

# **The *CESSDA Data Citation Guide* and how it might be strengthened in relation to legal and ethical aspects of data reuse**

André Jernung, Swedish National Data Service

 <https://orcid.org/0000-0002-9027-0113>



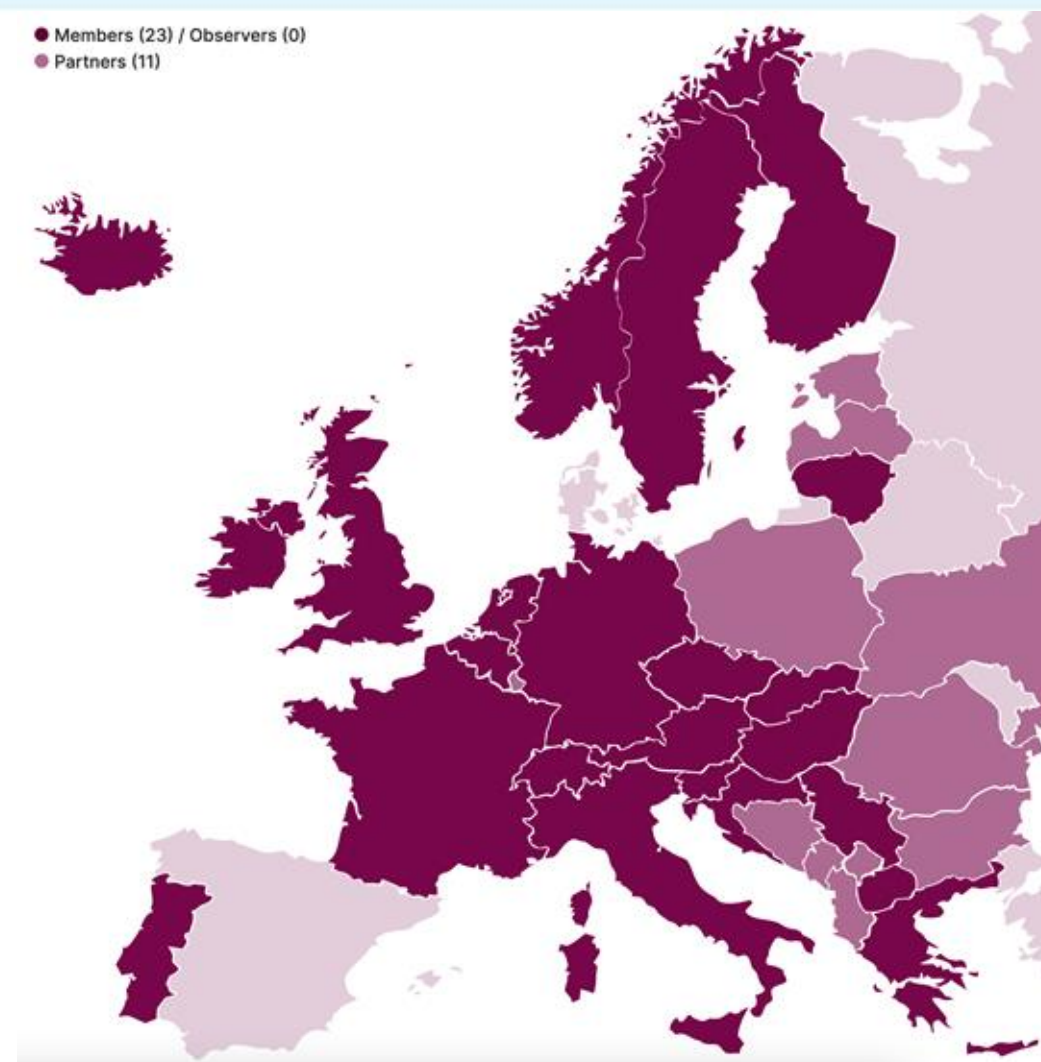


CESSDA, the Consortium of European Social Science Data Archives <https://www.cessda.eu/>

A European Research Infrastructure Consortium (ERIC) bringing together social science data archives and repositories

Its role is to:

- Ensure that data are preserved, curated, and reusable, fostering trust in research
- Facilitate cross-border access to research data
- Provide integrated and sustainable data services for the social sciences
- Foster collaboration among data infrastructures, researchers, and decision-makers



## Some of the other CESSDA services and resources



[CESSDA Data Catalogue](#)



[ELSST Thesaurus](#)



[CESSDA Resource Directory](#)



[Data Management Expert Guide \(DMEG\)](#)




[Data Archiving Guide \(DAG\)](#)


Coming soon...



