAR

2. Data transfer into digital archive

3. Findable

4. Accessible

5. Interoperable

6. Reusable

Project LaZAR

- Web platform for processing, publishing and digital preservation of multimedia field research data from regional studies

Project partners:

Institute for Slavonic and Caucasus studies
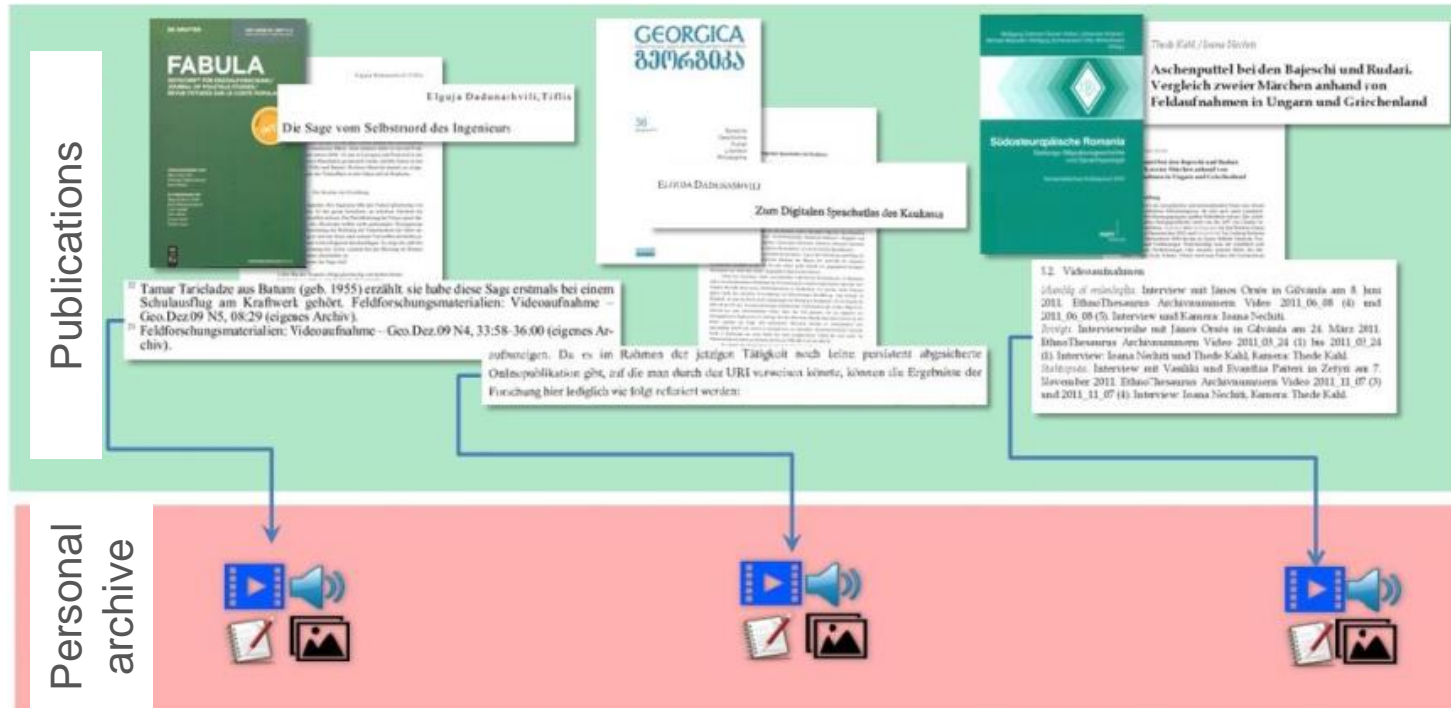**Professional competence and content**

Head Office of the GBV common library network
Development and managing of repository

