



# FAIRness of Research Data in Humanities



Ljiljana Poljak Bilić, University of Split Library  
Kristina Posavec, University of Zagreb University computing centre SRCE





# (Research) Data in humanities?

- Ambiguity of term/concept (Unsworth et al. 2006, Borgman 2009, Edmond & Tóth-Czifra, 2018):
  - source vs data
  - the concept of research data is understood differently by the scientists from different fields and authors vs. readers
  - data as blanket term

“When is data?” instead of “What is data?”

## SSH data types:

- text samples
  - tables
  - images
  - maps
  - photographs
  - statistics
- graphs, figures, charts, visualisations)
  - databases
  - timelines
- audio-visual media (Prost et al., 2015)

## DH data types:

- print paradigm publications
  - electronic paradigm publications
    - single or collected/curated primary sources
      - software
- [patents/licenses]
- [ephemera, such as exhibitions and performances] (Edmond, 2019)

## Humanities data types:

- publications
- other primary sources (manuscripts, artworks)
- digital representation of cultural objects
- events, websites
- software
- documentation
- standards
- digital infrastructures
- personal data
- corpora
- born-digital artifacts (e.g., tags, associations, texts) (Gualandi, Pareschi, Peroni, 2022)



# Research methodology

## Sample:

Repositories: CROSSDA, DABAR, OPENAIRE

Subject category: Humanities: philosophy, theology, philology, history, art history, art science, archaeology, ethnology and anthropology, religious studies, and interdisciplinary humanities

Type of object: Research Data / Dataset

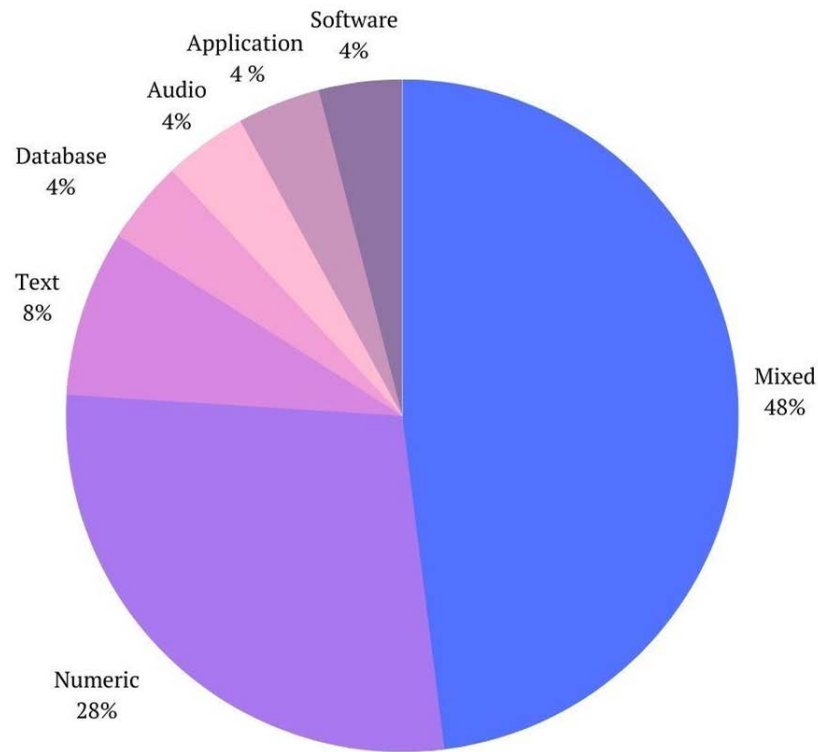
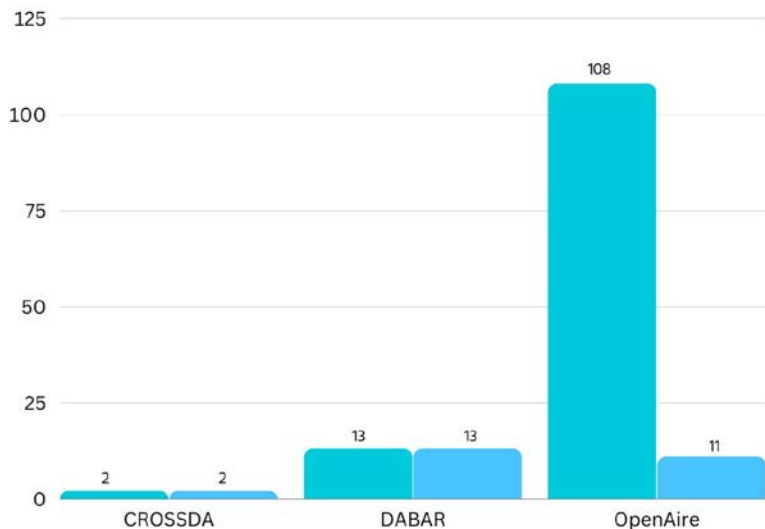
## Methodology:

