



FAIR in Practice

17.5.2024

Mari Kleemola

Finnish Social Science Data Archive (FSD),
Tampere University

with special thanks to Tuomas J. Alaterä, FSD

My FAIR story today



Introduction



FAIR Principles, assessment vs assistance



FSD and FAIR in practice



FAIR enough? FAIR + time



Reflections



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

Introduction

Finnish Social Science Data Archive (FSD)

CoreTrustSeal certified data repository

National research data infrastructure

The Finnish Service Provider for CESSDA ERIC

Founded in 1999, separate unit of
Tampere University

Free-of-charge services for users

Main user groups: researchers, students
and teachers in universities



FSD's Aila data portal's key numbers

- ~6000 registered users
- ~2000 data sets
- ~6000 access applications (2023)
- ~6200 publications citing data

CESSDA

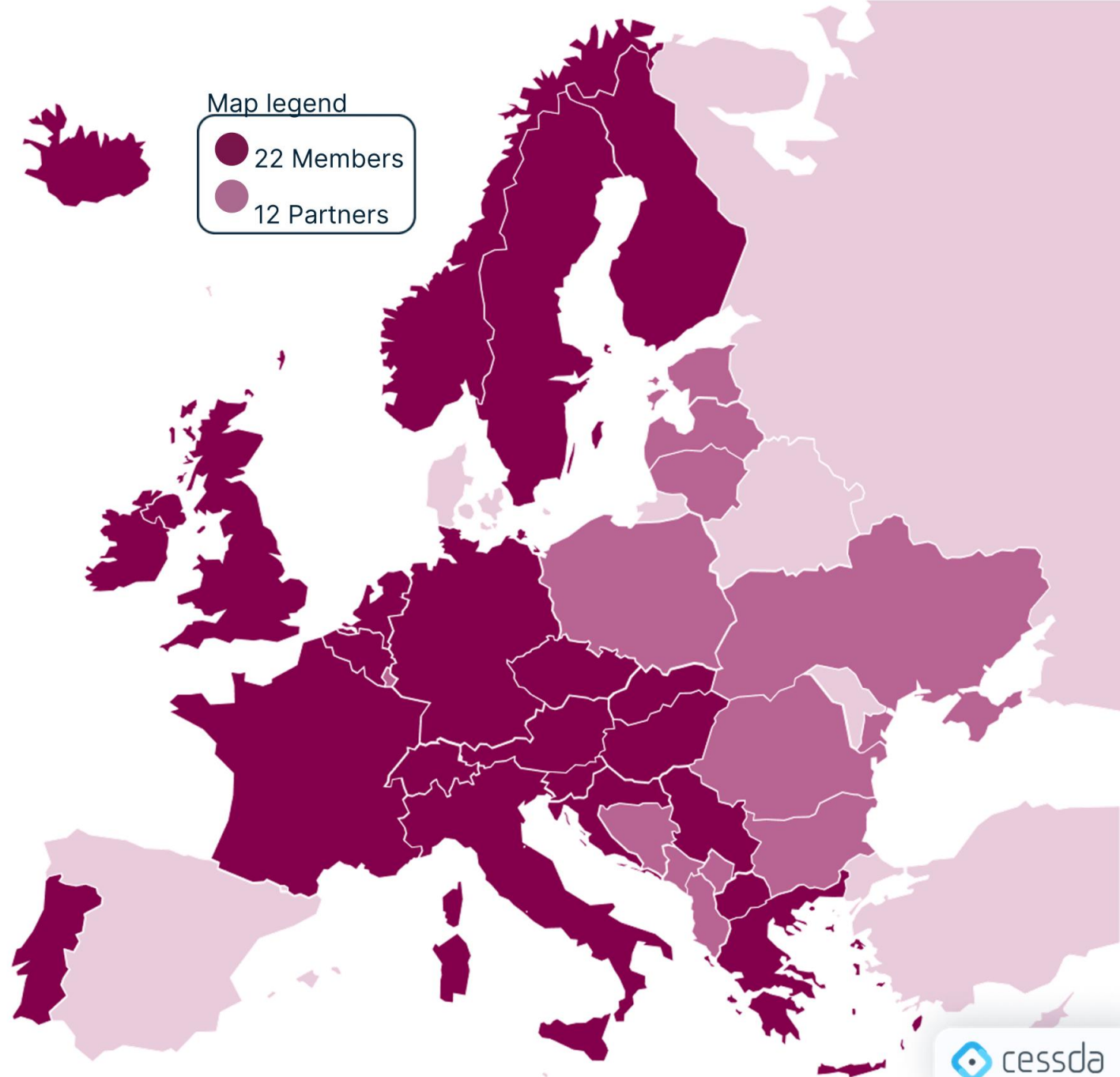
Is the Consortium of European Social Science Data Archives (Service Providers)

A true European research infrastructure:

1 out of 24 ERICs in Europe

Mission

To provide a sustainable research infrastructure that enables the research community to conduct high-quality research in the social sciences to contribute to effective solutions to the major challenges facing society today.