[European Question Bank \(EQB\)](#)

# New resource: CESSDA Data Citation Guide

 Data Citation Guide



**CESSDA Data Citation Guide**

- Why Cite Data?
- The Anatomy of a Data Citation
- Recommendations for Data Repositories
- Recommendations for Authors
- Recommendations for Journals and Publishers
- Recommendations for Research Performing Organisations (RPOs)
- Recommendations for High-Level Research Policy and Strategy Development Entities
- Recommendations for Ethics Committees
- References
- CESSDA Data Citation Working Group

## Welcome to the CESSDA Data Citation Guide!

Citing data is essential to ensuring transparency, reproducibility, and recognition in research. Just as scholarly articles are cited, datasets must be acknowledged to give proper credit to data authors and to support the discovery, access, and reuse of research data.

At [CESSDA](#), we advocate for **best practices in data citation** across the social sciences. This guide is designed to **support the entire research ecosystem**—from individual researchers to policymakers—in fostering a more coherent, FAIR-aligned, and sustainable data citation culture.

### What you will find in this guide

The Data Citation Guide begins by explaining [why data citation matters](#) and [what a data citation should include](#).

It then offers **targeted and practical recommendations** for different stakeholder groups:

- [Data repositories](#)
- [Authors](#) of publications mentioning and/or using data
- [Academic journals and publishers](#)
- [Research performing organisations](#) (RPOs)
- [High-level research policy and strategy development entities](#)
- [Ethics committees](#)



[datacitation.cessda.eu](https://datacitation.cessda.eu)

Research data should be cited to ensure transparency, reproducibility and recognition in research.

## Is there a need for yet another resource?

- Many good and pioneering resources on data citation out there
- Some are quite STEM-oriented
- Some are focusing on improving practices even further within research communities with a healthy data citation culture
- Many resources **target only one stakeholder group**, f.e. researchers/authors
- But... researchers have to consider **frameworks, limitations and incentives created by other stakeholders** such as publishers, funders, policy makers, research infrastructures and others
- Thus, **many stakeholders within the research ecosystem affect data citation behaviour**, creating further need for awareness-making efforts

# General goals of the Data Citation Guide



## Data Citation Guide

- **Enabling and facilitating data citation practices**
  - Identifying core requirements
  - How to do the right thing: give credit, enable transparency and reproducibility
  - Cite data that you produced yourself – Cite data being re-used or referenced
- **Enhancing data citation interoperability and FAIRness**
  - Describing and interconnecting the research ecosystem through metadata and relations
  - Enabling end users to do this through well-configured infrastructures
- **Raising awareness among key stakeholder groups**
  - Showing how structures and incentives affect data citation behaviour
  - Pinpointing counterproductive practices and blockers
- **Working towards cultural change for research data as a scholarly output**
  - Published data as valuable, independent research outputs
  - Targeting incentive structures (research assessment, merit and recognition, research output indexing)

# Definitions and anatomy of a data citation

## What is a data citation?

- A reference to the data themselves
- Placed in-text (author-date system) or as a footnote (notes system) whenever data are mentioned
- Listed in the reference list, like any other scholarly source
- The reference may take other shapes if not within a traditional research publication

## Six core components

1. Data author(s)
2. Title of the data
3. Year of publication
4. Version
5. Data publisher
6. Persistent identifier (e.g. DOI)

## Additional components

- Accession or dataset number
- Resource type (e.g. [Dataset])
- Place of publication



# Examples of references using the data citation components

Mapping the data citation components to the [APA 7th](#) reference style:

Data author(s) (Publication year). *Data title* (Data number; Version)  
[Resource type]. Data publisher. Persistent identifier

Examples:

ISSP Research Group (2025). *International Social Survey Programme: Family and Changing Gender Roles V - ISSP 2022* (ZA10000; Version 1.0.0) [Dataset]. GESIS. <https://doi.org/10.4232/5.ZA10000.1.0.0>

Mach, A., Ginalski, S., Davoine, E., Leimgruber, M., Strebel, M. A., Piguet, S., Araujo, P., & Benz, P. (2025). *OBELIS Swiss Local Elites: Basel, Geneva and Zurich* (2808; Version 1.0) [Dataset]. FORS. <https://doi.org/10.48573/b9p3-rz53>



# Key stakeholders addressed in the DCG



1. **Request data to be cited**
2. Provide all six core data citation components
3. Provide data **versioning and access information**
4. Provide **citation suggestions** including the core components
5. Facilitate data citation through export and download of **metadata for reference management software**
6. Make citation information and other publication metadata accessible through **generic structured metadata** files
7. Make data publication and **relations** to other entities machine-actionable through rich metadata and **PIDs**
8. Raise **awareness** about data citation

1. **Cite the data themselves**, citing associated reports or articles is not sufficient, data availability statements are not substitutes for citing
2. Include all **six core components** in the reference list
3. Cite **primary and secondary data** as well as **replication materials**

1. Allow data citations in the **standard list of references**
2. **Require** authors to cite data specifically
3. Include all **six core components** in data citations and provide citation examples
4. Encourage authors to **publish data and replication materials** as independent publications
5. Encourage authors to provide **PIDs** for published data
6. Allow **time for quality assurance** and data curation by data repositories

1. Encourage data citation practices
2. Encourage authors to **report data publications** as research outputs
3. Include data publication **metadata** in indexing services
4. Include and leverage the **added value of metadata** describing data publications in administrative systems
5. Take data publications into account when summarising outputs

# Research Policy and Strategy Development Entities

1. **Require** data to be cited
2. Include all six core components in guiding policies
3. Make the **benefits** of data citation clear in policies
4. Require **active reporting** of data publication outputs through research infrastructures

1. All research should follow **good citation practice**
2. Even **restricted** data should be **citable**
3. Address data citation in data management plans



## New stakeholder groups?

### Data Holders

- Legal / administrative responsibility for data
- May be decision-makers for reviewing legal and ethical considerations and allowing access to data on a case-by-case basis
- Important stakeholders for long-time governance of the data context being cited
- May allow further verification, provenance, transparency and reproducibility
- Maintaining links between data and physical objects, samples or undigitised materials
- May maintain links between data and individuals/research subjects if relevant
- May be hospitals, universities, cultural heritage institutions and others

New stakeholder groups?

## Research Output Aggregators & Catalogues

- Re-uses metadata and makes data visible and searchable
- May present their own metadata formats interpreted from a primary metadata source
- May present relations between data and other research outputs and phenomena
- Enables and influences incentive infrastructures: scientometric workflows, citation tracking and RPO rankings
- May present access levels, licences and legal/ethical restrictions in different ways
- May be databases, catalogues, scientific knowledge graphs etc.
- Examples: OpenAIRE Graph, Web of Science

New stakeholder groups?

## Reference Style Maintainers

- Important stakeholder creating constraints for what is able to be cited
- Influences authors as well as publishers and their submission guidelines
- May be bound to specific traditions, media types or historical limitations
- May create obstacles for sustainable data citation i.e. providing provenance, versioning or PIDs in citations

- Which issues and stakeholders do you believe should be addressed in the further development of the Data Citation Guide?
- What additional recommendations would facilitate data re-use and citation with regards to long-term governance as well as legal and ethical accountability?
- Did you notice something that you disagree with in the guide?

CESSDA Data Citation WG <https://datacitation.cessda.eu/>

# Thank you



[eden-fidelis webpage](#)



[/company/fidelis-eosc/](#)



[@fidelis-eosc](#)