FORS GUIDES
to survey methods
and data management

Data Citation: How and Why Citing (Your Own) Data

Christina Bornatici[1] and Nicolas Fedrigo[1]

[1] FORS

FORS Guide No. 19, Version 1.1 (last update January 2025)

March 2023

Abstract:

This FORS Guide calls on all parties involved to commit to best practices regarding data citation. It first explains why researchers and students should cite the data they use, and how to cite data correctly. It also focuses on the practice of major Swiss journals and data publishers in social sciences and closely looks at their data citation recommendations and requirements.

Keywords: citation of datasets, best practices in citation, authorship attribution, data findability, citing data sources

How to cite:

Bornatici, C. & Fedrigo, N. (2023). Data Citation: How and Why Citing (Your Own) Data. *FORS Guides*, 19, Version 1.1, 1-15. https://doi.org/10.24449/FG-2023-00019

The FORS Guides to survey methods and data management

The FORS Guides offer support to researchers and students in the social sciences who intend to collect data, as well as to teachers at university level who want to teach their students the basics of survey methods and data management. Written by experts from inside and outside of FORS, the FORS Guides are descriptive papers that summarise practical knowledge concerning survey methods and data management. They give a general overview without claiming to be exhaustive. Considering the Swiss context, the FORS Guides can be especially helpful for researchers working in Switzerland or with Swiss data.

FORS, Géopolis, CH-1015 Lausanne
www.forscenter.ch/publications/fors-guides
Contact: info@forscenter.ch

Acknowledgement:

We thank Sharon Bolton (UK Data Service), as well as Brian Kleiner and Emilie Morgan de Paula (FORS Data Management and Archive Services) for their attentive proofreading and valuable feedback on earlier versions of this Guide.

1. INTRODUCTION

Citing sources is a fundamental research practice that is taught to students from the beginning of their academic curriculum. One of the main reasons to cite sources is to acknowledge and give credit for someone else's or one's own previous work. However, this concerns mostly 'traditional' sources, such as academic publications (e.g., books and journal articles). Indeed, despite having emerged in the late 1990s, data citation is still an uncommon practice in academic publications across scientific disciplines (Late & Kekäläinen, 2020; Parsons et al., 2019; Vannan et al., 2020; Yoon et al., 2019; Zhao et al., 2018; Zuiderwijk et al., 2020).

Data are usually presented in the methodological section of a publication. Data authors, that is persons who contributed to the production of the data, also called data producers or data creators, are sometimes mentioned in a footnote. A data-related publication (e.g., methodological report) might also be cited. However, in such cases data are generally not correctly cited in publications. This Guide aims to show how data can be cited properly. We adopt the definition, partly based on FORCE11[1] (2014), that *'a correct data citation'* refers to the data itself, whether published or unpublished, and appears both in the text, tables and graphs in an abbreviated version, as well as in the bibliographic section of a publication in a formal and structured reference including at least the core citation components (i.e., data authors, title, version, year published and publisher) (see Part 3 of this Guide).

Data are essential for conducting empirical research. Given this, on the one hand, the lack of data citation seems odd (Vannan et al., 2020), especially when the value of data for research has gained more and more recognition lately. Good data management and sharing practices have been promoted by several initiatives driven by the open science movement, and more recently open research data strategies. For example, the first Swiss National Strategy for Open Research Data was published in 2021 (swissuniversities, 2021). Technologies and infrastructures, such as the Digital Object Identifier System (DOI) or data repositories, were also developed or improved to facilitate data sharing and reuse. This has resulted in more data being available to conduct research, and therefore has reinforced the need to attribute credit to data authors.