German National Library of Science and Technology (TIB)
Digital preservation

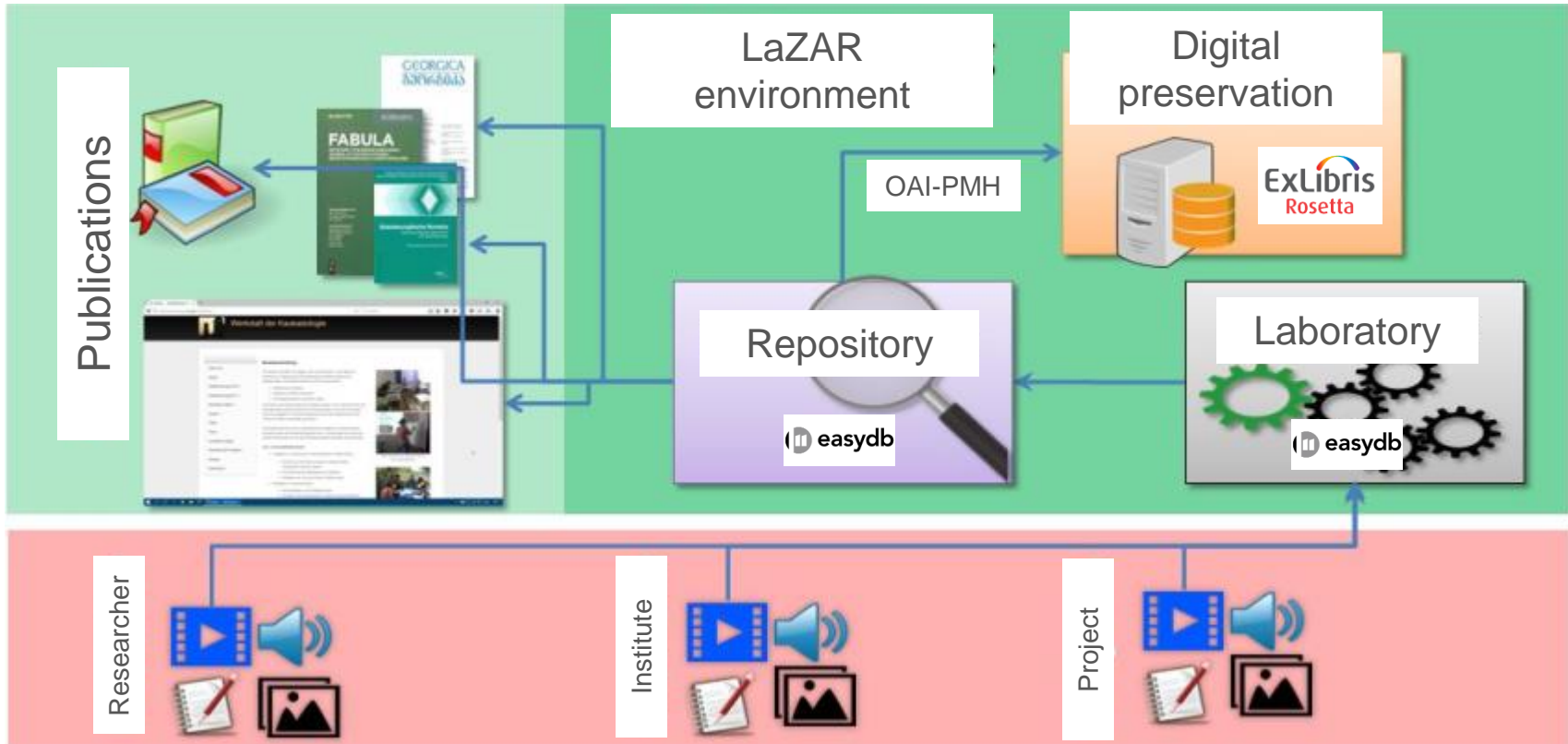# Multimedia citation in research publications.
Contemporary practice

Source: own representation, based on (Dadunashvili, Voß & Dührkohp, 2017)

- Lack of standardisation by referencing from personal archive

- Cited objects neither findable nor accessible and reusable (no DOI, no rich metadata according to any standards)

- Publication on Youtube, Dropbox, Google Drive only as provisional solution

Multimedia citation in research publications.
LaZAR Model

Publications

LaZAR environment

Digital preservation

ExLibris Rosetta

OAI-PMH

Repository

easydb

Laboratory

easydb

Researcher

Institute

Project

Source: own representation, based on (Dadunashvili, Voß & Dührkohp, 2017)

Hierarchical data model and citation

(1) Main collection

Feldforschung in Kachetien 2011

KONVOLUT · FREIGEGEBEN · CC BY-SA

(2) Sub collection

Verwandtschaftsbezeichnung

KONVOLUT · FREIGEGEBEN · CC BY-SA
LZA BEWILLIGT

(3) Sub collection

Herstellung des Besens

KONVOLUT · FREIGEGEBEN · CC BY-SA
LZA BEWILLIGT

(4) Media object

Herstellung des Besens_6

DATEI · FREIGEGEBEN · CC BY-SA · LZA BEWILLIGT

Main collection	Author (Year): Title. Main collection, LaZAR, Version, DOI
Sub collection/ Media object	Autor (Jahr): Title. In: Author, Main collection, LaZAR, Version, DOI