1. Searching and retrieving datasets
2. Data analysis
  - A priori coding (Connaway & Radford, 2021); Attribute / Framework Matrix Coding (Saldaña, 2013) with codes: name of the repository, scientific field, dataset title, dataset type, openness, licence, read-me file, file format, downloads No, citations No, methodology, PID, Findable score (FS), FS FAIR level, Accessible score (AS), AS FAIR level, Interoperable score (IS), IS FAIR level, Reusable score (RS), RS FAIR level, FAIR level overall, FAIR %
  - Two independent researchers analyzed the data; in the second stage, a consensus was reached in case there were any different values applied



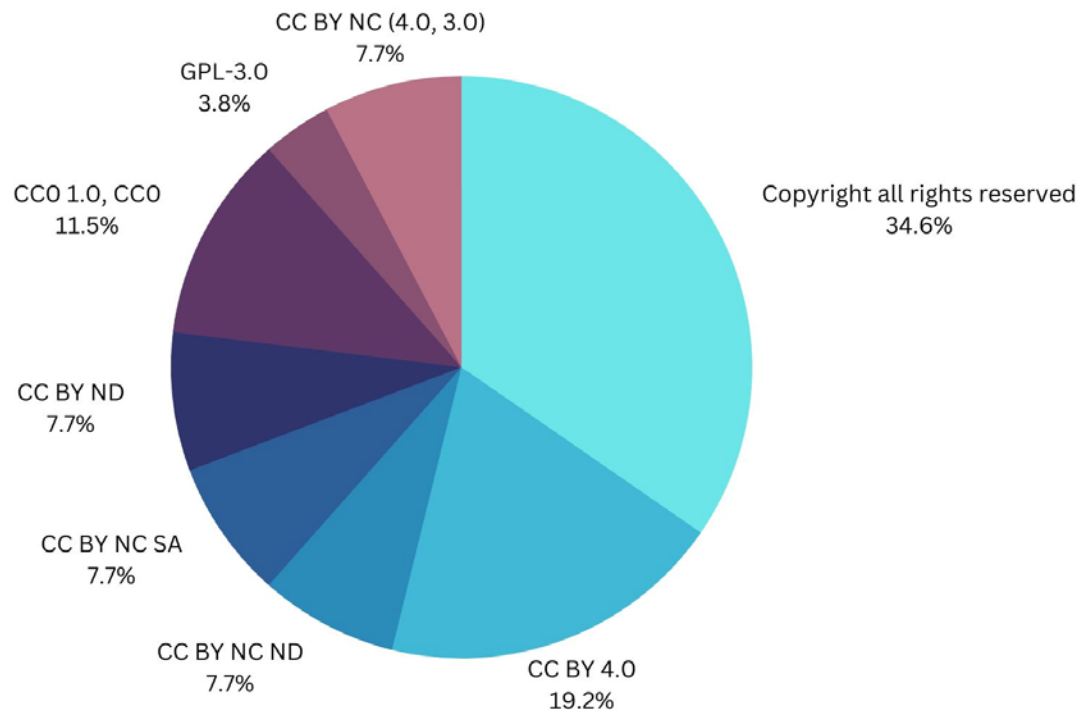
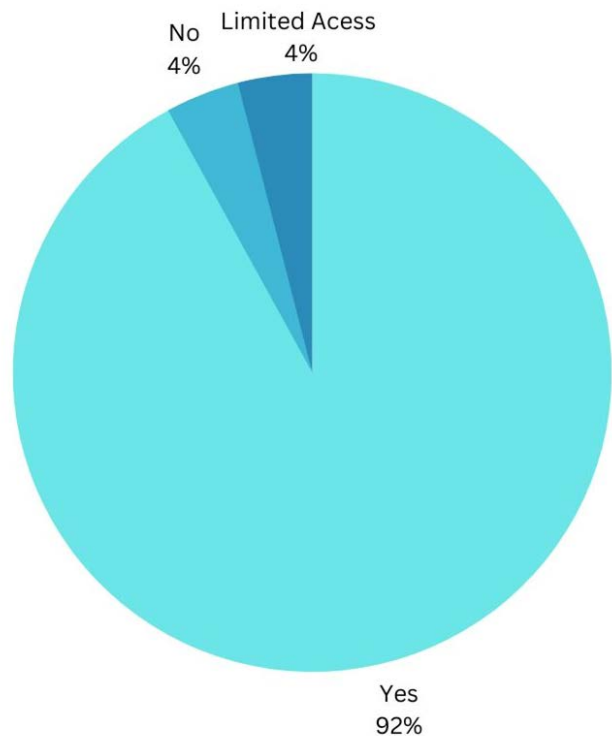
# RQ1: What types of research data are represented in humanities?

