Challenges and successes of CESSDA ERIC

Mari Kleemola

Finnish Social Science Data Archive (FSD) and CESSDA ERIC

CLOSER conference: Preparing for the future II: international approaches to challenges facing the longitudinal population studies.

Session C3 Data discoverability: issues in interoperability

London, 16 January 2020





YHTEISKUNTATIETEELLINEN TIETOARKISTO FINNISH SOCIAL SCIENCE DATA ARCHIVE

Tampere University



One of the goals of CESSDA ERIC is to be leading in data discovery. This presentation will give an insight to how CESSDA has been building a coherent infrastructure in a multilingual environment and with Service Providers with differing capacities. I will also shortly describe the approach to (meta)data interoperability taken by the Social Sciences and Humanities Open Cloud (SSHOC) project that contributes to the Open Science agenda and realising the EOSC.

CESSDA ERIC

Consortium of European Social Science Data Archives — European Research Infrastructure Consortium

Cross-European **resource discovery**

Improved **quality** of data and metadata

A wider selection of **comparable data**

Certification of data archiving organisations

Professional **training** for data archivists and scientific community

Improved mechanisms for **data access and analysis**

Strong involvement of organisations **outside Europe**



— CESSDA timeline



CESSDA on ESFRI Landmarks 2011, Steering Committee Agreed on 2008, ESFRI Roadmap 1st Update **Documents** cessda ESTR (essda)
 AND A Consortium of European Social Sciences Data Archives June 18th 2013, CESSDA AS as a 1976, Amsterdam 2006, ESFRI Roadmap Legal Institution, First General ZA, UKDA, DDA, Steinmetz, 2010. CESSDA ERIC BASS, NSD, ADPSS **Assembly Meeting** June 14th 2017, **Formal Process CESSDA-ERIC** 2003 DDI Alliance **1976 IFDO 1974 IASSIST** 2013, MoU 13 countries 2008-2010, EC Preparatory 1992, Limerick **Phase Project** Formalized 2014, ESFRI: Prioritization of Support to ESFRI Projects Dec. 5th 2013, Formal Opening for Implementation 2011, MoU 12 countries 2009, Meeting between Germany and Norway 2011, ESFRI Roadmap 2nd Update 2014, CESSDA AS Main Office staffed

BTG BRO



CESSDA infrastructure building

Challenges

- Service Providers with differing capacities, capabilities and maturity
- Multilingual environment
- Limited resources
- Infrastructure is never complete!

Solutions

- Common standards and practices (CMM, CVs, policies)
- FP7 and H2020 funding, CESSDA Work Plan projects
- Obligations and support activities
- CESSDA community has long history

CESSDA Tools & services











... and more, see <u>https://www.cessda.eu/Tools-Services</u>







Commonality Type Commonality Type

Describes the degree of similarity between two items or schemes (collections of items).

Version: 1.0 Download

Consortium of European Social Science Data Archives Consortium of European Social Science Data Archives CESSDA Data Catalogue Topic Topic Cessda Cess

https://datacatalogue. <u>cessda.eu/</u>

| 14105 results found | | | | Advanced search | Reset filters | Clear search | |
|--------------------------------------|---|--|-------|-----------------|---------------|--------------|---|
| ▶ Торіс | Ø | Results per page 30 T | | Sort by | Relevance | | ¥ |
| Collection years | 0 | < | 1 2 3 | 4 | | | > |
| ▶ Country | 0 | NorLAG easy-to-use, 2002–2008, panel Statistics Norway (SSB) NorLAG easy-to-use is a user-friendly data file from the Norwegian panel study on Life course, Ageing and Generation (NorLAG). NorLAG is an interdisciplinary and longitudinal study that aims to provide increased knowledge about variation and changes in the second half of life (40 years and older) by studying behaviour and life transitions in four central life areas: (1) work and retirement, (2) family and generations, (3) mental health and well-being, and (4) health and care. In all areas, ge | | | | | |
| ▶ Publisher | Ø | | | | | | |
| Language of data files | 0 | | | | | | |

Severe Acute Malnutrition and Mortality among Children 6-59 Months in the DRC, 1992

Schwinger, Catherine (University of Bergen)

These data are part of a large cohort study. For this analysis, the aim was to describe the mortality risk of children in the community who had severe acute malnutrition (SAM) defined by either a mid-upper arm circumference (MUAC)

Language: English

Danish Gallup Omnibus Data 1966, Omnibus no. 11 TNS Gallup

Danish Gallup Omnibus Data 1966 comprises 10 omnibus surveys carried out by the Danish Gallup Institute in 1966. This data material contains information from omnibus no. 11. # #Each survey contains a number of demographic variables common to all the surveys of that year, as well as a number of questions dealing with

- Need common shared understanding of what are the digital objects, what are the required characteristics of data and metadata, and how these change over time.
- Many CESSDA SPs and other data repositories are moving from mediating discovery and download to mediating use of data (for example providing analysis environments).
- In general, both CESSDA ERIC and the SPs need to become more mature to meet the growing complexity.
- In practice, succesful reuse of data does not always require complete understanding of all interoperability aspects!



About SSHOC Marketplace Training SSHOC & EOSC SSH Landscape News & Events Services Main impacts The Social Sciences and Availability of an EU-wide, EU-wide availability of high 日 quality "cloud ready" SSH Humanities are seamlessly easy-to-use SSH Open ż Marketplace, where tools tools and high quality SSH integrated in the European **Open Science Cloud** and data are openly data accessible State of the art Research Maximising reuse through EU-wide availability of trusted Infrastructure in several **Open Science and FAIR** and secure access •••• principles (standards, pilot domains advanced mechanisms for SSH data, through dedicated SSH common catalogue, access conforming to EU legal data pilots cluster projects control, semantic requirements techniques, training)



SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-04-2018, grant agreement #823782.

SSHOC Task 3.5 Interoperability Hub

- D3.1 Report on SSHOC (meta)data interoperability problems https://doi.org/10.5281/zenodo.3569867
- Provides an inventory of data and metadata formats currently relevant for SSHOC communities, recommendations, and prioritisations for providing conversions services.
- Pragmatic approach: focus on metadata and data formats.
- Focus on syntactic interoperability (common structures); semantic interoperability is out of scope for T3.5.
- Move from local practices to increased cooperation (e.g. crossdisciplinary, EOSC) will probably deliver a wave of new interoperability challenges.

SSHOC recommended metadata formats

Table 6. Recommended metadata standards.

| Domain / ERIC | Recommended metadata standards | | | |
|--------------------------------------|---|--|--|--|
| All | Dublin Core ¹³ , relaxed DataCite ¹⁴ | | | |
| Social Sciences / CESSDA, ESS, SHARE | DDI Codebook, DDI Lifecycle ¹⁵ | | | |
| Heritage Sciences / E-RIHS | CIDOC-CRM ¹⁶ (and its extensions, especially PEM ¹⁷) | | | |
| Language Sciences / CLARIN | CMDI ¹⁸ | | | |
| Arts and Humanities / DARIAH | CIDOC-CRM (and its extensions), EDM ¹⁹ , TEI (teiHeader) ²⁰ | | | |

• Many interoperability problems can be avoided if standards, formats, practices and tools are documented well and transparently.

- The basis for any standard/format interoperability are explicit and sufficient standard/format descriptions, i.e. references of all elements and values to accepted semantic registries.
- Although most of the standards/formats already have made a start with this, a great deal of work remains to be done.
- In SSHOC T3.5, areas of possible future investigation and elaboration include standard and format registries, rights management, version management, and coherent mapping between the features of different standards and formats.

Thank You!

mari.kleemola@tuni.fi