On the other hand, although data citation is far from being a new concept, a consensus has only been reached recently on setting out the core elements of good citation practice (Parsons et al., 2019). Indeed, in 2014 several libraries, data archives and repositories, in partnership with the information science communities, endorsed a Joint Declaration of Data Citation Principles (FORCE11, 2014). Guidelines and requests for citing data were already defined long ago by data publishers. For example, in common with other national data archives such as the British (UKDS), the German (GESIS) or the Finish (FSD) social science data archives, the Swiss data archive for social sciences FORS (2008-present) and its predecessor SIDOS (1993-2007) always requested that data accessed through their services is cited. However, it is only since 2014 that some journal publishers, bibliographic styles, and reference management software have also added guidelines, recommendations, or specific resource types for data citation (Parsons et al., 2019; Silvello, 2018). This has led to higher awareness

[1] FORCE11 stands for The Future of Research Communications and e-Scholarship. FORCE11 presents itself as "a community of scholars, librarians, archivists, publishers and research funder" aiming "to bring about a change in modern scholarly communications through the effective use of information technology", and therefore "to facilitate the change toward improved knowledge creation and sharing" (https://force11.org/info/about-force11/, accessed 01.12.2022).

of the importance of citing data and to progress in terms of data citation among researchers (Late & Kekäläinen, 2020; Parsons et al., 2019).

Through the open science movement and international initiatives such as the FORCE11 Joint Declaration, it has increasingly become recognised that citing any data used in a publication or assignment is equally important as citing a book, an article, or any other sources contributing to the research. Indeed, data citation aims to acknowledge the authors of the data and to facilitate access to the data (Parsons et al., 2019). Moreover, data citation increases transparency and, thus, gives credibility to the author of a scientific publication (such as a journal article). For these reasons (and many others described in Part 2), correct data citation is part of good and fundamental research practice and should be known and applied by all students and researchers and anyone using data.

The main aims of the present Guide are to explain 1) why data users (i.e., persons who are analysing the data and publishing results based on them) should cite the data they use, and 2) how to cite data correctly. While the primary audience of this Guide is researchers and students, as main data authors and/or data users, this Guide also represents a good opportunity to call on the other parties involved in social science research in Switzerland to commit to best practices regarding data citation. These other parties include the publishers of scientific articles (e.g., academic journals), as well as data publishers (e.g., data repositories and archives). Their role is essential, whether this be in requiring data to be correctly cited in published articles, or in providing recommendations and examples regarding the citation of data that are preserved and disseminated.

2. WHY CITE DATA?

Across scientific fields, it has been recognised that applying correct data citation practices has immediate and long-term benefits for data authors, data users, and research in general. The following main reasons for data citations are provided in the literature (Cousijn et al., 2019; Dosso & Silvello, 2020; Parsons et al., 2019; Silvello, 2018), as well as by data archives and repositories (e.g., Finnish Social Science Data Archive, 2022; UK Data Service, 2021), and university libraries (e.g., Columbia University, 2021; Cornell University, 2022; University of Geneva, 2022):

- **Data attribution:** Data citation allows the authors of a publication to properly acknowledge data authors (including themselves if they created the data) and to attribute credit to them for collecting, documenting, and sharing their data. It also allows identification of the authors of the data used in the publication, so that they can be contacted in case of questions.
- **Data findability and access:** Data citation can be an entry point to data otherwise not or hardly findable. Indeed, data citation makes it easier for the readers to find and access the data. With the use of persistent and unique identifiers (such as DOIs), location and access to data is quick and ensured for the long-term.
- **Data reuse:** Since data citation facilitates findability and access to the data used in a publication, it also helps and fosters reuse of the data by the readers in order to address new research questions.

- **Transparency:** Data citation indicates precisely which data, and its specific version, has underpinned the research. This gives credibility to the work of the data user, in accordance with the research standards for citations.
- **Reproducibility:** Data citation supports the reproducibility of science and, more specifically, of your research by enabling other researchers to find the data used for the analyses in order to reproduce and verify the results and, thus, assess their integrity.
- **Data recognition:** Citing data specifically contributes to recognising data as a legitimate scholarly contribution, that is, recognising the creation and sharing of data as important scientific contributions alongside scientific articles and monographs.
- **Data impact:** It is likely that data authors wish to know if, why, and in what manner the fruit of their work is used by others, as they do for other research outputs such as articles. Data citation ensures the traceability of reuse and facilitates measuring the impact of data. These two factors are particularly useful for data authors who need to demonstrate the wider value of their data, for example to obtain funding to collect new data. Moreover, as research data is rendered visible by the citation, it becomes a meaningful input to the scholarly record. Following this reasoning, the creation of data and its secondary use may provide an additional measurable means of appraising a researcher's impact. Needless to say that currently citations reveal only a fraction of the actual (re)use of the data (e.g., for the social sciences Late & Kekäläinen, 2020).
- **Data sharing:** Finally, citing the data and therefore acknowledging the data authors is an important incentive for future data authors to share their data. Indeed, if this practice enters more fully into the scientific culture, data authors will know that they will be credited for the extra work that data sharing implies (e.g., thorough documentation and anonymisation of the data to allow for reuse).

