
The importance of FAIR DDI-compliant metadata in optimising reuse and preservation of SSH research data

Alina Danciu, Nicolas Sauger, Sciences Po, Center for Socio-Political Data (CDSP), CNRS

Conference on Smart Metadata for Official Statistics (COSMOS), INSEE, 11-12 April 2024, Paris

CDSP, a pioneer in French data sharing

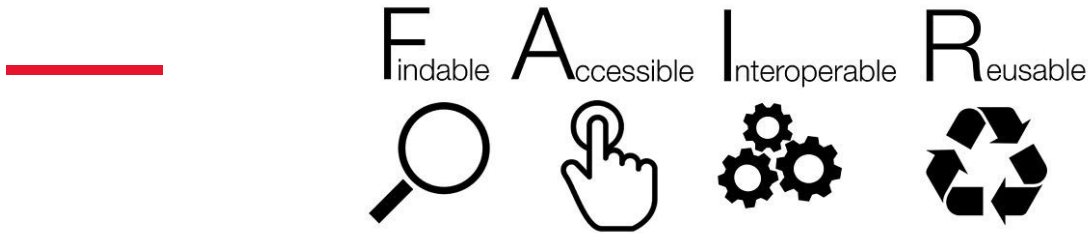
- Since **2005**, **preservation and sharing of sociology and political science surveys and data**
- More than 250 databases : political attitudes, gender, family, immigration, school, health, cultural practices, new technologies, etc.
- 150 recoded French election results databases from 1958 to 2012
- **Among the first French archives to adopt the Data Documentation Initiative (DDI)**
- The DDI suite can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving





FAIR principles all along the data lifecycle

- CDSP adopted **DDI-Codebook**, one of the products of the suite since the centers' creation. This product is used to document mostly “simple” survey data.
- Experimentation with **DDI-Lifecycle**, for managing longitudinal panel data (French [ELIPSS](#) probabilistic panel).
- Active role in supporting DDI use in France: the first to provide online DDI training in French in 2019.
- CDSP respects the FAIR principles all along the data lifecycle and was granted the **CoreTrustSeal** for its data bank in August 2023 (first SSH French archive to have it)



- One of the very important aspects in being eligible for the CoreTrustSeal is the **respect of FAIR principles**
- First three FAIR criteria are now widely explored by the research data community, the R principle is much less so

What makes data reusable, once it is findable, accessible, and interoperable?

What makes data reusable? GO FAIR recommendations

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings. ()

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.

R1.1. (Meta)data are released with a clear and accessible data usage license.

R1.2. (Meta)data are associated with detailed provenance.

R of FAIR: data quality

- Respecting an internationally accepted **metadata schema for metadata**, along with **quality criteria for metadata and data**, are key elements for both reuse and preservation.
- What do we mean by data quality?
The CDSP accepts data that have demonstrated their importance to the social science community through their substantial research and/or teaching value, and their legacy value or uniqueness
- Need for **formal criteria in terms of data quality itself**, which would make it possible to prejudge the reproducibility (obtaining same results using the same data and same process) and reuse of data (the R of FAIR).
- Efforts to be made in terms of metadata describing the origin and **processing of data & publication of codes and algorithms**.

To move on

- More **formal data criteria**
- Publication of **codes and algorithms**
- Description of the **larger or cultural context of data collection**
- Explore new standards such as SDTL (Structured Data Transformation Language) for representing data transformation commands

Get in touch!

alina.danciu@sciencespo.fr

nicolas.sauger@sciencespo.fr