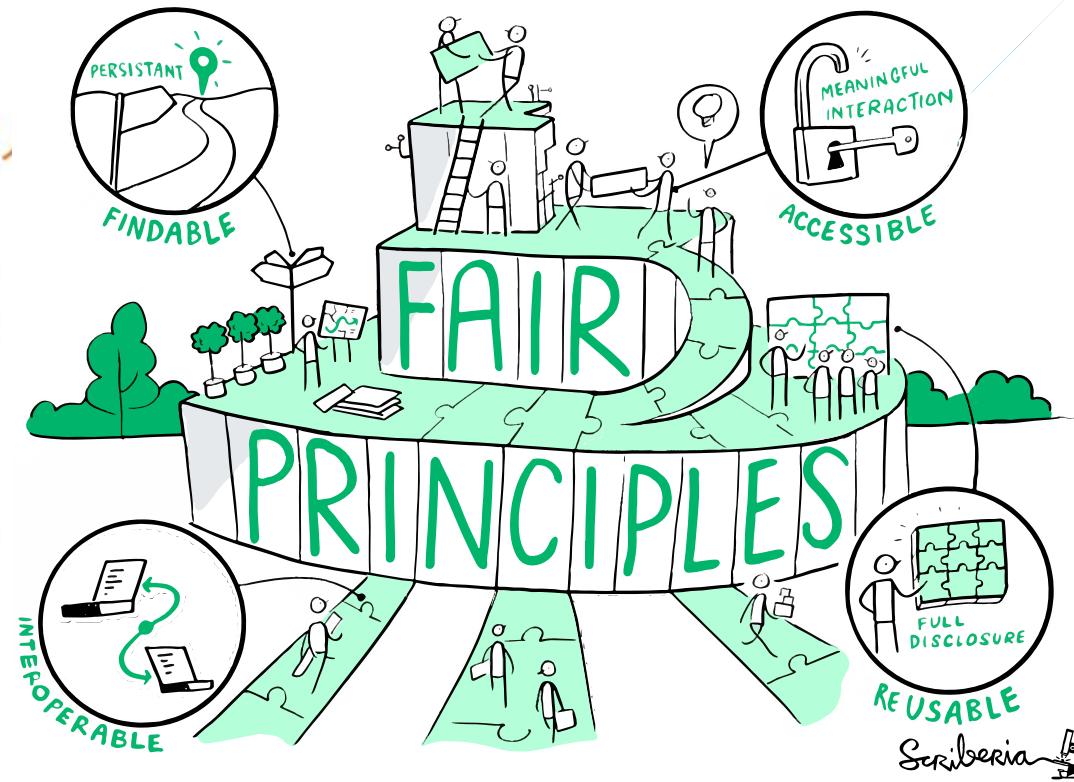
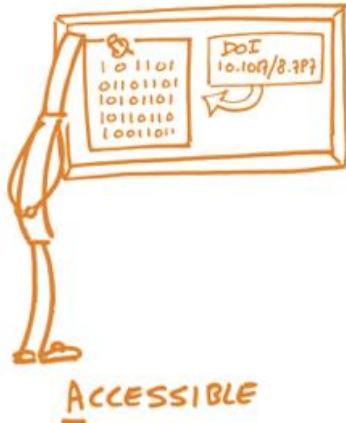




TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

FAIR Principles

FAIR DATA PRINCIPLES



FAIR PRINCIPLES

Findable Magnifying glass icon	Persistent Identifiers (PIDs) ID icon	Rich metadata Document icon with 'm'	Indexed data repositories Database icon with magnifying glass	PIDs in metadata ID icon with document
Accessible Hand pointing icon	Standard communications protocol Network icon	Open, free protocol Open lock icon	Authentication, where necessary User icon with shield	Metadata is always available Infinity symbol
Interoperable Gears icon	Vocabularies Tree diagram icon	Vocabularies are FAIR Gears and document icon	Linked metadata Network of documents icon	
Reusable Recycling symbol icon	Metadata have multiple attributes Document icon with arrows	Usage license Open lock icon with checkmark	Provenance Document icon with arrows	Community standards Group of people icon