Data Transfer into TIB's digital archive

- Rosetta is configured to harvest the objects and their metadata via OAI-PMH Interface from easydb (harvesting can be scheduled)
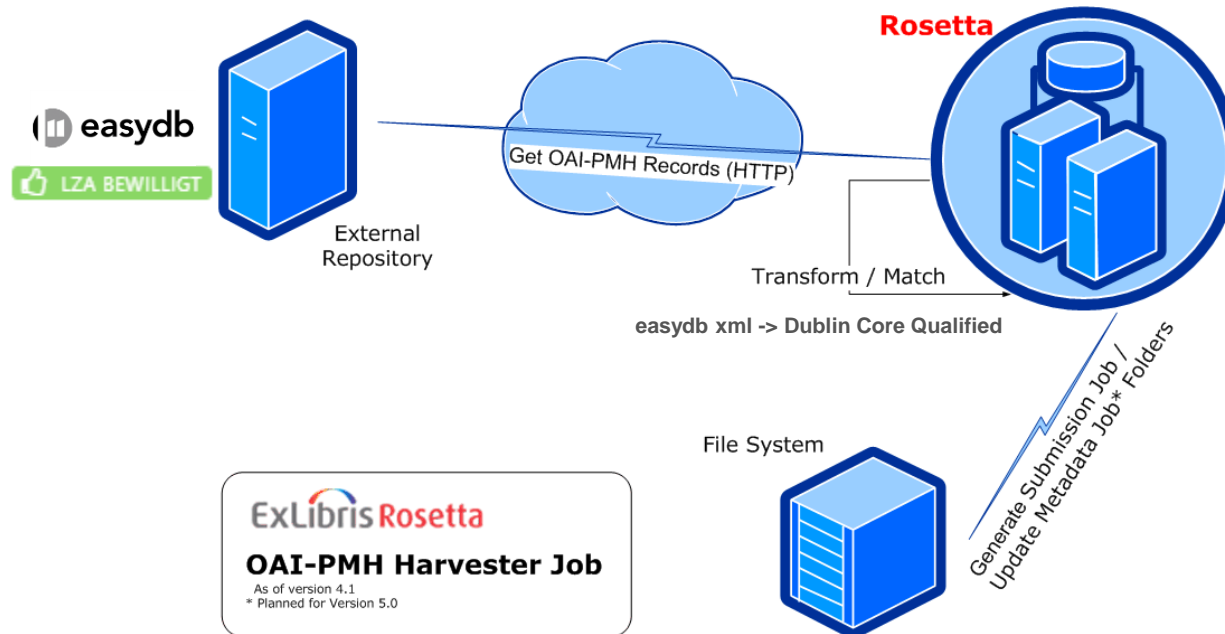
- TIB ensures the access to, readability and usability of the data in the long term according to OAIS reference model in a trustworthy digital archive

- Use of collections in Rosetta to reproduce the certain hierarchical structure of the data from easydb

Data Transfer into TIB's digital archive

```xml
1 <?xml version="1.0" encoding="UTF-8"?><dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:easydb="https://schema.easydb.de/EASYDB/1.0/objects/" xmlns:x
2 <dc:source>465</dc:source>
3 <dc:identifier xsi:type="dcterms:URI">https://lazardb.gbv.de/api/v1/objects/uuid/b5da3a88-217d-4be1-91b6-779a7f41f0d8/file/id/1001963724/file version/name/c
4 <dc:identifier>b5da3a88-217d-4be1-91b6-779a7f41f0d8</dc:identifier>
5 <dc:ispartof>252/253/254</dc:ispartof>
6 <dc:subject>Feldforschung in Kachetien 2011/Schafhaltung/Markierung der Herde</dc:subject>
7 <dcterms:license xmlns:dcterms="http://purl.org/dc/terms/">CC BY-SA</dcterms:license>
8 <dc:rights>LZA bewilligt</dc:rights>
9 <dc:title>Markierung der Herde</dc:title>
10 <dc:subject>Ohr,http://d-nb.info/gnd/4133604-5</dc:subject>
11 <dc:subject>Schafhaltung,http://d-nb.info/gnd/4116366-7</dc:subject>
12 <dc:subject>Markierung,http://d-nb.info/gnd/4443932-5</dc:subject>
13 <dc:creator>Fieber, Marco , ORCID: http://orcid.org/0000-0001-5491-4842(Friedrich-Schiller-Universität Jena)</dc:creator>
14 <dc:contributor>Dadunashvili, Elguja, ORCID: http://orcid.org/0000-0002-5434-0815, GND: http://d-nb.info/gnd/133557154, (ProjectLeiter)</dc:contributor>
15 <dc:contributor>425 v3_m_43(Informant)</dc:contributor>
16 <dc:contributor>424v1_m_52(Informant)</dc:contributor>
17 <dc:date> anfang: 2011-09-21</dc:date>
18 <dc:language>Georgisch, Glottolog: http://glottolog.org/resource/languoid/id/nucl1302, GND: http://d-nb.info/gnd/4124679-2</dc:language>
19 <dc:description>Bei dem unstrukturierten Interview wird erläutert welche Bedeutung der Ohrzeichen haben und wie das einzelne Zeichen heißt.</dc:description>
20 <dc:description>VisuelleDokumentation</dc:description>
21 <dc:coverage>Chantlisqure, 41.906562, 45.914137</dc:coverage>
22 <dc:relation>IstDokumentiertVon Interview zum Thema Markierung der Herde, 21.09.2011 </dc:relation>
23 </dc:dc>
```

