TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

Co-funded by the Erasmus+ Programme of the European Union

Open Science, data archiving and FSD, important legal aspects in data archiving

16.4.2024

Helena Laaksonen
Director at Finnish Social Science Data Archive (FSD), Tampere University

Contents

- Finnish Social Science Data Archive (FSD) & CESSDA ERIC
- Data Archiving, Harmonisation & Open Science
- SSH & Open Science
  (+ legal aspects concerning research data)
- Data Archiving in Practice & Examples
- F.A.I.R. & T.R.U.S.T.

Finnish Social Science Data Archive (FSD)

- Separate unit of the Tampere University since 1999
- National Research Infrastructure
- Finland's national service provider for CESSDA
- CTS-sertified (CoreTrustSeal) = trustworthy data repository
- Services free of charge
- Expert organisation: data and IT specialists (~20 persons)

CORE TRUST SEAL

Findable
Accessible
Interoperable
Reusable

cessda

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

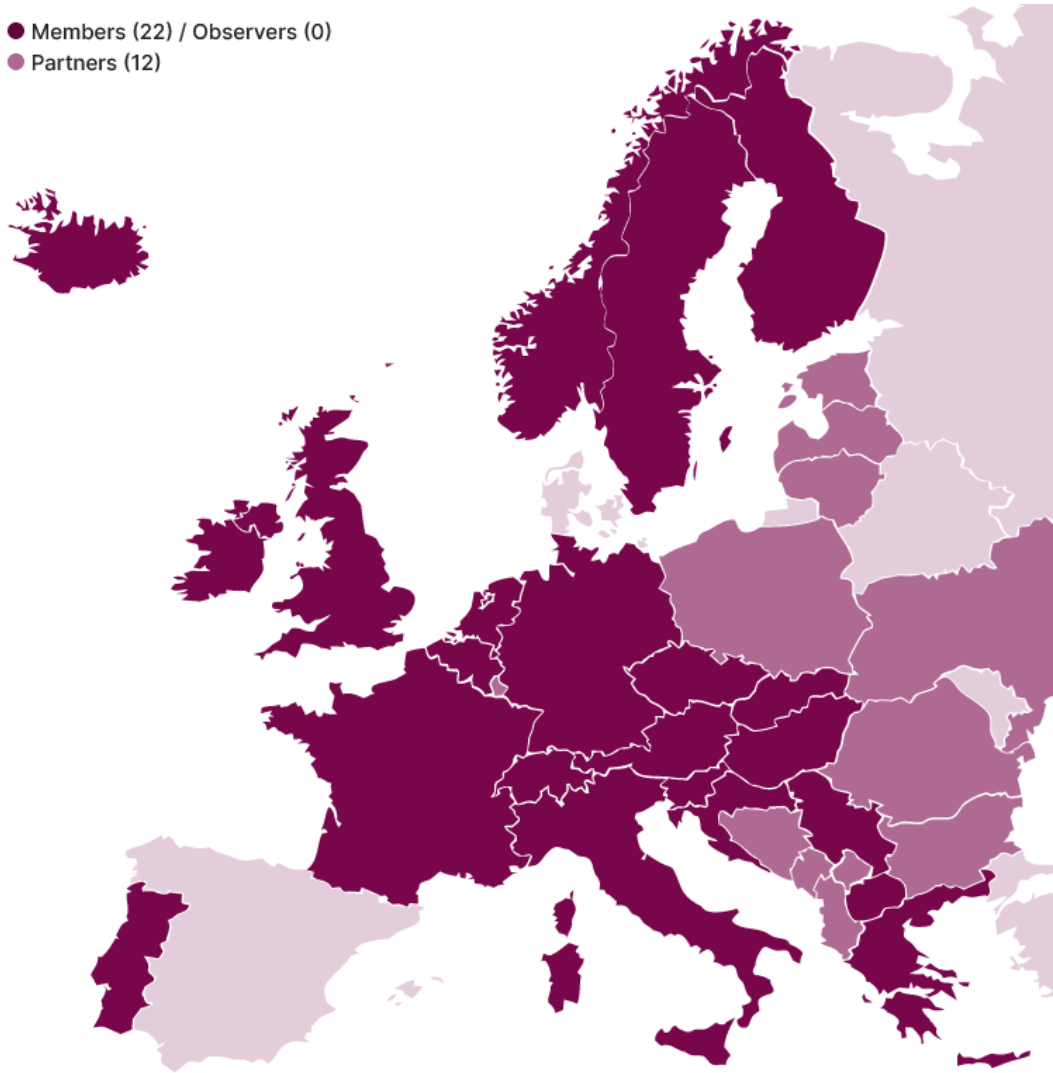
Co-funded by the Erasmus+ Programme of the European Union