F indable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

A ccesible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

I nteroperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

R eusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016).
<https://doi.org/10.1038/sdata.2016.18>

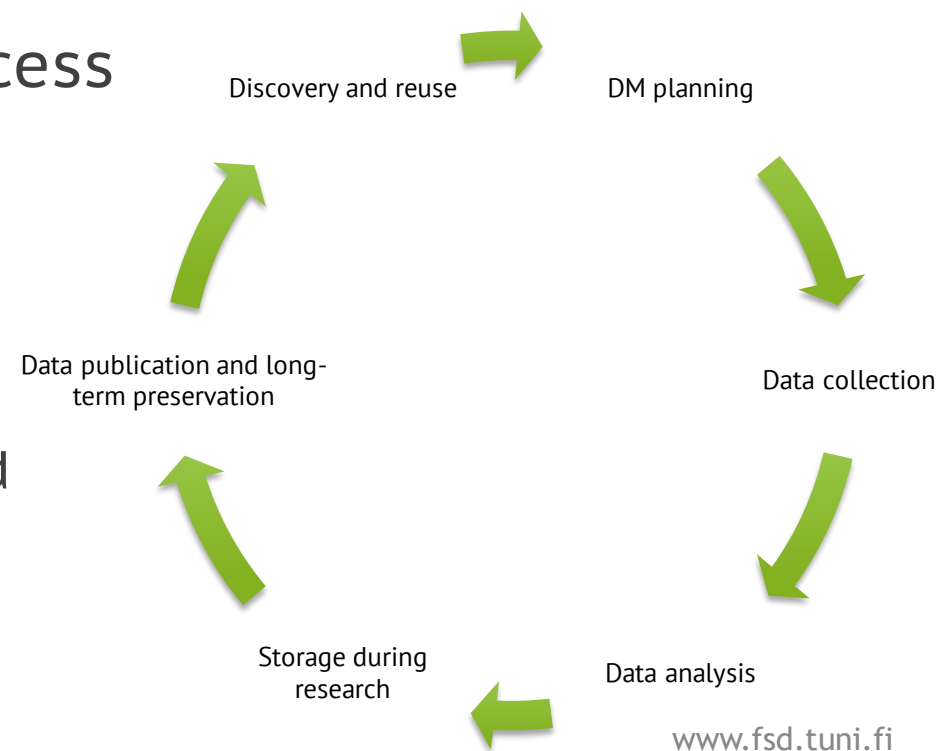
Sources:

- https://open-science-training-handbook.github.io/Open-Science-Training-Handbook_EN/
- <https://the-turing-way.netlify.app/reproducible-research/rdm/rdm-fair.html>
- <https://www.linkedin.com/pulse/why-your-data-must-fair-david-crosswell>
- <https://dmeg.CESSDA.eu/Data-Management-Expert-Guide/1.-Plan/FAIR-data>

Whole research process should be FAIR

- ▶ FAIR should not be separate from the research process
- ▶ Data management starts from the plan (DMP)
- ▶ “FAIR by design” = the entire research process planned so that FAIR becomes possible.
For example:

- ▶ metadata about the data and workflows are accumulated automatically
- ▶ the necessary identifiers tags are easily created
- ▶ the data file formats are commonly used
- ▶ barriers to data sharing are taken into account
- ▶ ”Make FAIR, don’t FAIRify”



