Diary of our initiatory journey on the continent of data citation in SSH

Nicolas Larrousse, Huma-Num CNRS

Edward J. Gray, Huma-Num CNRS / DARIAH

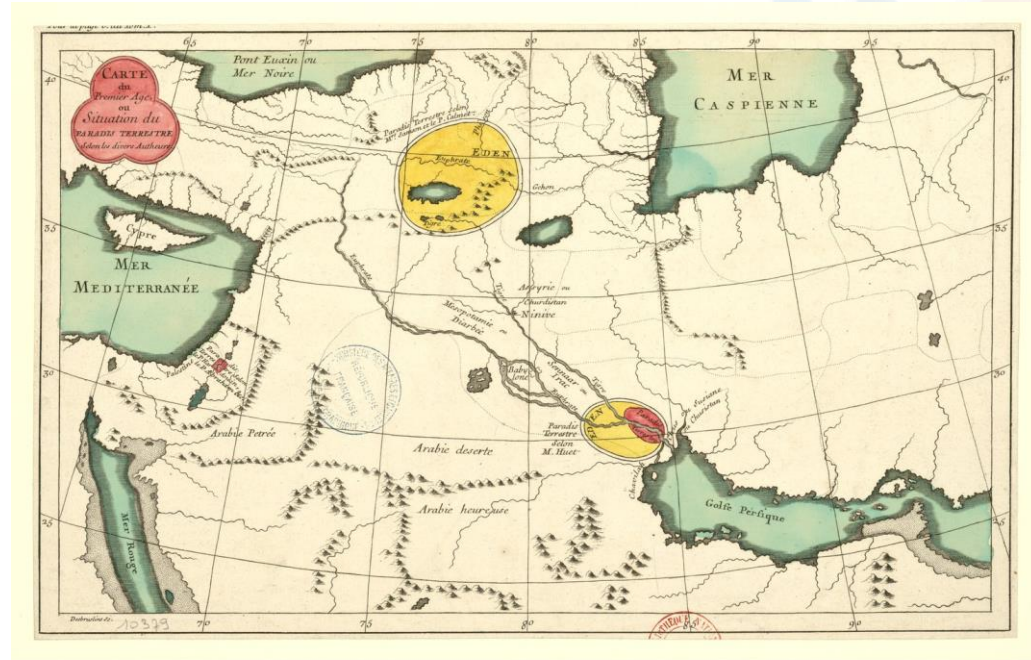Cesare Concordia, ISTI-CNR

DH 2022 Conference - Tokyo

SSHOC
social sciences & humanities open cloud

Data Citation in SSH … A Terra Incognita?

Data citation is not very common in SSH world

Citation practices are very diverse within SSH communities

Community of practices but not real standards

Source gallica.bnf.fr / Bibliothèque nationale de France

Work carried out during SSHOC Project funded by H2020 Program - Task 3.4 "Making Data FINDable by being Citable"

SSHOC
social sciences & humanities open cloud

Data Citation in SSH … An initial exploration

Overview of existing data-citing mechanisms

In SSH research communities: practices are seldom standardized

Citations are generally based on text

=> Citations are not really "actionables"

Domaine public, https://commons.wikimedia.org/w/index.php?curid=41497

See https://doi.org/10.5281/zenodo.3595964

SSHOC
social sciences & humanities open cloud

Data Citation in SSH … A first "Map"

Production of a set of Recommendations based on Force11 citation principles

Explanations of these general principles in the context of SSH practices in a "Why" "How" "What to expect" practical way

Importance: Current lack of a "data citation culture", "disappearance" of research material by non publication.	➢ Outline the importance of data articles & data journals. ➢ Preserve and disseminate to avoid disappearance and minimize costs. ➢ Select a "trusted" repository and use standards	➢ Raising awareness on data citation ➢ Minimizing costs of data creation ➢ Endorsing reuse and standardization of data	End users/researchers

See https://doi.org/10.5281/zenodo.5361718

Source gallica.bnf.fr / Bibliothèque nationale de France

SSHOC
social sciences & humanities open cloud

Data Citation in SSH … A deeper Exploration

Analysis of 85 SSH Data Repositories

Checks for:
- Presence of PID, and PID type
- Presence of Landing Page
- Presence of Structured Metadata encoded in webpage
- Presence of "Cite As" Feature
- Use of Standardized Vocabularies (such as ORCID)
- Use of Versioning
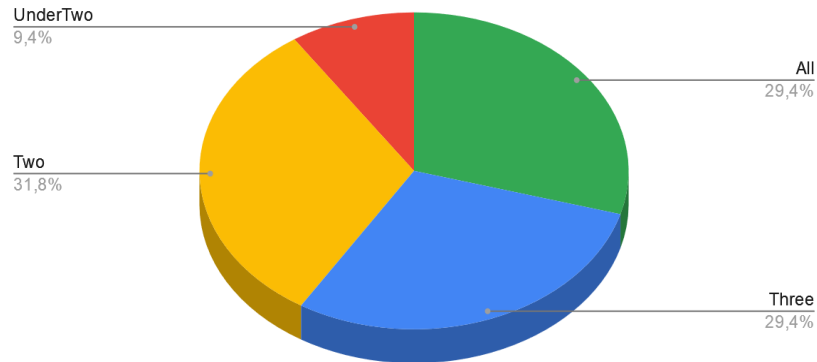- Presence of Links to Related Publications

