



HAL
open science

Challenges for the management and integration of historical data

Sébastien Oliveau

► **To cite this version:**

Sébastien Oliveau. Challenges for the management and integration of historical data. International Workshop on Historical Big Data, Dec 2022, Madrid (Université Carlos III), Spain. hal-03921607

HAL Id: hal-03921607

<https://hal.science/hal-03921607v1>

Submitted on 3 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

International Workshop on Historical Big Data: Research Data and infrastructures for socio-economic history and social sciences

Challenges for the management and integration of historical data

the example of a French research infrastructure, Progedo

Sébastien OLIVEAU, Progedo's director - sebastien.oliveau@cnrs.fr



DATA
INFRASTRUCTURE

Progedo is member of

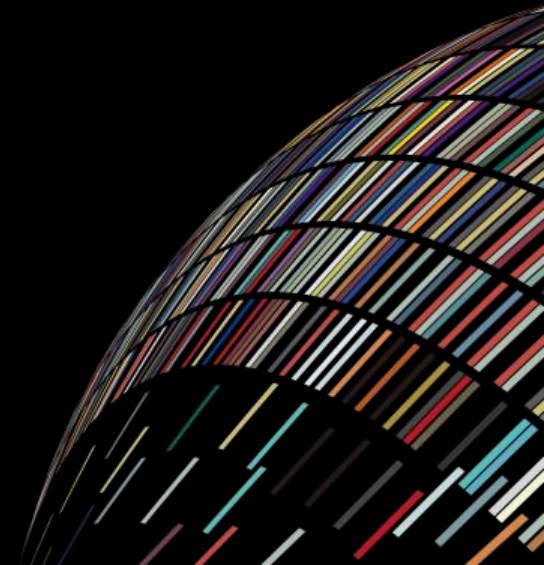


CESSDA



FAIR-IMPACT

Expanding FAIR solutions across EOSC



What's an historical data?

Data Used by historians?

Old Data?

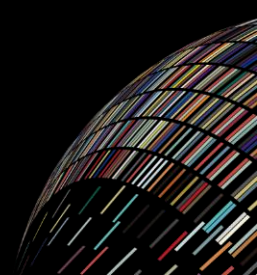
Data build from primary sources?

Incomplete data?

And what are their characteristics?

Data without specific legal issues?

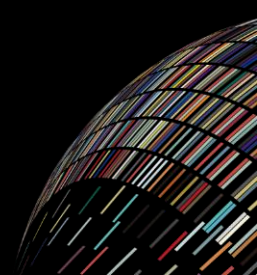
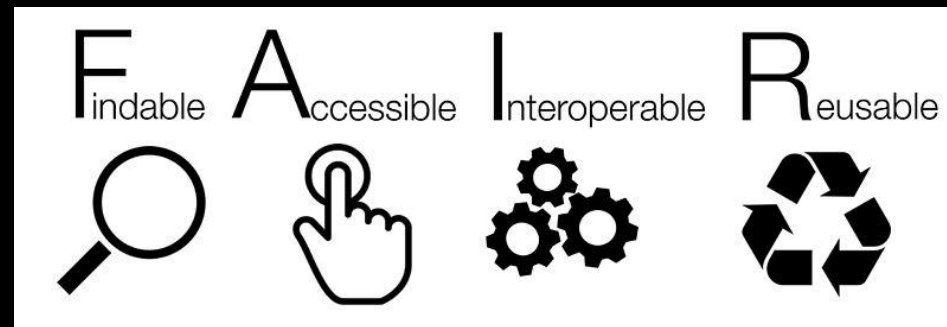
Diachronic/longitudinal Data?