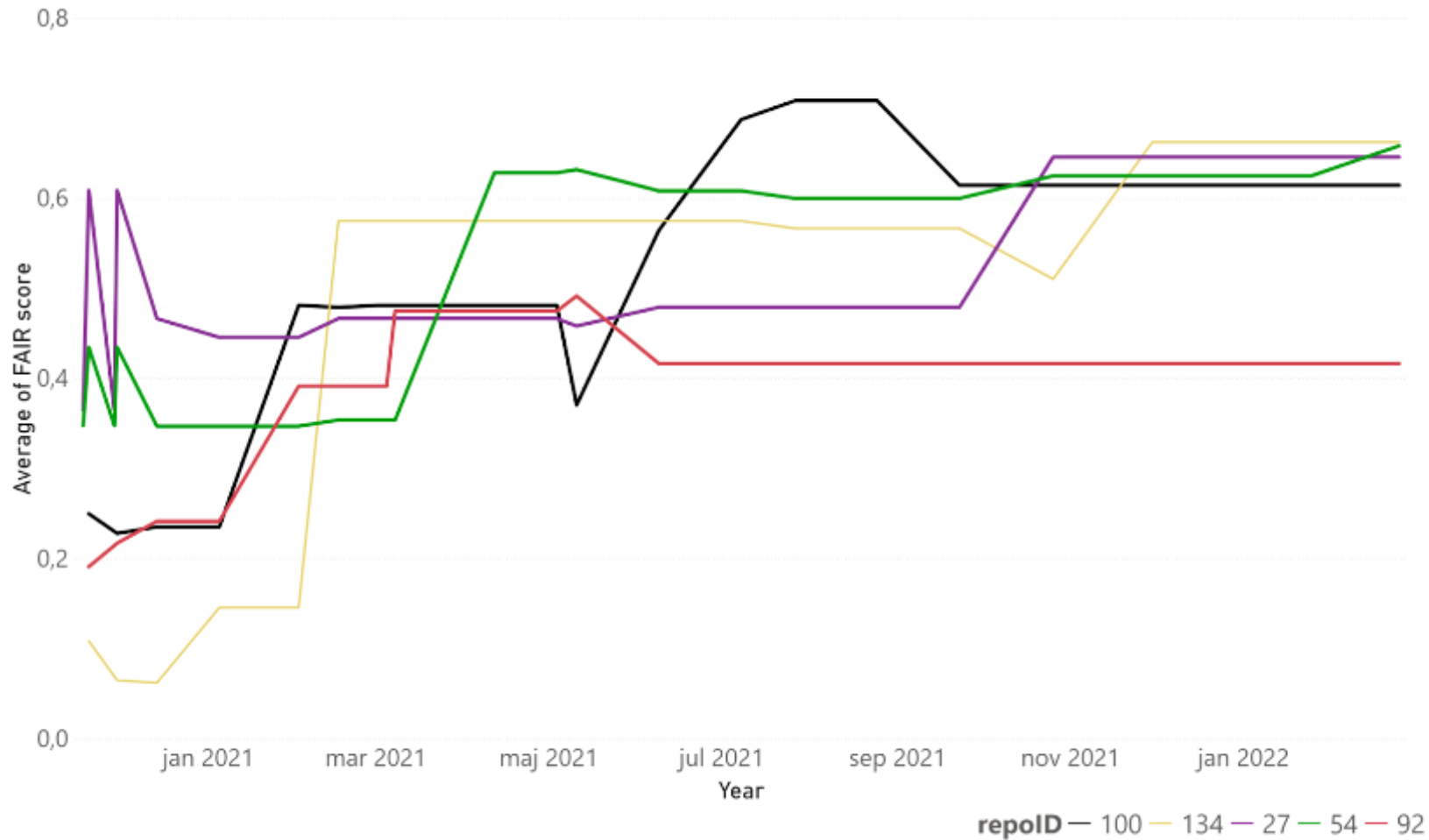


TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

Assessment vs. assistance

Assessment vs. assistance

- ▶ FAIR Principles are principles, not metrics
- ▶ FAIR has unified communities; conceptually, FAIR has been endorsed across disciplines
- ▶ Several activities and initiatives have focused on developing metrics and indicators, tests and tools ([FAIRrassist.org](https://fairrassist.org) lists 28 tools)
 - ▶ There is considerable flexibility in interpretations of the FAIR principles. Different tools give very different FAIR scores for the same Digital Object.
- ▶ Interpretation of numerical scores is not straightforward and requires knowledge. Binary judgements must be avoided.
- ▶ Uptake is best achieved via good guidance that provides appropriate remedies in case of test failures.
- ▶ OSTRails (<https://ostrails.eu/>) aims to transition from pass/fail tests towards assistance aimed at improving the results of an assessment where possible.



Evolution of FAIR score over time of five supported repositories in the EOSC-Nordic project.

