



Sharing, storing and preserving research data

Alina Danciu, Cynthia Pedroja

► To cite this version:

Alina Danciu, Cynthia Pedroja. Sharing, storing and preserving research data. Doctoral. Data management, Sciences Po Paris, France. 2019. hal-03921103

HAL Id: hal-03921103

<https://sciencespo.hal.science/hal-03921103>

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

27-28 May 2019

DATA MANAGEMENT

Sharing, storing and preserving research data

Session 4

Alina Danciu | Data manager | Center for Socio-Political Data (CDSP)

Cynthia Pedroja | Project manager in research data management | Library

“Data management is key for success” Nicolas Sauger, April 2019

General frame

Syllabus 

27th May 2019

13:00-14:30 | Session 1 - Introduction to data management and DMPs

14:30-17:30 | Session 2 - Organizing and cleaning data

28th May 2019

09:00-12:00 | Session 3 - Introduction to Security, GDPR Compliance, Intellectual Property and Ethic

12:00-13:00 | Lunch

13:00-16:00 | Session 4 - Sharing, storing, and preserving research data

16:00-17:30 | Session 5 - Conclusion

Digital preservation and trusted repositories

Finding data

1. Begin with the definition of your topic
2. What? **Population** that you want to study
3. When? **Time** you want to study
4. Where? **Space**: Geography or place

What is your search strategy?

1. Search in a **data archive**
2. Identify potential **producers**
3. Turn to the **literature**
4. Ask for help

Evaluate Data

Once you've chosen a data set that you believe will work, take care to carefully evaluate it.

- Is it **appropriate**?
- Does it come from an **authoritative source**?
- Does it fit your **needs**?
- Does it cover your **Where, When, and Who or What** requirements?
- Are you willing to compromise your requirements or manipulate the data to fit your needs?

Always **read the documentation and codebook** to ensure that the analysis you are planning to do really measures what you want it to.



Data sources

Mission of data archives: to embed data into the research lifecycle in such a way that data are **published, shared, discovered and reused**. Trusted domain repositories, such as the CESSDA (Consortium of European Social Science Data) Archives, design their data infrastructures to **follow the FAIR (Findable, Accessible, Interoperable and Reusable) data principles**.

[CESSDA Data Catalogue](#)

Examples of Non-European Data Archives:

- Ontario Data Documentation, Extraction Service and Infrastructure ([Odesi](#))
- Center for Studies on Public Opinion, Brasil ([CESOP](#))
- Australian Data Archive ([ADA](#))
- US Inter-university Consortium for Political and Social Research ([ICPSR](#))



Source: CESSDA Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC.
Retrieved from <https://www.cessda.eu/DMGuide>

Other important data sources (1)

European longitudinal research projects

- [The European Social Survey](#) (ESS-ERIC)
- [Generations & Gender Programme](#) (GGP)



Eurostat

- **The Eurostat data collection** (in aggregate form) can be accessed [here](#). **eurostat** 
- **Eurostat microdata**

Eurostat microdata (including the [European Union Statistics on income and living conditions](#) (European Commission, n.d.a) can be accessed under specific conditions, frequently through some form of secure access (especially in case of confidential data). More information on how to access Eurostat microdata can be found in the publication '[How to use microdata properly](#)' (European Commission, 2018).

Source: CESSDA Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. Retrieved from <https://www.cessda.eu/DMGuide>

Other important data sources (2)

- [OECD iLibrary](#) (Organisation for Economic Cooperation and Development)  OECD
- [UNICEF data](#)  
- [UNdata](#) (United Nations)
- [World Bank Open Data](#)  THE WORLD BANK

Other examples

- [Data resources on aging](#), [political behaviour](#), [migration](#)
- Data resources on [international comparisons](#)

Source : CESSDA Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC.
Retrieved from <https://www.cessda.eu/DMGuide>

How about your data? Important terms

- What is digital preservation and why do we need it ?
- What is long-term archiving and why do we need it?

Tips



Timing is everything!

In data archiving and publishing timing is everything. If you archive or publish your data as soon as data collection ends, your knowledge about your data is still very high. As such, it will take you the least time to prepare your data for deposition while simultaneously guaranteeing the highest possible data quality for future users.

Publish a data paper

For high-quality datasets consider publishing a data paper in a data journal. In this way, you can describe your datasets in more detail which will increase its visibility and chance it will be reused. The data journal does not hold the datasets (they are in a data repository). See '[Promoting your data](#)' for more information on this route.

Archive with the help of an expert: https://youtu.be/66oNv_DJuPc

There is a difference between self-archiving without any help and archiving with the help of an expert. While self-archiving is a quick and easy way to publish data, archiving with the help of an expert will enhance data quality. Expert help is most likely to be available at a trusted domain repository and an institutional repository. Check to see whether that is the case.