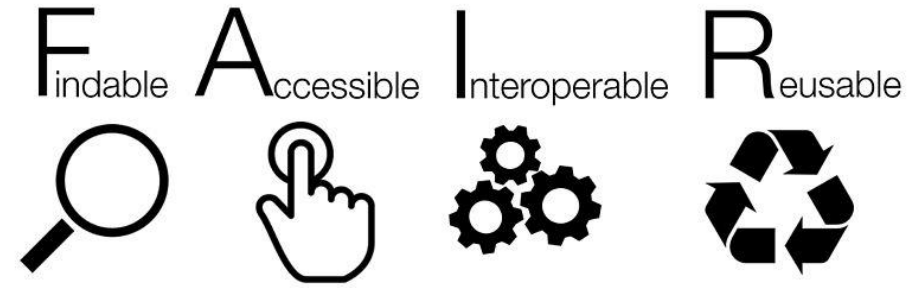
Statistical data Management - Principles

About FAIR Data

About FAIR statistical data



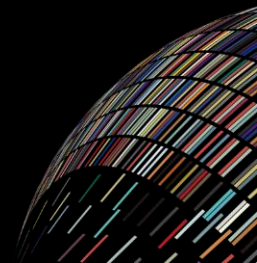
Findable



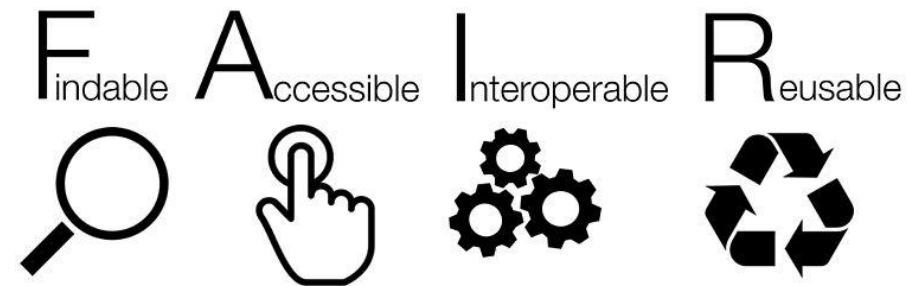
Findable (so searchable) online

=> catalogue (<https://data.progedo.fr>)

With a **Persistent identifier** (Handle) such as DOI (which is also actionable).



Accessible



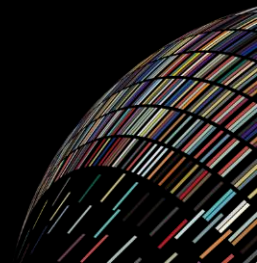
In a (public) repository

Openly available or through an explicit access procedure
(for restricted data: registration and terms of law)

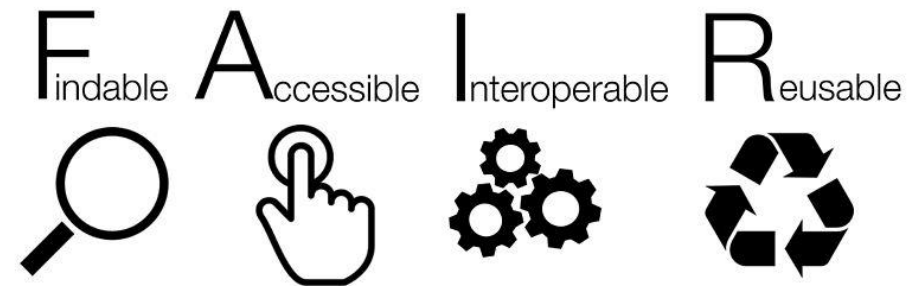
Harvard Dataverse <https://data.harvard.edu/dataverse>

Zenodo <https://zenodo.org>

Quetelet-Progedo-Diffusion <https://data.progedo.fr>



Interoperable



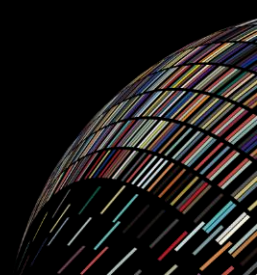
Technological and semantic interoperability

The standards employed should be explicitly mentioned as this is the first requirement for interoperability

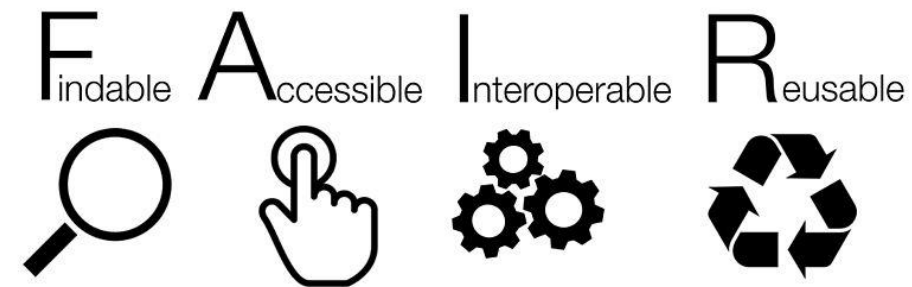
Documentation standards

Citation Standards

Harvesting standards



Interoperable



Technological and semantic interoperability

The standards employed should be explicitly mentioned as this is the first requirement for interoperability