Nordling, J., Mihai, H., Meerman, B., Alaterä, T., Kleemola, M., & Livenson, I. (2022). D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR. Zenodo. <https://doi.org/10.5281/zenodo.6880904>



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

Example: FSD and FAIR

FAIR at FSD (F)

▶ F – findability

- ▶ The datasets archived in FSD are described in detail in Finnish and English
 - ▶ The experts at FSD compile metadata based on the researcher's data documentation
- ▶ Metadata is published online (CC BY 4.0)
- ▶ We generate a persistent identifier (PID) for each dataset (currently URN, working on DOI)
 - ▶ PID for each data and metadata version coming
- ▶ Data sets can be found in FSD's [Aila](#), CESSDA [data catalogue](#), [Finna](#), [Etsin](#), and [Research.fi](#) –websites (among others)

FAIR at FSD (A)

▶ A – accessibility

- ▶ Metadata are available on Aila and on FSD's OAI-PMH-interface (Kuha2)
 - ▶ We use international DDI2.5 format for data descriptions
 - ▶ Metadata in DDI Codebook, OAI Dublin Core ja EAD3 -formats
- ▶ (Registered) users may download data from Aila according to the terms of use
- ▶ Aila supports HAKA identity federation identification
 - ▶ Users outside of HAKA-network may be granted a user account manually (for example people from universities outside of Finland)

FAIR at FSD (I)

▶ I – interoperability

- ▶ Datasets are described using an international DDI Codebook metadata format
 - ▶ Metadata includes variable level description
- ▶ National and international controlled vocabularies and classifications are used
 - ▶ YSO (from FINTO), and ELSST for keywords, classification of research disciplines, CESSDA topic classification, vocabularies from DDI (sampling procedure, collection method...)
- ▶ Citations to other metadata, data and publications (including PIDs)
- ▶ Commonly used data formats (including formats accepted to long-term preservation)
- ▶ Researcher id's, ORCID (we are working on organisational identifiers)
 - ▶ app. 290 data sets on Aila have at least one ORCID-id, all in all app. 550 ORCIDs in Aila

FAIR at FSD (R)

▶ R – reusability

- ▶ Every dataset has an availability class, that has been agreed on with the depositor
 - ▶ Terms of use are included in the metadata
- ▶ Datasets in class A are available freely (CC BY 4.0)
- ▶ Metadata on Aila available on CC BY 4.0 –license, reduced EAD- ja DC- data descriptions on CC0-license

- (A) freely available under the CC BY 4.0 license
- (B) for research, teaching, and studying
- (C) for research only (inc. Master's, doctoral, and University of Applied Sciences Master's thesis)
- (D) only with permission from the data depositor

How to: INTEROPERABLE (I) different expressions

Study title

Finnish Voter Barometer 1973

Dataset ID number

FSD1000

Persistent identifier

[urn:nbn:fi:fsd:T-FSD1000](https://nbn.fi/urn:fi:fsd:T-FSD1000)