DARIAH-DE: Research data in humanities as all those sources/materials and results collected, written, described, and/or evaluated in the context of a research and research question in the field of human and cultural sciences and in machine-readable form for the purpose of archiving, citation and for further processing.





## RQ2: To what extent are datasets in the field of humanities represented in repositories by level of openness and under which license?



# RQ 3 To what extent does research data align with FAIR principles?

The evaluation of datasets FAIRness was conducted using the **FAIR Data Maturity Model** developed by **RDA Working Group “FAIR Data Maturity Model”**

The tool used for FAIR assessment is **F-UJI a web service** for assessing datasets and it is based on metrics developed by the **FAIRsFAIR project** which are proposed by the **RDA FAIR Data Maturity Model Working Group**.

Assessment is based on:

- URL or persistent identifier of dataset
- aggregated metadata:
  - embedded in the data (landing) page
  - retrieved from a PID provider.

# RQ 3 To what extent does research data align with FAIR principles?

Assessment metrics implemented in F-UJI tool:

## Findable

- Data is assigned a globally unique identifier.
- Data is assigned a persistent identifier.
- Metadata includes descriptive core elements.
- Metadata includes the identifier of the data it describes.
- Metadata is offered in such a way that it can be retrieved programmatically.

## Accessible

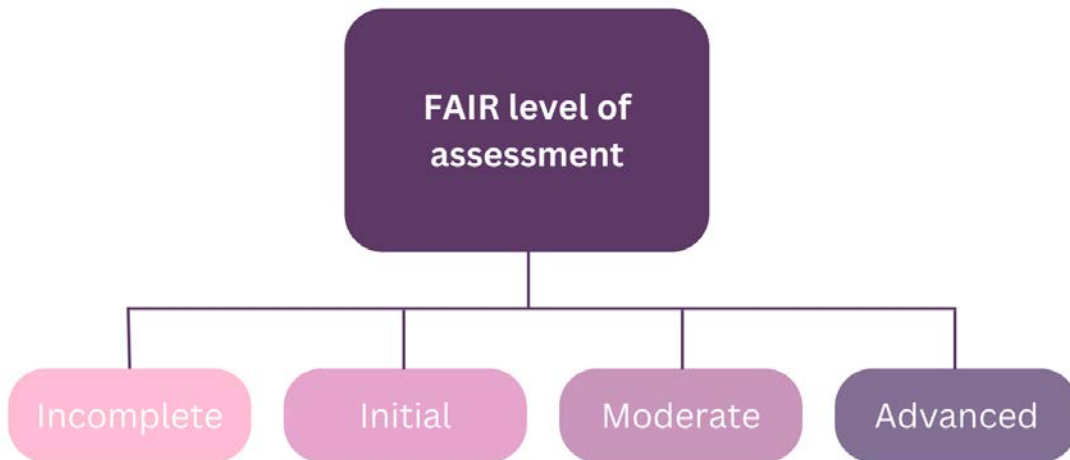
- Metadata contains access level and access conditions of the data.
- Data is accessible through a standardized communication protocol.
- Metadata is accessible through a standardized communication protocol.