3. HOW TO CITE DATA CORRECTLY?

Three aspects should be considered to cite data correctly. First, the citation should refer to the data itself (sometimes the data are also referred to as dataset). A dataset is composed of one or several data files (e.g., interview transcripts, survey data) and can include documentation files (such as a methodological report on data collection and data management). Generally, datasets are accessible from a discipline-specific, institutional or journal data archive or repository. A data citation should, thus, link to the page where the dataset is published. In this sense, citing a methodological article or report that explains sampling, survey questionnaire developments, and data collection is not citing the data itself. Further, this does not constitute a good data citation practice and cannot be considered sufficient for correctly citing the data.

Second, the data should be cited in-text, in an abbreviated version, and included as a full version in the general reference section, as is the case for any other sources and materials contributing to a publication, such as books, articles, reports, conference presentations, or websites (Silvello, 2018). Therefore, using footnotes, notes on tables, and figures mentioning the data is not sufficient to correctly cite data.

Third, to be considered correct the full citation must include at least the core components that are explained in more detail below.

3.1 CORE COMPONENTS OF A DATA CITATION

Citing data correctly also implies choosing the appropriate citation components, which should meet some standards. Different institutions – including specific international consortia and associations such as DataCite[2] or IASSIST[3], data repositories and archives, university libraries and scientific journals – have proposed elements that should constitute a data citation (see e.g., IASSIST, 2012; UK Data Service, 2022). Whilst in the past, recommendations lacked harmonisation, a consensus has now been reached on the core elements of data citation. The six elements below are essential components of a data citation and are generally considered as the minimal requirements for data identification and retrieval (see e.g., Jessop, 2021; Silvello, 2018; University of Geneva, 2022):

- **Author:** Name of each individual or organisational entity responsible for the creation of the data. This could also be the rights holder.
- **Title of the data:** Complete title of the data.
- **Year of publication:** Year the data was published or disseminated for the version of the data used.
- **Version:** Version or edition number associated with the data.
- **Data publisher:** Organisational entity (e.g., data repository or archive) that makes the dataset available by preserving and/or disseminating the data.
- **Persistent identifier:** Unique electronic identifier used to locate and access the data (such as a DOI). If available, the persistent identifier should be included as a link (e.g., https://doi.org/xxxxxx) and used instead of the URL or other Internet link.

These core components are very similar to the ones used in the citation of other materials. For example, a book citation is composed of the book's author(s), title, publication year, publisher, and place of publication (although the latter tends to disappear, see e.g., APA 7th edition), as well as the book's persistent identifier (i.e., an ISBN) when it is available online. For journal articles, the standard citation displays in addition the journal title. Thus, the only specific additional component to data citation is the version. The version number changes as new data is added (e.g., longitudinal data) or existing data is corrected. It is therefore important to refer to the exact data version used for the analyses to allow for its retrieval.

To properly arrange and format the above elements, authors of a publication (e.g., a scientific article) should follow the publisher's (e.g., the scientific journal's) guidelines and bibliographic style (e.g., APA, Chicago, etc.).

3.2 ADDITIONAL COMPONENTS OF A DATA CITATION

Depending on the citation format requested by the guideline or style, additional components, such as the data number, the resource type, the place of publication, or the date of access, might be requested:

- **Data number**: Unique numerical identifier for the data, provided by the data publisher. The data number is not a persistent identifier. It is only used within a data archive or repository.

[2] DataCite is a DOI provider for research data, https://datacite.org/ (accessed 01.12.2022).
[3] IASSIST stands for International Association for Social Science Information Service and Technology, https://iassistdata.org/ (accessed 01.12.2022).