Data type

Quantitative

Authors

Gallup Finland

University of Helsinki

<titlStmt>

<titl>Puolueiden ajankohtaistutkimus

1973</titl>

<parTitl xml:lang="en">Finnish Voter Barometer

1973</parTitl>

<IDNo agency="FSD">1000</IDNo>

<IDNo agency="URN">urn:nbn:fi:fsd:T-

FSD1000</IDNo>

</titlStmt>

<rspStmt>

<AuthEnty>Gallup Finland </AuthEnty>

<AuthEnty>University of Helsinki </AuthEnty>

</rspStmt>

<prodStmt>

<producer>Gallup Finland </producer>

<producer>University of Helsinki</producer>

</prodStmt>

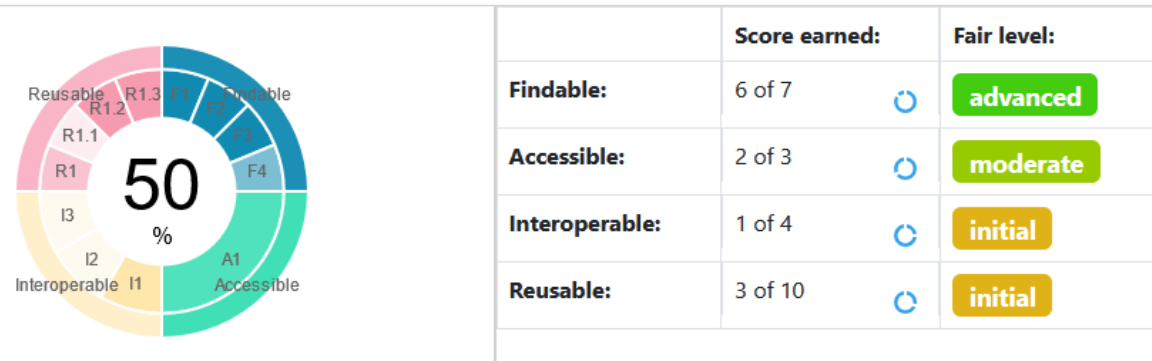
<distStmt>

1. Human readable
2. DDI XML
3. JSON-LD

```
{
  "@context": {
    "@vocab": "https://schema.org/"
  },
  "@type": "Dataset",
  "creator": [
    {
      "@type": "Organization",
      "name": "Gallup Finland"
    },
    {
      "@type": "Organization",
      "sameAs": "https://ror.org/040af2s02",
      "name": "University of Helsinki"
    }
  ],
  "@id": " http://urn.fi/urn:nbn:fi:fsd:T-FSD1000 ",
  "name": [
    {
      "@value": " Finnish Voter Barometer 1973 ",
      "@language": "en"
    },
    {
      "@value": " Puolueiden ajankohtaistutkimus 1973 ",
      "@language": "fi"
    }
  ],
}
```

Force PID/URL:	https://services.fsd.tuni.fi/catalogue/FSD3770?tab=description&study_language=fi