Consortium of European Social Science Data Archives (CESSDA ERIC)

cessda

- CESSDA provides large-scale, integrated and sustainable data services to the social sciences.
- Service Providers in member countries provide national services.
- Aim to promote the results of social science research and support national and international research and cooperation.

https://www.cessda.eu/



Sevices

FSD

- Aila
  - Data Ingest/Deposit
  - Data dissemination
- Data Management Guidelines
  - Management Instructions for data life cycle -> improved quality
- Information service on the archived research data
- ...
- https://www.fsd.tuni.fi/en/services/

CESSDA ERIC

- Data Catalogue (CDC)
  - Search and browse data
  - Share metadata: for reusers in CDC & for harvesting into other catalogues
- Tools for the Service Providers:
  - Vocabulary service
  - ELSST Thesaurus
  - ...
- https://www.cessda.eu/Tools

Data Archiving, Harmonisation & Open Science

- Shared Metadata Models/Standards e.g.
  - DDI: https://ddialliance.org/
  - CESSDA Metadata Model
- Vocabularies & Thesauri
- PIDs: URN, DOI, ORCID, ROR … (+ data PID resolvable to a metadata record that is kept long term)
- Versioning (data/metadata)
- Harmonised research data: questionnaires, background variables, (hierarchical) concepts (classification of occupations, education level…)
- Objective: data understandable to humans and machines: enhanced findability, reusability, comparability, interoperability
- Need to be understandable and reusable to the main target audience

SSH data & Open Science & Legal Aspects

- Data gathered from human subjects by interviewing, measuring, observing, or the research subjects' own creation (written accounts, photographs, recordings...)
- Research Subjects should not be identifiable (GDPR & national law)
  - anonymisation
  - Pseudonymisation
- Data ownership & licensing: data belongs to the depositors or meets the threshold of originality
- Restricted access supports openness =>
  data is as open as possible, as restricted as necessary
- Read more in Data Management Guidelines

Archiving process in practice

- Researcher plans the data collection and informs the respondents about archiving
  - Dataset's availability class affects the informing
  - Researcher documents every phase of the data collection and modification
- Researcher delivers anonymised "raw" data, and documentation that should be preserved
  - Data matrices, interview transcriptions…
  - Questionnaires, interview themes… (instrument)
  - Description of data and data collection (check on Aila which information is needed for DDI metadata)
- After FSD has processed the dataset and the researcher has confirmed it, metadata is published on Aila and data is ready for dissemination

Example I: Aila Data Service

Aila

- Search and browse data
- Licensing: A, B, C, D
- Quantitative & Qualitative
- Metadata goes deep into the content
  - Dataset description
  - Variable description
  - Metadata is published (CC BY 4.0)
- Downloading: A CC by 4.0, restricted B – D.
  - Haka for (most) Finnish users
  - FSD's identifications service for foreign customers

Example II: CDC & Topical/Multidisciplinary catalogues

- CESSDA Data Catalogue
  - Covid-19 datasets in Aila
  - Covid-19 datasets in CDC
- COVID-19 Data Catalogue
  - CDC data in COVID-19 data catalogue
- Google Datasetsearch (searching Kansalaispulssi)

What is needed?
Harmonised metadata
+ OAI-PMH interfaces
+ expert personnel
(IT & SSH content)

FAIR Principles

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

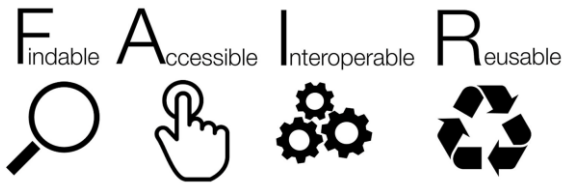
Co-funded by the Erasmus+ Programme of the European Union

What is FAIR?

- Findability
- Accessibility
- Interoperability
- Reusability

- Support good data management and research quality, reproducibility and impact.
- As an idea, FAIR is ~10 years old and have developed from an article (Nature 2016) into an "ideal/ideology".

- Intended for **all data**, both quantitative and qualitative.
- Apply **also to metadata**, which is used to describe the data and make it discoverable.
- Also applies to software, tools, code, algorithms and workflows that generate data*
- The goal is always **machine actionability** and interoperability of the data.
- FAIR data is understandable to humans and, to a certain extent, can be processed programmatically.
- Discoverable, **accessible,** and interoperable data that is **reusable** facilitates reproducibility.



Why FAIR?