## Interoperable

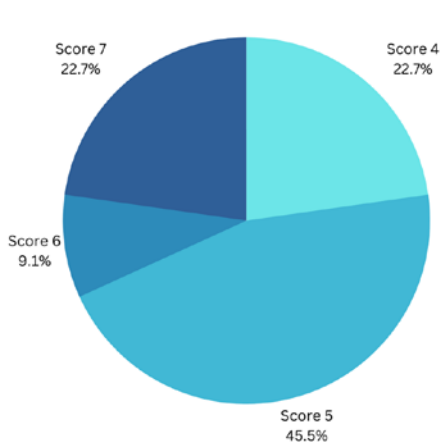
- Metadata is represented using a formal knowledge representation language.
- Metadata uses semantic resources.

## Reusable

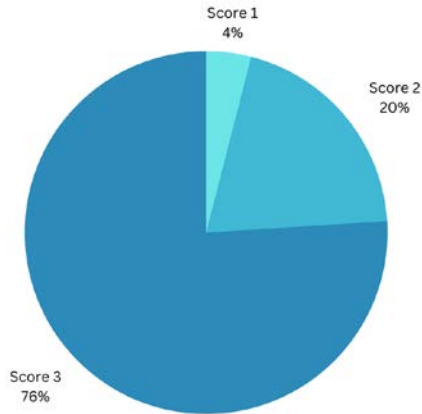
- Metadata specifies the content of the data.
- Metadata includes license information under which data can be reused.
- Metadata includes provenance information about data creation or generation.
- Metadata follows a standard recommended by the target research community of the data.
- Data is available in a file format recommended by the target research community.



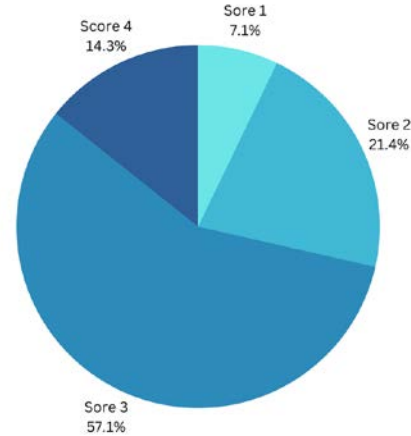
# RQ 3 To what extent does research data align with FAIR principles?