Cite support:	enabled
Force Version:	metrics_v0.5
Force Specification:	https://doi.org/10.5281/zenodo.6461229
Force version:	3.1.0
Download assessment results:	{JSON}
Download and share assessment results:	

Summary:



FAIR Evaluator assessment of FSD3770, May 2024

test ID

Summary:

Description: FAIR Metrics Evaluation: test; Tested Identifier: https://services.fsd.tuni.fi/catalogue/FSD3770?tab=description&study_language=fi; generated by <https://orcid.org/mari.keemola@tuni.fi>

Resource: https://services.fsd.tuni.fi/catalogue/FSD3770?tab=description&study_language=fi

Collection: 6

Observations: Ran 22 tests (14 succeeded, 8 failed).

JSON response: http://fairdata.services:3333/FAIR_Evaluator/evaluations/13350.json

Tests passing and failing

- GEN2 UNIQUE IDENTIFIER +
- GEN2 METADATA IDENTIFIER PERSISTENCE +
- GEN2 DATA IDENTIFIER PERSISTENCE +
- GEN2 STRUCTURED METADATA +
- GEN2 GROUNDED METADATA +
- GEN2 DATA IDENTIFIER IN METADATA +

F-UJI assessment of FSD3770, May 2024

Over the years, FSD's highest score has been (if my memory serves right) 77%



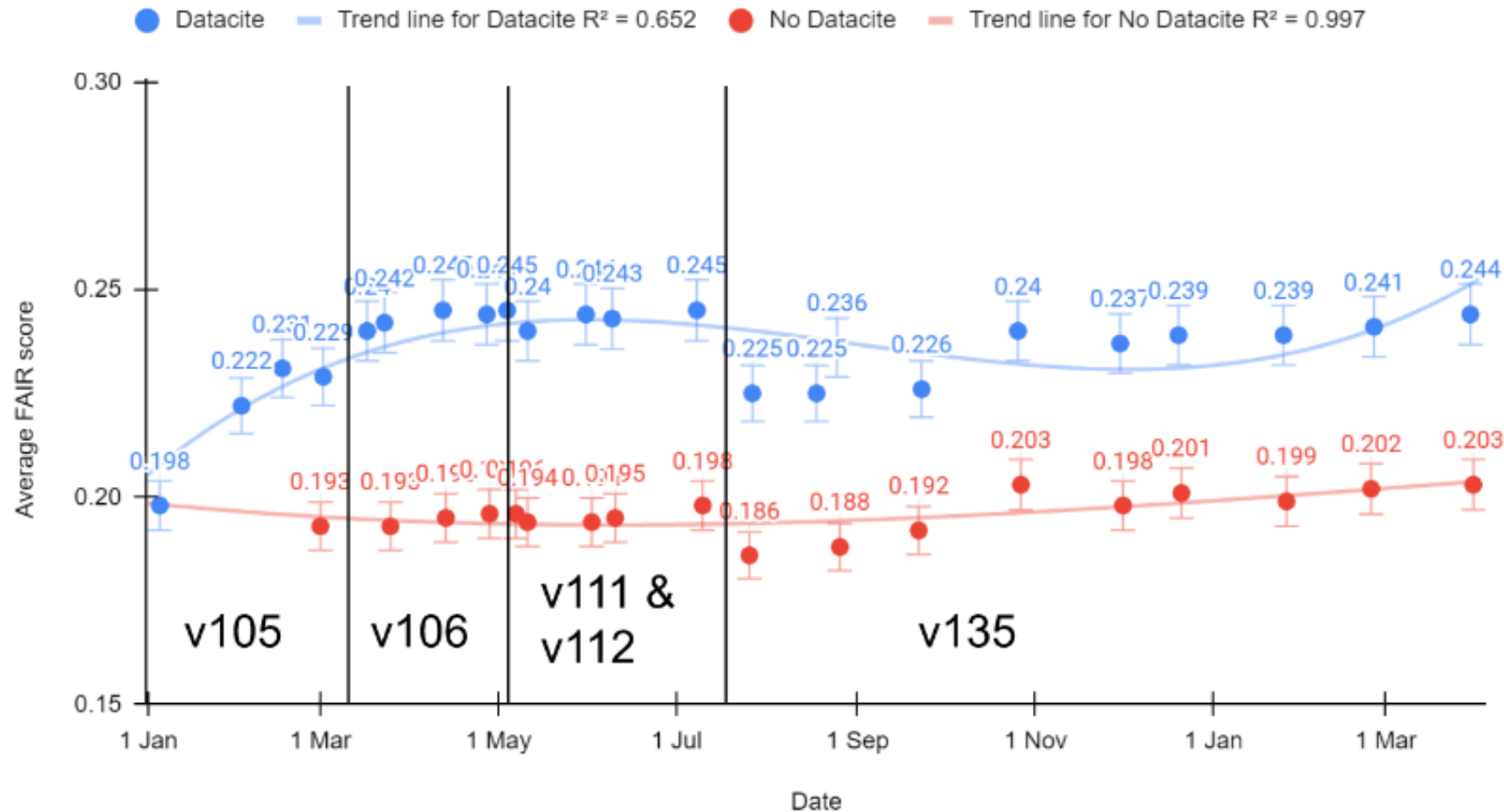
TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