FAIR ≠ Open (data)

- For the (digital) data to be usable from the beginning and remain usable in the long term.
- To guarantee the scientific reproducibility, the discovery of new knowledge and the transparency of the conducted research by opening access to data and/or the metadata describing them.
- FAIR is **not a goal or standard** for open science or data sharing, but a principle-level framework of how research data, methods, code and other related material should be organised so that their understandability and reuse and programmatical processing are possible.
  - Provides a framework for data management: what to consider, when and what services are needed.
- Research funders or publishers may require that the FAIR principles (or openness of data) are implemented.
- Research (data) cannot be FAIR without research community / domain-relevant perspective or approval.
  - Data will not "open up" if the designated community does not understand it.
  - Combining data/methods/code also requires discipline-specific knowledge.
  - Therefore discipline-specific repositories are also needed (e.g., FSD).
- But the purpose of FAIR is to promote interdisciplinary solutions.

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

Co-funded by the Erasmus+ Programme of the European Union

Examples of implementing FAIR

- How to FAIR (Danish National Forum for Research Data Management)
  - https://howtofair.dk/
  - Six central topics for FAIR accompanied with videos:
    1. Documentation
    2. File formats
    3. Metadata
    4. Access to data
    5. Persistent identifiers
    6. Data licenses
  - Approx. two hours of learning material

TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

Co-funded by the
Erasmus+ Programme
of the European Union

TRUST Principles

- **T**ransparency
- **R**esponsibility
- **U**ser Focus
- **S**ustainability
- **T**echnology

A common framework to facilitate discussion and implementation of best practice in digital preservation.

See Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* **7**, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

Co-funded by the Erasmus+ Programme of the European Union

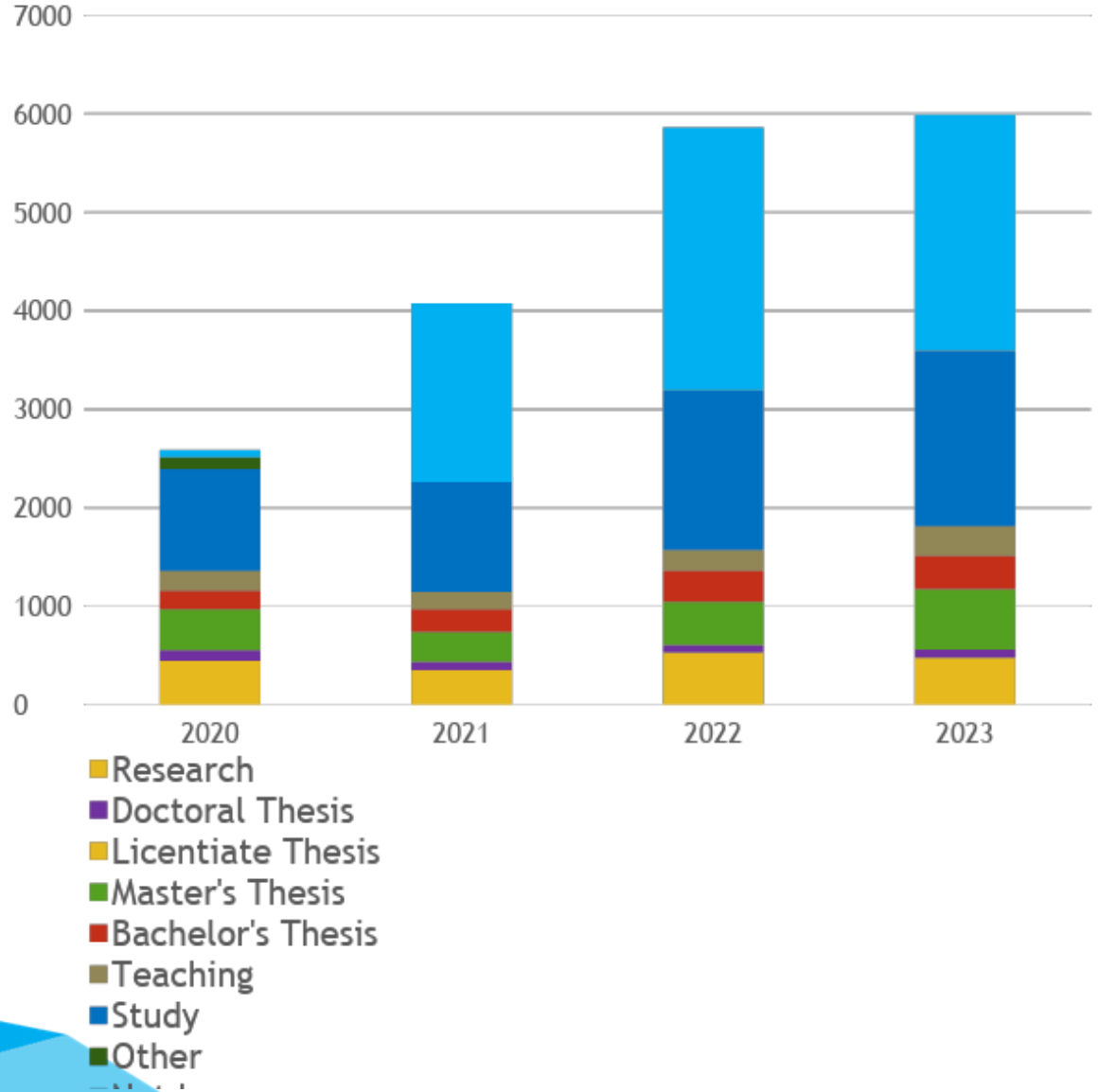
Statistics – Responsibly Open Data in Practice