- **Resource type**: General resource type, often provided in square brackets, e.g., [dataset], [data file and documentation]. Adding the resource type in brackets after the title could help readers to instantly differentiate data citations from other source citations.
- **Place of publication**: Physical location of the data publisher.
- **Access date**: The date of access to the data. If no version number is provided by the data repository or archive, then the access date should be used instead.

Moreover, authors should pay attention to the additional requirements from the data authors or the data publisher. Apart from the citation components, these additional requirements might also be an acknowledgment note, a data availability statement, or a request to cite a methodological report on the data collection or data article besides the data citation.

Finally, one should remember that it is always better to provide more information for the data citation rather than less. The information supplied should allow readers *to identify, retrieve, and access the same unique dataset.*

3.3 CONCRETE EXAMPLES

Some common bibliographic styles have now developed a specific citation style for data (e.g., APA 7th edition). If a specific citation style for data is not provided (e.g., MLA 7th edition, Chicago author-date 17th edition), the format for books is often considered a generic format that can be modified for other source types. In such cases, it is important to make sure that the six essential components are included in the citation. Here are basic forms and examples of data citations using the APA, MLA, and Chicago author-date styles.

APA (7th edition): Author(s) (Publication year). *Data title* (Data number; Version) [Resource type]. Publisher. Persistent identifier

- Ernst Stähli, M., Sapin, M., Pollien, A., Ochsner, M., & Nisple, K. (2022). *MOSAiCH 2021. Measurement and Observation of Social Attitudes in Switzerland. Study on Health and Health Care and related topics* (Ref. 2033; Version 1.0.0) [Dataset]. FORS. https://doi.org/10.48573/t659-e039

MLA (7th edition): Author(s). *Data title*. Version. Place of publication: Publisher, Publication year. Medium of publication. Date accessed. Persistent identifier

- Ernst Stähli, Michèle, Sapin, Marlène, Pollien, Alexandre, Ochsner, Michael, & Nisple, Karin. *MOSAiCH 2021. Measurement and Observation of Social Attitudes in Switzerland. Study on Health and Health Care and related topics*. Version 1.0.0. Lausanne: FORS, 2022. Web. 25 Jul 2022. https://doi.org/10.48573/t659-e039

Chicago author-date (17th edition): Author(s). Publication year. *Data title*. Version. Place of publication: Publisher. Persistent identifier

- Ernst Stähli, Michèle, Marlène Sapin, Alexandre Pollien, Michael Ochsner, and Karin Nisple. 2022. *MOSAiCH 2021. Measurement and Observation of Social Attitudes in Switzerland. Study on Health and Health Care and related topics*. Version 1.0.0. Lausanne: FORS. https://doi.org/10.48573/t659-e039

3.4 USING REFERENCE MANAGEMENT SOFTWARE TO FACILITATE DATA CITATION

More and more data archives and repositories have a tool to download citations' components in RIS or EndNote formats that can be uploaded in a reference management software (e.g., Zotero, EndNote, Mendeley). EndNote has a specific reference type for data called 'Dataset'. In Mendeley or Zotero, a generic reference type should be used, such as the 'Document' item type in Zotero.

When there is no automatic download available, all essential and additional components of a data citation should be filled in manually in the reference management software. Data archives and repositories, such as the FORS Data Service, usually suggest a citation for the data that can be copied and pasted. If no citation is provided, the citation components should be copied from the different descriptive fields of data archives' and repositories' catalogues.

Using a reference management software allows for a quick and easy referencing in-text and in the reference list. However, it is important to always check that the bibliographic style used in the publication displays all essential components, and additional ones (if any requested), in the reference list.

3.5 SPECIAL CASES

Citing unpublished data

It is always preferable that the data is published before the release of your article, book, or other publication to include all core components of the data citation and to make the data findable. If the used data, for various reasons, have not yet been published, the phrase "unpublished data" should be used instead of the publication year. In this case, since the data is unpublished, you will not be given the option of adding the version, publisher, or persistent identifier.