Source: CESSDA Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC.
Retrieved from <https://www.cessda.eu/DMGuide>

Promoting your data

How can you attract people to use your data and make them as impactful as possible?

1. **Choose open access**

2. **Licence your data**

Stating clear reuse rights is like having a warm 'Welcome' on the doormat of your dataset. It is an important aspect in making sure your data meet the R (Reusable) in [FAIR data management](#).

3. Always **cite your data** and link your data to scientific publications which are based on this data.

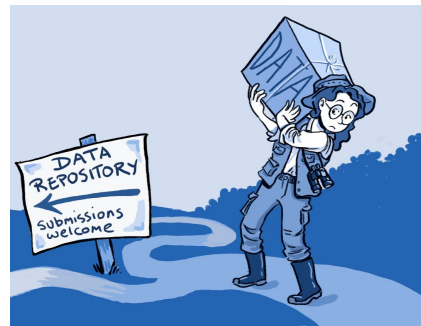


Source: CESSDA Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC.
Retrieved from <https://www.cessda.eu/DMGuide>

4. Publishing your data

There are different ways to publish your data. Your preference may depend on the customs in your discipline or on the expectations of your funder.

According to a [survey by Wiley](#) (2014), the preferred way of publishing data is as **supplementary material of a journal article**. That may change as more data repositories become available, and more journals recommend depositing in them. A **data repository** is a digital archive collecting, preserving and displaying datasets, related documentation and metadata ([OpenAIRE, 2017](#))



Source: CESSDA Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC.
Retrieved from <https://www.cessda.eu/DMGuide>

4.2 Choosing a data repository

A (trusted) domain repository

Use a (trusted) repository already established for your research domain.

The [CESSDA archives](#) are examples of domain-specific trusted repositories. Do note that not all individual datasets may be accepted or only certain types of data (e.g. surveys but not qualitative data). As a general rule, high-quality data with a potential for reuse and that can be publicly shared are submitted to this kind of repositories.

A general purpose repository

If none of the above is available, use a general purpose repository like [Zenodo](#) (n.d.), [Figshare](#) (n.d.) or [Harvard Dataverse](#) (2017). Here you can store, share and register your research data. Do take note that long-term preservation of your data collection isn't always guaranteed. Check the repository in question to find out.

Find your own at re3data.org

Search [Re3data.org](#) (n.d.), a registry of over 1500 research data repositories, to discover other data repositories. You can search by subject, content type and country. In addition, you can select whether you want to search for data archives with a certificate (a trusted repository), with data sets that are available via open access or for data sets that have a persistent identifier.

Source: CESSDA Training Working Group (2017 - 2018). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC.
Retrieved from <https://www.cessda.eu/DMGuide>

Trusted repositories

Definition

- A certification meets a range of issues and standards relating to trusted data and services
- Certifications can be granted by a self-audit and a peer review or by a formal audition by an external neutral institution (ISO)
- Most of the CESSDA archives are certified (Core Trust Seal)

DataSPIRE

Sciences Po's institutional repository

Exercice

<https://demo.dataverse.org/>

The screenshot displays the DataVerse demo interface. On the left, a sidebar contains filters for 'Publication Status' (Draft (2), Unpublished (2)), 'Author Name' (Bianchi Suzanne (1), Gougou Florent (1), Lesnard Laurent (1), Nazio Tiziana (1), Tiberi Vincent (1)), 'Author Affiliation' (Centre de recherches sociales (FORS), UMI Centre études europe Po Paris (1), Observatoire Sociolo (OSC), Sciences Po Université de Californie, Université de Tübingen (1)), 'Subject' (Social Sciences (2)), 'Keyword Term' (Campagnes électorales (1), Famille (1), Mariage (1), Mobilisation électorale (1), Politisation (1)), and 'Deposit Date' (2018 (2)).

The main content area shows '1 to 2 of 2 Results'. The first result is 'Attitudes à l'égard des formes conjugales alternatives' by Lesnard, Laurent, Bianchi Suzanne, Nazio, Tiziana, Voorspoel, Mariëke, 2019, 'Attitudes à l'égard des formes conjugales alternatives', https://doi.org/10.5072/FK2JCKVWR, Demo DataVerse, DRAFT VERSION. The second result is 'Dynamique de mobilisation - vague 18' by Tiberi, Vincent, Gougou Florent, 2019, 'Dynamique de mobilisation - vague 18', https://doi.org/10.5072/FK2JCKVWR, Demo DataVerse, DRAFT VERSION.