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

Co-funded by the Erasmus+ Programme of the European Union

Access applications in Aila



Purpose of use	2020	2021	2022	2023
Not known (A-data)	76	1813	2672	2395
Other	119	0	1	1
Study	1035	1117	1622	1784
Teaching	202	175	205	299
Bachelor’s Thesis	188	231	316	339
Master’s Thesis	415	303	441	614
Licentiate Thesis	1	2	0	2
Doctoral Thesis	108	83	79	80
Research	444	351	527	477
Sum	**2588**	**4075**	**5863**	**5991**

TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

Co-funded by the
Erasmus+ Programme
of the European Union

Added after the presentation

FSD & CSC

- CSC provides bit-level long-term preservation for the archived data; FSD maintains usability of the archived data
- FSD's metadata is pushed into Etsin (the general national data catalogue). An example. (~35 % of the metadata in Etsin)
- FSD's metadata is harvested into Finna catalogue as well.
- Neither of the catalogues provides the functionalities Aila does. The metadata is limited on a very general level.

Functionalities in Etsin:

- provides Citations for the datasets in several formats
- Possible to browse metadata in FSD by depositor organization.
- Please, see also the notes page of this slide.

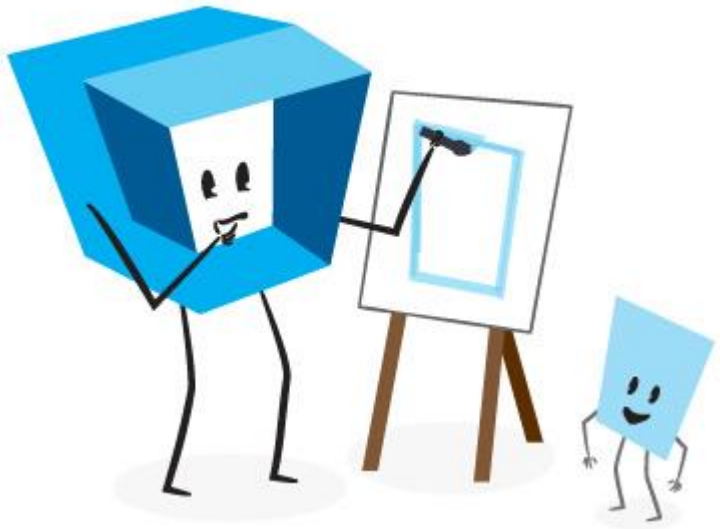
TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

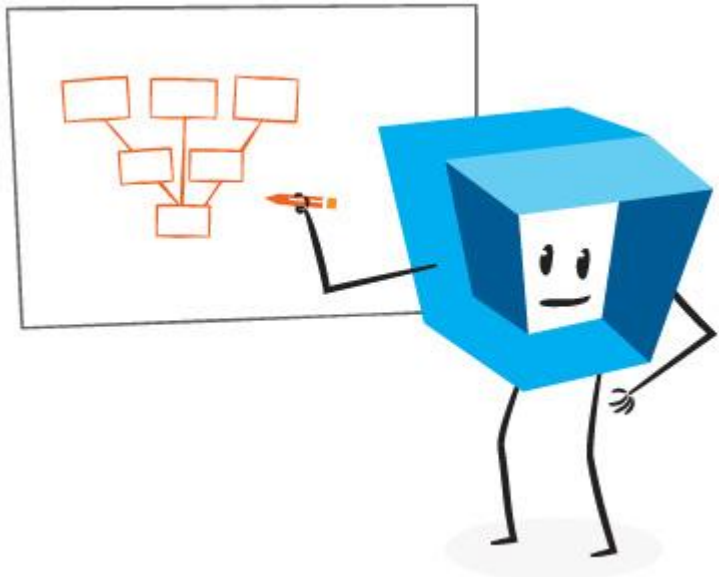
OPEN ASIA
UNIFYING SCIENCE, EMPOWERING INNOVATION

Co-funded by the Erasmus+ Programme of the European Union

Thank you!

https://www.fsd.tuni.fi/en/





Slides #8, 12 – 14, 18 by Henna Juusola & Tuomas J. Alaterä

User-services.fsd@tuni.fi · +358 29 452 0411
Helena Laaksonen
helena.laaksonen@tuni.fi
ORCID: 0000-0002-1312-1958