Documentation standards

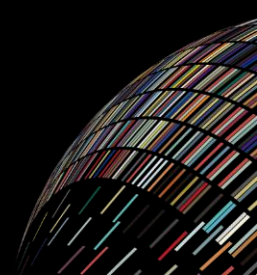
Dublin Core, DDI

Citation Standards

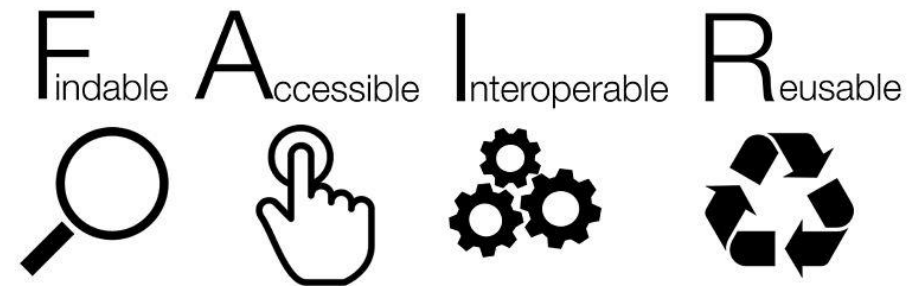
DataCite, etc.

Harvesting standards

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)



Interoperable



Technological and semantic interoperability

The standards employed should be explicitly mentioned as this is the first requirement for interoperability

Documentation standards

Dublin Core, DDI

Citation Standards

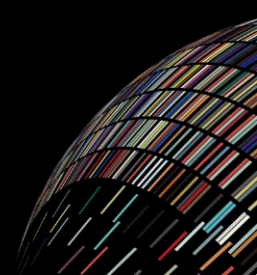
DataCite, etc.

Harvesting standards

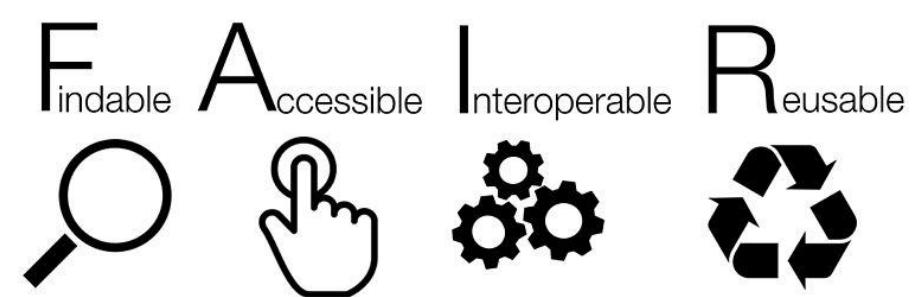
Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

Formats (open formats such as XML, CVS, etc.)

Controlled vocabularies



Interoperable



Technological and semantic interoperability

The standards employed should be explicitly mentioned as this is the first requirement for interoperability

Documentation standards

Citation Standards

Harvesting standards

Formats (open formats such as XML, CVS, etc.)