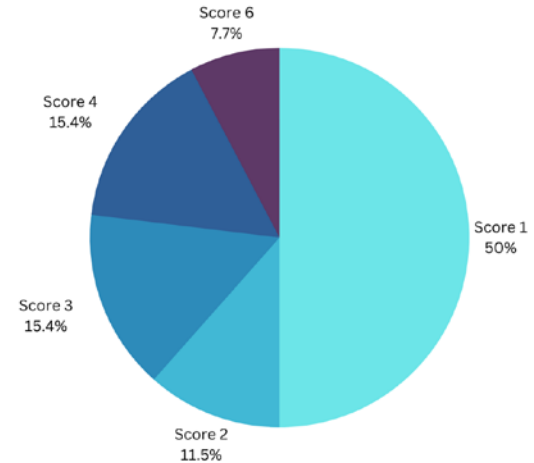
F



A



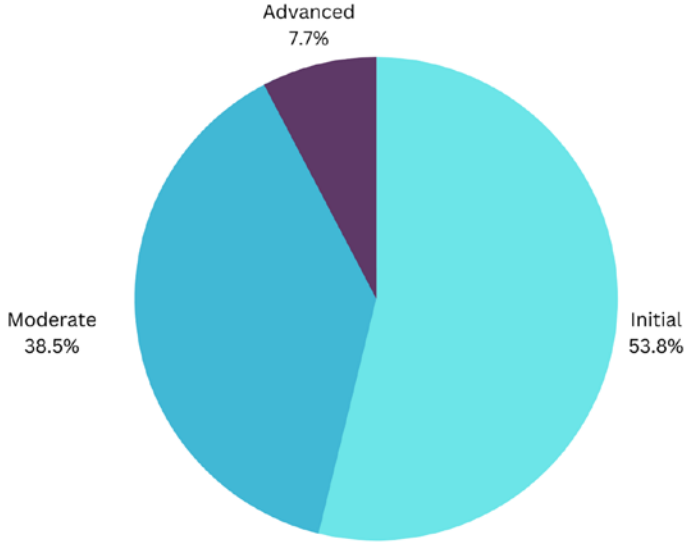
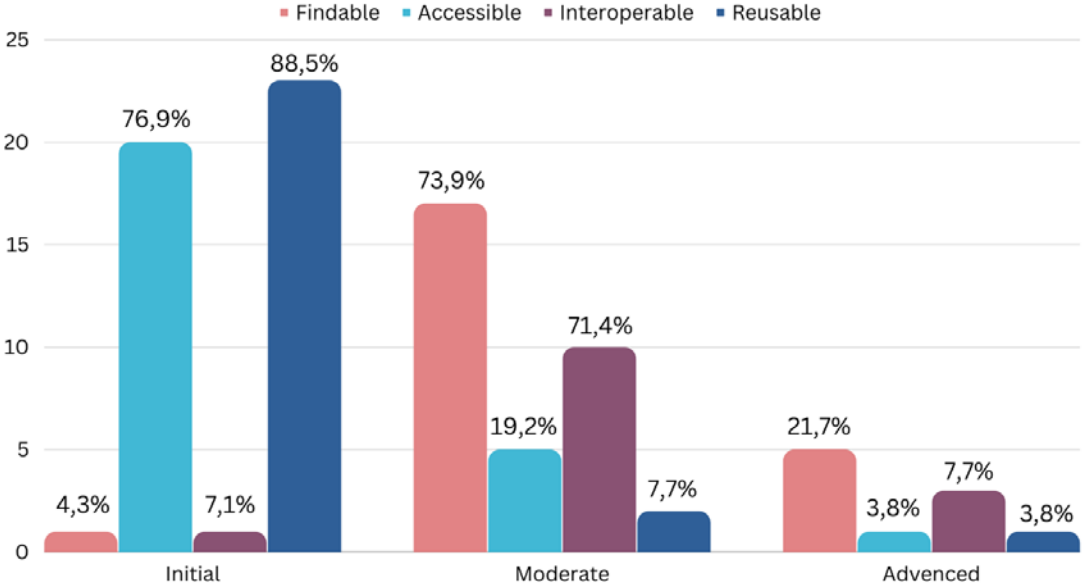
I



R



# RQ 3 To what extent does research data align with FAIR principles?





# Conclusion

## **RQ1: What types of research data are represented in humanities?**

- 123 objects are stored as data sets, but only 20 % are actual datasets
- Majority (48%) are mixed-type data sets which contains textual and numerical data and 28% are numerical datasets

## **RQ2: To what extent are data sets in the field of humanities represented in repositories by level of openness and under which license?**

- 92% of datasets are openly available but 34,6% are licensed with copyright all right reserved
- 11,5% of datasets have CCO licenses which serves as a public domain and is the most suitable for research data

## **RQ3: To what extent does research data align with FAIR principles?**

- all data sets are aligned with FAIR principles
- majority of datasets (53,8%) are on initial FAIR level, just one data set has highest FAIR score with 83%
- the most frequent percentage of FAIR level is 29% which is equivalent to initial FAIR level



# Limitations and future research

## Limitations

- difficulties in finding datasets from humanities
- large number of objects is stored in repositories as datasets but further analysis shows that majority are not datasets but research papers
- limited filter and search option in repositories
- in this research authors conducted analysis of 3 databases: DABAR, CROSSDA and OpenAIRE
- use of different FAIRness tool could provide different results
- automated testing with F-UJI tool depends on clear machine assessable criteria and some aspects (e.g. rich, plurality, accurate, relevant) specified in FAIR principles still require human mediation and interpretation
- F-UJI tests are designed in consideration of generic cross domain metadata standards such as dublin core, dcat, datacite, schema.org and conduct only cursory examination

## Future research

- research could be conducted on a larger scale with more databases and repositories



**Thank you for your attention!**