The selected result, 'Dynamique de mobilisation - vague 18', is shown in detail. It includes a description: 'La majeure partie des connaissances produites en sociologie politique quantitative provient des enquêtes conduites au moment des élections. L'étude longitudinale DYNAMOIS est proposée par Florent Gougou, Vincent Tiberi et vingt-cinq politistes français. Elle se distingue des enquêtes classiques en couvrant aussi bien les périodes électorales que les périodes ordinaires. Ce dispositif vise à mesurer le poids relatif des facteurs de court terme (effet des campagnes électorales, notamment) et des facteurs de long terme sur le vote des électeurs. La dix-huitième et dernière vague de DYNAMOIS s'est déroulée en décembre 2017. Il s'agit d'une interrogation des panelistes hors périodes électorales qui se traduit par la passation d'un questionnaire « noyau ». Aucun module spécifique n'a été prévu pour cette vague. (2018-07)'. It also lists the subject as 'Social Sciences', the keyword as 'Elections, Vote, Politisation, Mobilisation électorale, Campagnes électorales, Temps ordinaire', and the related publication as 'Florent Gougou, Vincent Tiberi, équipe Dynamob, équipe ELIPSS [auteur] Dynamiques de mobilisation - vague 18 (décembre 2017) [fichier électronique], Fondation Nationale des Sciences Politiques (FNSP) [producteur], Centre de Données Socio-Politiques (CDS) [éditeur], Version 1'.

At the bottom, there are tabs for 'Files', 'Metadata', 'Terms', and 'Versions'. A search bar is also present with the text 'Search this dataset...' and a 'Find' button.

Importance of metadata

Impact of a mistake : [Case Dominique Cardon](#)

Make the most of your data, deposit it!

Why?

Preserving, sharing, citing your research data.

Where?

Trusted repository.

DataSPIRE institutionnal repository of Sciences Po

- le dépôt de données pour préserver et sécuriser
- l'attribution d'un identifiant unique (DOI) pour pérenniser
- la citation automatique pour référencer
- la maîtrise des droits d'accès
- l'exploration des données et de leur contexte (métadonnées)
- de lier données et publications



Ouverture DataSPIRE : automne 2019

Chef de projet : Geneviève MICHAUD (CDSP)



The screenshot shows the DataVerse website interface. At the top, there's a navigation bar with 'Datasets (2)', 'Files (2)', and 'Unpublished (2)'. Below this, a search bar displays '1 to 2 of 2 Results'. Two search results are listed:

- Attitudes à l'égard des formes conjuguées alternatives** (25 av. 2019) by LESNARD Laurent, BIANCHI Suzanne, NADIO Tidana, VOORPOSTEL Mariëtte. Description: L'enquête ATAPA interroge l'évolution de la définition même de la famille au moment où ont lieu, en France et plus généralement en Europe, de nombreuses transformations dans la formation des types d'union (cohabitation, PACS, etc.) et de parentalité (homoparentalité, hors-mariage...).
- Dynamique de mobilisation - vague 18** (25 av. 2019) by TIBERJ Vincent, GOUSSOU Florent. Description: La majeure partie des connaissances produites en sociologie politique quantitative provient des enquêtes conduites au moment des élections. L'étude longitudinale DYNAMOB est proposée par Florent Goussou, Vincent Tiberj et vingt-cinq politistes français. Elle se distingue des enquêtes classiques en couvrant aussi bien les périodes électorales que les périodes ordinaires. Ce dispositif vise à mesurer le poids relatif des facteurs de court terme (effet des campagnes électorales, notamment) et des facteurs de long terme sur le vote des électeurs. La dix-huitième et dernière vague de DYNAMOB s'est déroulée en décembre 2017. Il s'agit d'une interrogation des panelistes hors périodes électorales qui se traduit par la passation d'un questionnaire « royaume ». Aucun module spécifique n'a été prévu pour cette vague (2018-07).

Below the search results, there's a detailed view for 'Dynamique de mobilisation - vague 18'. It includes fields for Description, Subject (Social Sciences), Keyword (Élections, Vote, Politisation, Mobilisation électorale, Campagnes électorales, Temps ordinaire), and Related Publication (Florent Goussou, Vincent Tiberj, équipe Dynamob, équipe ELIPSIS [auteurs]. Dynamiques de mobilisation - vague 18 (décembre 2017). [Fichier électronique]. Fondation Nationale des Sciences Politiques (FNSP) [producteur]. Centre de Données Socio-Politiques (CDSP) [diffuseur]. Version 1).

DataSPIRE démo (copie d'écran)

PUBLIER POUR ÊTRE LU ET VU OU COMMENT VALORISER SES TRAVAUX DE RECHERCHE PENDANT ET APRÈS SA THÈSE

3 octobre 2019, 9h30-17h30

Cette journée vise à sensibiliser les doctorants à plusieurs aspects de la publication scientifique tout en leur donnant un espace adapté de discussion avec des intervenants professionnels de l'institution et des experts invités.

- Pourquoi et comment diffuser sa thèse en ligne, comment publier sa thèse, etc.
- Comment publier dans une revue à CL et notamment en anglais, pourquoi ouvrir un carnet de recherche, comment médiatiser ses travaux, etc.

More information and registration: <https://formations-bibliotheque.sciencespo.fr>