FAIR enough?

What is FAIR enough?

- ▶ The research community's needs determine what to aim for!
- ▶ Findability is typically easier than interoperability – think about your use case when setting the goal!
- ▶ Regardless of FAIR, exercise good data management practices
- ▶ Think about FAIR as not a one goal, but a continuous process: "A bit FAIRer every day!"
- ▶ Discipline-specific metadata is important. Too often metadata include only administrative and organisational information without any useful specific descriptors.
 - ▶ See e.g. Mark A. Musen. *Without appropriate metadata, data-sharing mandates are pointless*. Nature 609, 222 (2022). <https://doi.org/10.1038/d41586-022-02820-7>
- ▶ A digital object cannot really be made FAIR or evaluated for FAIRness in isolation from its context, it is dependent on the (data) ecosystem.
 - ▶ For example, the persistence of an identifier is determined by the commitment of the organisation that assigns and manages it.

Development of average FAIR score over time



Nordling, J., Mihai, H., Meerman, B., Alaterä, T., Kleemola, M., & Livenson, I. (2022). D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR. Zenodo. <https://doi.org/10.5281/zenodo.6880904>

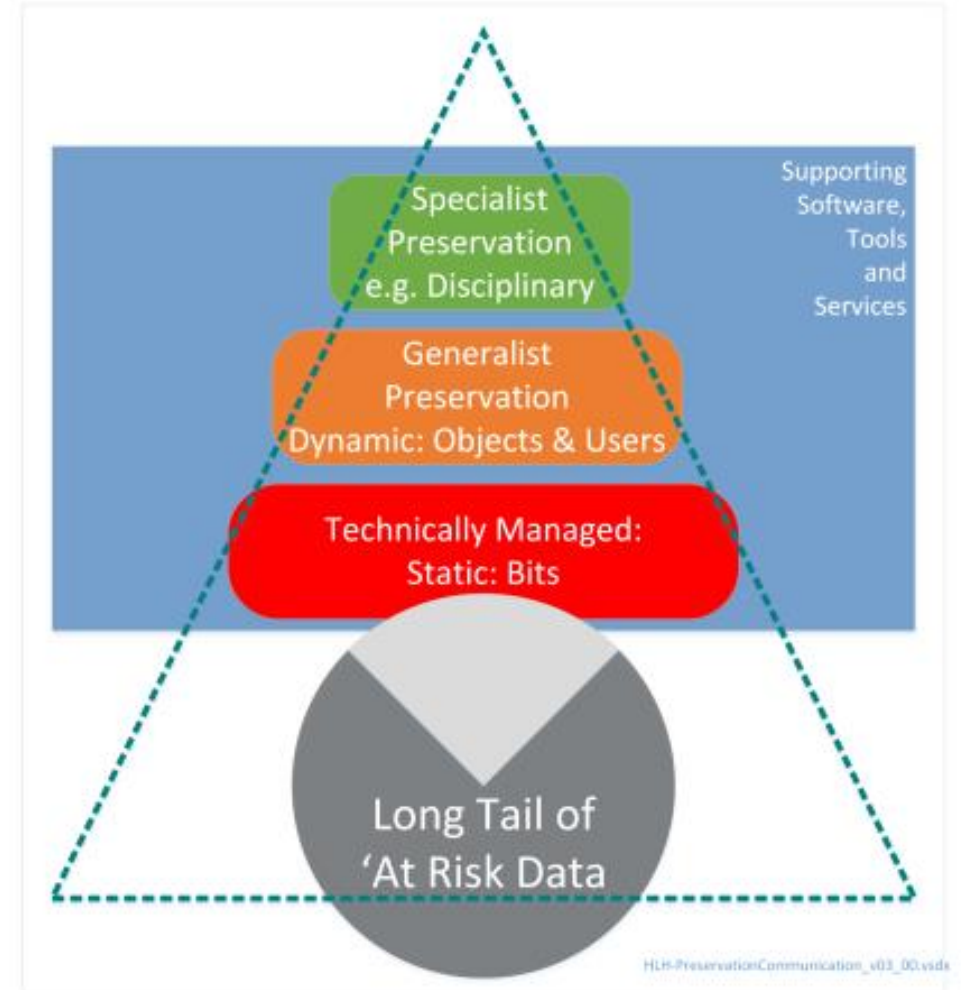
- ▶ Based on EOSC-Nordic, a FAIR score over 50% could be said to be good, **but:**
 - ▶ Depends on your use case and community needs
 - ▶ Tests, tools and repositories have improved since 2022!
- ▶ Check also the The Finnish national Open Science [Guide for Making Your Data FAIR](#)



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

FAIR + time

- ▶ FAIR principles don't take into account the inevitable changes to the data environment or the users' needs
- ▶ Research data will not become or stay FAIR by magic
- ▶ We need skilled people, transparent processes, interoperable technologies and collaboration
- ▶ Trustworthy Digital Repositories (TDRs) play a key role in enabling data to become, and to remain FAIR over time.
- ▶ L'Hours, H., Kleemola, M., von Stein, I., van Horik, R., Herterich, P., Davidson, J., Rouchon, O., Mokrane, M., & Huber, R. (2022). FAIR + Time: Preservation for a Designated Community (02.00). Zenodo. <https://doi.org/10.5281/zenodo.5797776>



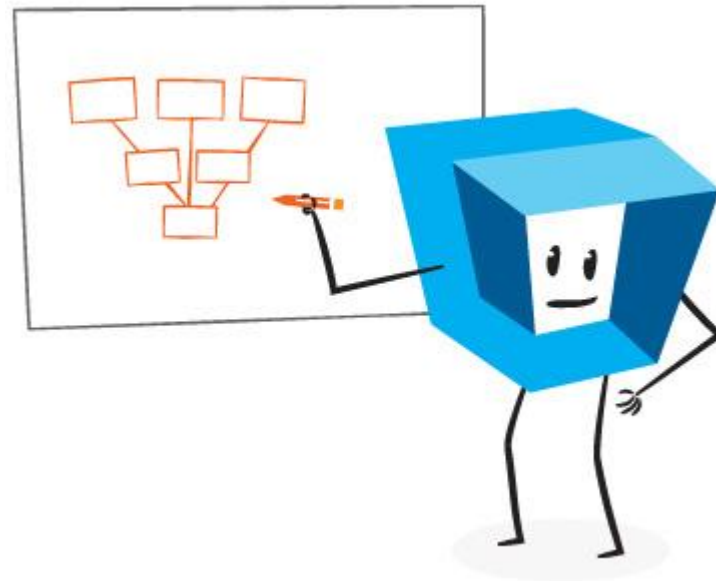


TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

Some reflections

Thank you!

Mari Kleemola, mari.kleemola@tuni.fi
<https://orcid.org/0000-0001-8855-5075>



Except the logos and dude above, content in this presentation is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).