Source gallica.bnf.fr / Bibliothèque nationale de France, Département des Manuscrits, Français 2810

See https://doi.org/10.5281/zenodo.5603306

SSHOC
social sciences & humanities open cloud

Data Citation in SSH … A deeper Exploration

Analysis of 85 SSH Data Repositories

Number of (4) Main Criteria Fulfilled

UnderTwo
9,4%

All
29,4%

Two
31,8%

Three
29,4%

Four Main Criteria:

- Presence of PID, and PID type
- Presence of Landing Page
- Presence of Structured Metadata encoded in webpage
- Presence of "Cite As" Feature

Encouraging, but still areas of improvement

See https://doi.org/10.5281/zenodo.5603306

SSHOC
social sciences & humanities open cloud

Data Citation in SSH … A Tool to Harvest Resources

Development of a prototype to retrieve enriched information for citation purposes

It is composed of three components:
- a harvester which grabs information about a dataset and normalizes it
- an API to disseminate the metadata of the citation thereby making it actionable
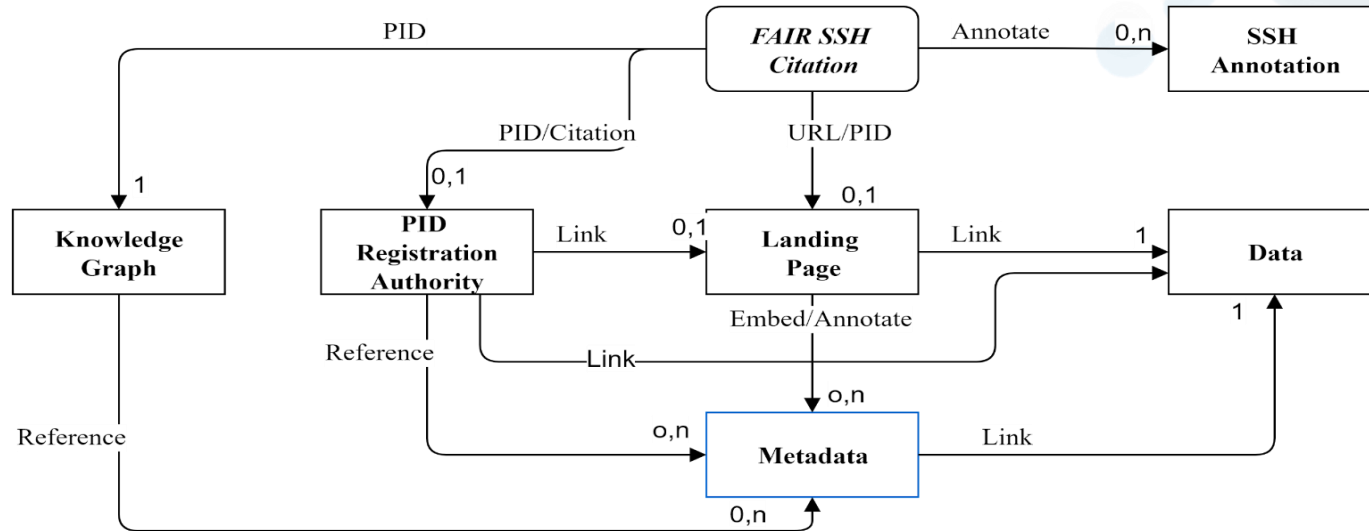- a citation viewer for human purposes

Source gallica.bnf.fr / Bibliothèque nationale de France

See http://v4e-lab.isti.cnr.it/citview/demo/index.html

SSHOC
social sciences & humanities open cloud

Data Citation in SSH … A Tool to Harvest Resources

In practice: getting metadata to build "FAIR SSH citations"

See http://v4e-lab.isti.cnr.it/citview/demo/index.html

Data Citation in SSH: Still "Uncharted Territories"

Recommendations for FAIR Data Citation in the Social Sciences and Humanities link

Consultation in events & expert review

Source.gallica.bnf.fr / Bibliothèque nationale de France

Inventory of current practices about Data Citation within SSH

link

Review of SSH Data Repositories

link

Data Citation Prototype link

Automatically extracts citation metadata from PIDs (persistent identifiers)

SSHOC
social sciences & humanities open cloud