It should be noted that *publishing is not synonymous with sharing*. Indeed, it is worth mentioning that data can be subject to an embargo, which makes access impossible before a date set by the data authors. This option is generally offered by data archives and repositories. With the FORS Data Service, data authors have the possibility to publish their data under an embargo using the SWISSUbase platform.[4] Under these circumstances, information on the data will appear in the catalogue and the data will receive a DOI and have a version number and a formal publisher. This means that all information needed to cite data correctly are available, while the data will not be accessible to third parties before the embargo has expired.

Citing the replication materials and the data

Sharing replication materials, which may contain data, codes and documentation, is increasingly required by academic journals, as this enhances transparency of research findings (Heers, 2021). The authors should cite the replication materials following the data citation guidance previously mentioned.

In a publication, authors could either analyse data collected by themselves (primary data) or data collected by other researchers for another research purpose (secondary data). We argue that no matter whether they use primary or secondary data, *authors should always cite both the replication materials and the (primary or secondary) data, accurately and separately*.

[4] SWISSUbase is a national cross-disciplinary data repository for Swiss research institutions: https://www.swissubase.ch/en/ (accessed 01.12.2022).

In the interest of transparency and accountability, the authors using secondary data must cite the secondary data directly in their publication. Citing only the replication materials that are authored by themselves[5] is not sufficient in this case.

For authors using their own data (primary data), both the data and the replication materials have the same authors, contrary to the last case. However, these two research outputs contain different information and have different purposes. If the replication materials often include only the subset of data used in the analyses (e.g., some observations, variables, or other components), there are generally more data collected within the research project that could be shared in a data archive or repository. Therefore, the data shared in the replication materials is mainly meant to verify and reproduce the results, while the 'full' data could be reused in the future to answer new research questions.

Since the 'full' data and documentation on one hand and the replication materials on the other hand have different purposes and contain different information, the FORS Data Service offers different solutions to preserve and disseminate them. The 'full' data and documentation are archived in and available through SWISSUbase, while the replication materials are shared through the FORS replication service.[6]

4. DATA CITATION IN THE SWISS CONTEXT

We have seen in the introduction that, across fields and contexts, data are not yet sufficiently cited in scientific publications. Social science researchers in Switzerland are certainly no exception. However, currently statistics and studies are lacking regarding their practices. To address the matter, we aimed to report on the status of data citation recommendations and guidelines in the social sciences in Switzerland. We focused on journals and data archives and repositories since they have an important role to play in fostering data citation.

4.1 SOCIAL SCIENCE JOURNALS

First, it seemed worth considering the positions of Swiss journals regarding research data citation. What recommendations or requirements do they issue, notably through the submission process? We assessed submission guidelines for authors from all journals issued by social sciences societies that are members of the Swiss Academy of Humanities and Social Sciences (SAGW/ASSH). These are members related to the anthropology, economy, social sciences, and society sections of the SAGW/ASSH.[7] In total, 10 journals were identified.

For each journal, we checked the online submission guidelines for the presence of 1) data citation requirements (i.e., the journal explicitly requests to cite data in-text and in the reference section), 2) a concrete example on how to cite data (i.e., indicating which components and how

[5] The syntaxes to recode variables and conduct the analyses are created by the authors of the journal article or book chapter for example. It is worth mentioning that replication material can contain a subset of the secondary data used in the analyses *only if* the data usage licence allows it. If this is not the case, then only the data citation including a permanent link leading to the exact data used in the analyses can be shared in the replication materials.
[6] The FORS replication service is dedicated to replication materials related to publications by Swiss researchers, in Swiss journals or using Swiss data: https://resources.swissubase.ch/replication/ (accessed 01.12.2022).
[7] See sections 4 to 7 of the SAGW/ASSH member societies: https://www.sagw.ch/fr/assh/reseau/societes-membres (accessed 01.12.2022).

to arrange them), and 3) any additional requirements regarding data citation. Table 1 summarises the outcomes.