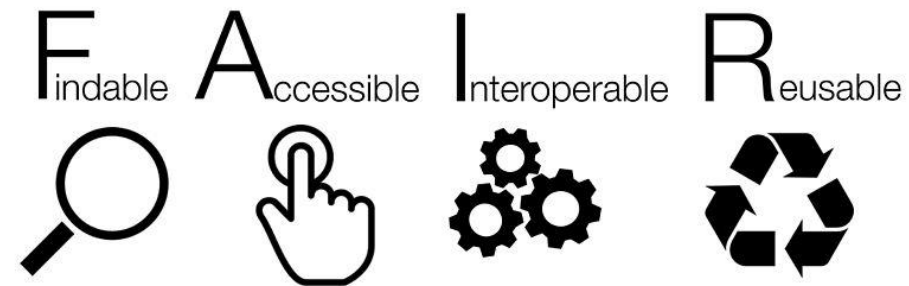
Controlled vocabularies

=> Machine actionnable





Reusable

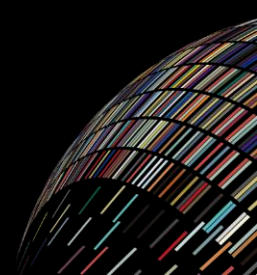


Metadata

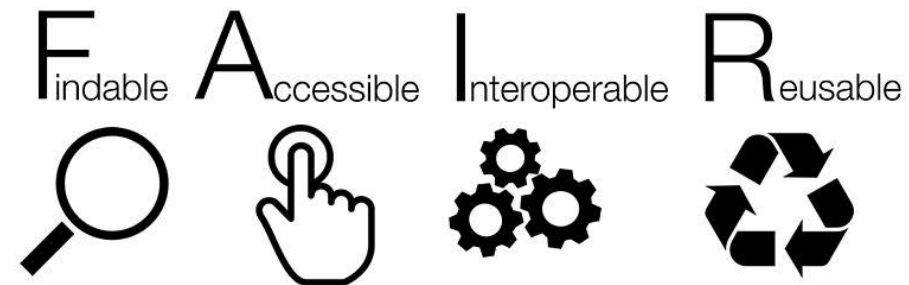
The Data Documentation Initiative (also known as DDI) is an international standard for describing surveys, questionnaires, statistical data files, and social sciences study-level information.



DATA DOCUMENTATION INITIATIVE



Reusable



Metadata

The Data Documentation Initiative (also known as DDI) is an international standard for describing surveys, questionnaires, statistical data files, and social sciences study-level information.



Licence and sources need also to be clearly indicated

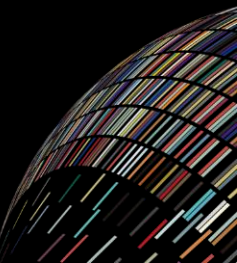
Statistical data Management - Application

Making the Data FAIR is not so easy... need to be helped

⇒ Research infrastructure

⇒ Local, national or international level?

⇒ What is the use of European infrastructure?

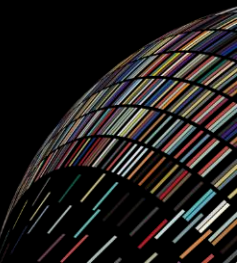


Statistical data Management - Progedo

Local relay: “University Data Plateforme” (PUD in french)

A national repository and catalogue, managed by a team of research assistants “Public Stastics Data Archive (ADISP in French)

A European Integration through Cessda (FAIR Practices, Catalogue, etc.)



Progedo, Cessda & the EOSC

Findable

=> catalogues <https://datacatalogue.cessda.eu>

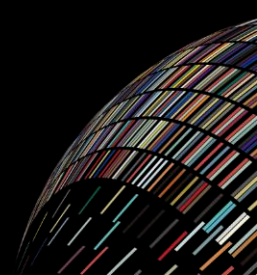
=> SSHOC  **SSHOC**
social sciences & humanities open cloud

Accessible

Interoperable

=> Definition of common standards for Data Archive

Reusable



National pluridisciplinary

Recherche Data Gouv

Harvest metadata

international disciplinary

Cesda Data Catalogue

Harvest metadata

Quetelet-Progedo-Diffusion repository and national catalogue

Feed the repository

Harvest data

ADISP

National Institute of Statistics

Produce data

Ministries

Produce data

Universities and research Centers

Produce data

INED

Produce data



International Workshop on Historical Big Data: Research Data and infrastructures for socio-economic history and social sciences

Challenges for the management and integration of historical data

the example of a French research infrastructure, Progedo

Sébastien OLIVEAU, Progedo's director - sebastien.oliveau@cnsr.fr

**PRO
GE
DO**

DATA
INFRASTRUCTURE

Progedo is member of



cessda

eosc

FAIR-IMPACT

Expanding FAIR solutions across EOSC