Mapping to Dublin Core

FAIR data principles: Rosetta and easydb
F-findable

	LaZAR Repository (easydb)	TIB Digital Archive (Rosetta)
Persistent identifier	• DOI (not implemented yet), internal object-ID	• internal object-ID from easydb and DOI are included in the descriptive metadata to each object
Rich Metadata	• data are described by the researchers themselves in a standardized metadata form with controlled vocabularies and free text fields	• objects are described with descriptive, legal, administrative, technical, structural and event metadata
Searchable resource	• database search interface for internal and external queries	• detailed internal search interface for descriptive, legal, administrative, technical, structural and event metadata
Reference to the object in it's metadata	• Internal use of object-ID in the metadata to each object	• internal object-ID from easydb and DOI are included in the descriptive metadata to each object

F-findable

Metadata form. Source: LaZAR (easydb)

FAIR data principles: Rosetta and easydb
A-Accessible

	LaZAR Repository (easydb)	TIB Digital Archive (Rosetta)
data retrieval using standardized communication protocoll	• accessible web user interface via HTTPS • metadata and data released for access are retrievable via OAI-PMH	• accessible web user interface via HTTPS
open and free protocoll	• yes	• yes
allows authentication and authorization	• yes	• yes
metadata accessible if the object no longer available	• not planned yet	• rudimentary metadata are kept in Rosetta even the object was deleted

FAIR data principles: Rosetta and easydb

I-Interoperable

	LaZAR Repository (easydb)	TIB Digital Archive (Rosetta)
metadata use formal, accessible, shared, broadly applicable knowledge rep. language	• data are described using controlled vocabularies and free text fields • multiple metadata formats can be picked when using the OAI-interface (DataCite, Dublin Core, easydb xml)	Rosetta uses widely accepted standards: • a METS-XML file as a container for all structural information regarding an object • Dublin Core for descriptive metadata. • DNX, which is a PREMIS-derivative, for legal, administrative, technical and event metadata
metadata use Vocabularies that follow FAIR principles	• uses vocabularies to describe languages (ISO-Code, Glottolog, GND), locations (Geonames and -coordinates), creator (GND, ORCID), keywords (GND, Getty)	• vocabulary tags from easydb are included in the descriptive metadata to each object
metadata include qualified references to other metadata	• relations between objects are described with an easydb-internal controlled vocabulary	• relation metadata are mapped to Dublin Core during Ingest

FAIR data principles: Rosetta and easydb

R-Reusable

|	LaZAR Repository (easydb)	TIB Digital Archive (Rosetta)
rich metadata description	• the internal metadata schema is specific for easydb, but metadata can be exported via OAI-PMH as Dublin Core or Datacite which are richly described with relevant attributes	• Rosetta uses different metadata standards such as METS, DNX, Dublin Core which are richly described
Metadata contains clear and accessible data usage license	• accessible objects in easydb are assigned to a usage license, e.g. Creative Commons	• access rights are assigned to every ingested object to regulate access • the corresponding license tag is stored in the descriptive metadata
Metadata assocociated with detailed provenance	• Metadata about the method of survey, interviewees and creation date are described.	• Detailed audit trail which tracks every defined event

FAIR data principles: Rosetta and easydb

R-Reusable

	LaZAR Repository (easydb)	TIB Digital Archive (Rosetta)
(Meta)data meet domain-relevant community standards	• data and metadata are created according to a uniform data model (collections, data records) and metadata form.	• TIB's digital archive complies with the OAIS-standard and is certified as a trustworthy digital archive according to Data Seal of Approval and nestor-Seal (2017).

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB
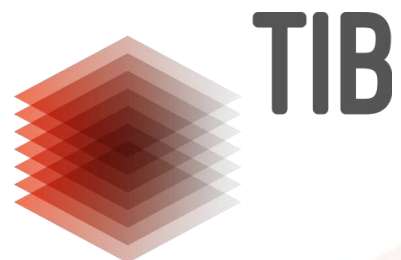
MORE INFORMATION
www.tib.eu

Contact
Oleg Nekhayenko
T +49 511 762-14518, Oleg.Nekhayenko@tib.eu

Leibniz
Association