Table 1. Review of Swiss journals policies and guidelines regarding data citation

	Data citation required	**Example of a data citation**	**Additional requirements**
European Journal of Psychology Open (formerly Swiss Journal of Psychology)	Yes	Yes	A data and materials availability statement is required.
Swiss Journal of Economics and Statistics	Yes, but only for publicly available data	Yes	A data and materials availability statement is required.
Swiss Political Science Review	Yes	No	A data and materials availability statement is required.
Die Unternehmung: Swiss Journal of Business Research and Practice	No	No	No
SComS: Studies in Communication Sciences	No	No	No
socialpolicy.ch	No	No	No
Swiss Journal of Educational Research	No	No	No
Swiss Journal of Social Work	No	No	No
Swiss Journal of Sociocultural Anthropology (formerly TSANTSA)	No	No	No
Swiss Journal of Sociology	No	No	No

Note: Last review done on the 23rd of February 2023.

At the time of the website consultation, three journals explicitly required data to be cited (although with some differences regarding the type of data) and required data availability statements. Of these three journals, two also gave concrete examples on how data citation should be constructed. These three journals are rather exceptions or the most innovative in the Swiss landscape. Indeed, most social science journals (7 out of 10) did not yet offer guidelines on how research data should or could be referenced. Article authors are generally requested to fulfil the guidelines for bibliographic reference sources of the chosen bibliographic style and to state the DOI, if available, in the bibliography. If journals provide concrete examples about how to cite more traditional research outputs, this is not the case for data.

Overall, researchers are not specifically requested to cite the data in the reference list, and no concrete examples are given about data citation. In light of this, it is strongly recommended that greater weight is given by journals to good data citation practice guidance.

4.2 SOCIAL SCIENCE DATA REPOSITORIES AND ARCHIVES

Data repositories and archives as infrastructures giving access to research data also play an important role in promoting and encouraging data citation. As with journals, we analysed the data citation requirements of data repositories and archives that share social science research data and that are located in Switzerland.

To select them, we used the Swiss National Science Foundation (SNSF) list of data repositories (SNSF, 2022). This list gathers commonly used data repositories and archives (27) by the Swiss research community that fulfil the SNSF open research data criteria. In addition, we included the most mentioned data repositories and archives (12) in funded data management plans in the Social Science and Humanities in the latest SNSF monitoring report (Milzow et al., 2020, p. 9). Among these 12 data repositories and archives, two were additional repositories, compared to the list of 27. From the final list of 29 data archives and repositories, 7 are located in Switzerland and can share social science data, and, thus, appear in our review. Among these reviewed data repositories and archives, one is specialised for social science data, two are generalist repositories, and the remaining 4 are institutional repositories (Table 2).

Table 2. Review of Swiss data repository and archive practices regarding data citation.

Data repository or archive		**Citation**		
Name	**Type**	**For each dataset**	**Includes all core components**	**Export RIS (or similar formats)**
FORS Data Service (access to data via SWISSUbase)	Discipline-specific: Social sciences	Yes	Yes, since Dec. 2022 (previously version not always stated)	No
Zenodo	Generalist	Yes	No, version not always stated	No
LORY (access to data via Zenodo)	Institutional: Lucerne's HEIs	Yes	No, version not stated	No
OLOS	Generalist	Yes	No, version not stated	No
Yareta	Institutional: Geneva's HEIs	Yes	No, version not stated	No
BORIS Portal	Institutional: University of Bern	No	-	No
ETH Research Collection	Institutional: ETHZ	No	-	No

Note: Last review done on the 27th of February 2023.

First, we assessed their data catalogues to find out whether, for each dataset, they suggested a specific citation and offered an export service of the citation elements (as suggested on page 7). Table 2 summarises the results. Five repositories provide a citation for each dataset they share: FORS Data Service, LORY, OLOS, Yareta, and Zenodo. Since December 2022, the citations provided by FORS Data Service, which uses the SWISSUbase platform, include all core components for a correct data citation. Citations of data available before this date do not

always include the version. Zenodo's citations vary in their inclusion or not of the dataset version. LORY – Lucerne Open Repository, OLOS and Yareta repositories provide citations without version. The last two data archives and repositories, BORIS Portal, and ETH Research Collection, do not provide any data citations. Regarding the possibility for users to export all necessary information from the data archive or repository and import it to their reference management software (e.g., Zotero, Endnote, Mendley, etc.), through for example a RIS document, this is currently not implemented at any of the repositories or archives. Yet, such a tool would facilitate and improve data citation practice by making the data citation process identical to the citation process of other sources (e.g., research articles).

Second, we have checked whether the data usage policies and licenses adopted by the data archives and repositories request explicitly data to be cited (Table 3). Among the seven repositories and archives, we did not find specific data usage policies or licenses for two of them. Among the other five data archives and repositories, FORS Data Service is currently an exception in the Swiss social science data archive and repository landscape. Indeed, whether shared in open access or in restricted access, all FORS Data Service licenses require data to be cited. Regarding data shared in open access, only Zenodo also requires their data to be cited. FORS Data Service and Zenodo use the Creative Commons licenses for open access data that at least require attribution (CC-BY licenses). The other repositories, BORIS Portal, ETH Research Collection and Yareta, allow the CC0 licence (and other similar licenses outside Creative Commons), which does not require citing the original authors. Regarding restricted access data, generally the data depositors set the conditions for the data access and usage, and, thus, citation requirements. At FORS, the data usage policy for restricted access always require data to be cited. In other data repositories, using a CC-BY licence is encouraged but not mandatory.

Table 3. Data usage policies and licenses requiring data citation.

Data repository or archive	**Policies and licenses requiring data to be cited**	
	Open access	**Restricted access**
FORS Data Service	Yes	Yes
Zenodo	Yes	No
BORIS Portal	No	No, but if not stated otherwise, BORIS portal content must be cited according to its general policy.
ETH Research Collection	No, but use of CC-BY licenses is encouraged.	No, but use of CC-BY licenses is encouraged.
Yareta	No, but use of CC-BY licenses is encouraged.	No, but use of CC-BY licenses is encouraged.
LORY	-	-
OLOS	-	-

Note: Last review done on the 27th of February 2023.

5. PRACTICAL RECOMMENDATIONS TO THE SOCIAL SCIENCES COMMUNITY

Based on the above, the following recommendations can be made for the social science community in Switzerland.

To researchers, students, and any data users:

Recommendation 1 – Cite all sources used in your publication or assignment, both in-text and in the reference list. This concerns both primary and secondary data. Acknowledgments and data availability statements in a footnote or at the end of the text are not sufficient.

Recommendation 2 – Cite the data directly, that is, where the data has been published and is accessible (for published data). Citing a methodological report or a data article is not enough.

Recommendation 3 – While citing the data, make sure that all the core components are included in the citation: data authors, publication year, title, version, data publisher, and persistent identifier. These are the minimal elements to be included. Arrange the citation elements and add extra elements and information based on the requirements of the bibliographic style, journal guidelines, or data archive and repository.

To academic journals:

Recommendation 4 – Scientific journals should always require that all materials used in an article are correctly cited. They should add information and examples on how to cite data in their guidelines for authors. The data citation should at least include the core elements set out above.

To data archives and repositories:

Recommendation 5 – Data archives and repositories should always require that the data they publish are cited and inform their users on correct citation practices. Data citation example(s) should include at least the core elements. In addition, data archives and repositories could facilitate data citation by allowing automatic export of core and additional data citation elements to reference management software. This ensures that the data citation process is identical to the citing process of other sources.

6. FURTHER READINGS AND USEFUL WEB LINKS

Jessop's (2021) guide offers more detailed explanations on why and how to cite data. Parsons et al. (2019) propose a good and concise history of data citation in research, as well as interesting ideas on future best practices. To go further, Dosso and Silvello (2020) discuss citation of complex data like relational databases and how data authors' credit could be fractioned based on their involvement. Also, the short ICPSR[8] (2018) video presenting why data should be cited could be used in trainings to introduce the subject.

[8] ICPSR stands for the Inter-university Consortium for Political and Social Research, https://www.icpsr.umich.edu/web/pages/ (accessed 01.12.2022).

REFERENCES

Columbia University, Health Science Library. (2021). *Citing data sources - Why is it good and how to do it?* https://library.cumc.columbia.edu/insight/citing-data-sources

Cornell University, Research Data Management Service Group. (2022). *Data citation*. https://data.research.cornell.edu/content/data-citation

Cousijn, Helena, Feeney, Patricia, Lowenberg, Daniella, Presani, Eleonora, & Simons, Natasha. (2019). Bringing Citations and Usage Metrics Together to Make Data Count. *Data Science Journal, 18*(1). https://doi.org/10.5334/dsj-2019-009

Dosso, Dennis, & Silvello, Gianmaria. (2020). Data credit distribution: A new method to estimate databases impact. *Journal of Informetrics, 14*(4), 101080. https://doi.org/10.1016/j.joi.2020.101080

Finnish Social Science Data Archive. (2022). *Citing Data*. https://www.fsd.tuni.fi/en/data/downloading-and-using-data/citing-data

FORCE11, Data Citation Synthesis Group. (2014). *Joint Declaration of Data Citation Principles*. FORCE11. https://doi.org/10.25490/a97f-egyk

Heers, Marieke. (2021). *Replication in the Social Sciences*. FORS. https://doi.org/10.24449/FG-2021-00016

IASSIST, Special Interest Group on Data Citation. (2012). *Quick Guide to Data Citation: Identify, Retrieve, Attribute*. https://www.icpsr.umich.edu/files/ICPSR/enewsletters/iassist.html

ICPSR. (2018). *Why Should I Cite Data?* Youtube. https://www.youtube.com/watch?v=jiCZKV-aIC0

Jessop, P. (2021). *Data Citation : A guide to best practice*. Publications Office of the European Union. https://doi.org/10.2830/59387

Late, Elina, & Kekäläinen, Jaana. (2020). Use and users of a social science research data archive. *PLoS One, 15*(8), e0233455. https://doi.org/10.1371/journal.pone.0233455

Milzow, Katrin, von Arx, Martin, Sommer, Cornélia, Cahenzli, Julia, & Perini, Lionel. (2020). *Open Research Data: SNSF monitoring report 2017-2018*. http://doi.org/10.5281/zenodo.3618123

Parsons, Mark A., Duerr, Ruth E., & Jones, Matthew B. (2019). The History and Future of Data Citation in Practice. *Data Science Journal, 18*(1), 52. https://doi.org/10.5334/dsj-2019-052

Silvello, Gianmaria. (2018). Theory and practice of data citation. *Journal of the Association for Information Science and Technology, 69*(1), 6-20. https://doi.org/10.1002/asi.23917

SNSF. (2022). *Open Research Data: Which data repositories can be used?* https://www.snf.ch/en/WtezJ6qxuTRnSYgF/topic/open-research-data-which-data-repositories-can-be-used

Swissuniversities. (2021). *Swiss National Open Research Data Strategy* https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/ORD/Swiss_National_ORD_Strategy_en.pdf

UK Data Service. (2021). *Reasons to cite data correctly*. https://ukdataservice.ac.uk/help/new-users/citethedata

UK Data Service. (2022). *New to using data: Cite data correctly*. https://ukdataservice.ac.uk/learning-hub/new-to-using-data/#cite-data-correctly

University of Geneva. (2022). *Research Data: Cite data*. https://www.unige.ch/researchdata/en/collect-organise/citation

Vannan, S., Downs, R. R., Meier, W., Wilson, B. E., & Gerasimov, I. V. (2020). *Data Sets Are Foundational to Research. Why Don't We Cite Them? - Eos*. https://eos.org/opinions/data-sets-are-foundational-to-research-why-dont-we-cite-them

Yoon, JungWon, Chung, EunKyung, Lee, Jae Yun, & Kim, Jihyun. (2019). How research data is cited in scholarly literature: A case study of HINTS. *Learned Publishing, 32*(3), 199-206. https://doi.org/10.1002/leap.1213

Zhao, Mengnan, Yan, Erjia, & Li, Kai. (2018). Data set mentions and citations: A content analysis of full-text publications. *Journal of the Association for Information Science and Technology, 69*(1), 32-46. https://doi.org/10.1002/asi.23919

Zuiderwijk, Anneke, Shinde, Rhythima, & Jeng, Wei. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PLoS One, 15*(9), e0239283. https://doi.org/10.1371/journal.pone.0239283