

## Chapter 6

# Archive & Publish

## Contents

Main take-aways .....	156
6.1 Towards archiving & publication .....	157
6.2 Selecting data for publication .....	159
6.3 Data publishing routes .....	160
6.4 Publishing with CESSDA archives .....	164
6.4.1 Citing your data .....	165
6.4.2 Licensing your data .....	167
6.4.3 Access categories .....	169
6.5 Promoting your data .....	172
6.6 Adapt your DMP: part 6 .....	174
Sources and further reading .....	175

## Main authors of this chapter

Sonja Bezjak, Slovenian Social Science Data Archives (ADP)

Irena Vipavc Brvar, Slovenian Social Science Data Archives (ADP)

## CITATION

CESSDA Training Team (2017 - 2019). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. DOI: 10.5281/zenodo.3820473

Retrieved from <https://www.cessda.eu/DMGuide>

## LICENCE



The Data Management Expert Guide by CESSDA ERIC is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. All material under this licence can be freely used, as long as CESSDA ERIC is credited as the author.

## Introduction



High-quality data have the potential to be reused in many ways. Archiving and publishing your data properly is at the core of making your data FAIR and will enable both your future self as well as others to get the most out of your data.

In this chapter, we venture into the landscape of research data archiving and publication. We will guide you in making an informed decision on where to archive and publish your data in such a way that others can properly access, understand, use and cite them.

## Main take-aways

After completing your journey through this chapter on archiving and publishing data, you should:

- » Understand the difference between data archiving and data publishing;
- » Be aware of the benefits of data publishing;
- » Be able to differentiate between different data publication services (data journal, self-archiving, a data repository);
- » Be able to select a data repository which fits your research data's needs;
- » Be aware of ways to promote your research data publication;
- » Be able to answer the DMP questions which are listed at the end of this chapter and adapt them to your own DMP.

## 6.1 Towards archiving & publication

This chapter in the tour guide is all about securing your research data's future for the following purposes:

» **Archiving data for future reference**

Research data archiving is about storing and preserving research data for the long term. When you archive your data, you make sure you can read and access the data later on. You can then also allow access to others for verification purposes when such a request arrives. In all cases, you should store your data safely, in a suitable file format, with adequate documentation.

» **Publishing data for reuse**

To make your data reusable for purposes beyond the one for which you collected them, you should publish your data. Publishing your data is the act of publicly disclosing the research data you have collected, making them findable, accessible and reusable.

The incentives that motivate interviewed researchers to archive and publish their research data (Van den Eynden & Bishop, 2014; Hahnel et al., 2017) fall into four main categories:

### Career benefits

Data publication may lead to increased visibility, reuse and citation and therefore recognition of scholarly work.

A number of studies show the impact of data publication on citation rates. Articles for which the underlying data is published are more frequently cited than articles for which this is not the case. Studies from social science (Pienta, Alter & Lyle, 2010), genetics (Piwowar and Vision, 2013; Botstein, 2010), astronomy (Henneken and Accomazzi, 2011; Dorch 2012) and oceanography (Sears 2011, Belter 2014) confirm this effect.

Be aware that whenever you use the published data you are obliged to cite them. For more information see the paragraph on data citation.

### Scientific progress

Data archiving and publication has direct benefits for the research itself (more robust), for the discipline and for science in general by enabling new collaborations, new data uses and establishing links to the next generation of researchers.

A tweet (ESS ERIC, 2017) from the European Social Survey (n.d.) is just one of the many, many examples of how sharing high-quality datasets leads towards new insights. The European Social Survey is widely accessible and used by many researchers.

### Norms

Norms of the project, research group, and/or discipline may determine whether a researcher is prone to publish his/her data. Overall, the openness of research data is at the heart of scientific ethics as is illustrated by the quote below.

*Sociologists make their data available after completion of the project or its major publications, except where proprietary agreements with employers, contractors, or clients preclude such accessibility or when it is impossible to share data and protect the confidentiality of the data or the anonymity of research participants (e.g., raw field notes or detailed information from ethnographic interviews) | American Sociological Association (1999).*

## External drivers

External drivers like research data management policies from research funders and publishers have a significant influence on data archiving and publication:

### » Funders

Some funders consider costs related to data archiving and publication eligible and require a DMP. For a list of funder requirements see the 'European diversity in funder requirements' section of this tour guide.

### » Publishers

Scientific journals are increasingly adopting data availability policies that advise or even request authors of manuscripts to make the research data, on which a manuscript is based, available. For example, PLOS One says in its data availability statement:

*All data and related metadata underlying the findings reported in a submitted manuscript should be deposited in an appropriate public repository unless already provided as part of the submitted article. Repositories may be either subject-specific (where these exist) and accept specific types of structured data, or generalist repositories that accept multiple data types, such as Dryad | PLOS One (2014a).*

In the coming paragraphs, the main focus will be on securing high-quality datasets for the future by combining data archiving and data publishing.

## 6.2 Selecting data for publication

Should you publish your data or shouldn't you? And if so, which part of it? Sometimes this question is rather straightforward to answer. For example, due to research funders' demands. Or because you strongly believe it is not up to you to withhold any latent knowledge hidden in your data from future researchers.

Even so, not all data are created equal and data publishing does involve an investment of time and money. Some datasets have a more obvious reuse potential than others. By following the arguments below (Tjalsma and Rombouts, 2011) you can see for yourself whether (part of) your dataset is an obvious candidate for data publication.

### » **Does your dataset have reuse potential?**

Does your data have potential value in terms of reuse, national/international standing and quality, historical importance, uniqueness (the data contain non-repeatable observations), originality, size, scale, costs of data production or innovative nature of the research? Could you foresee that secondary analyses on your data would benefit science? If your answer to any of these is yes, your dataset has serious reuse potential.

### » **Is your dataset reusable?**

To be suited for re-use your dataset should be functionally usable. Can the data be read and used? Are metadata available and are they sufficient to enable future users to understand your data? Are there any legal objections which prevent the data from being published? If you are sure you got these practicalities covered and your data have potential value in terms of reuse, you are 'good to go'.

If your data are not 'good to go' at this point but do have re-use potential, do not worry. Most of the time it is not too late to document your data properly and address other issues like gaining consent for sharing retrospectively (see 'Informed Consent').

## 6.3 Data publishing routes

*It is expected that a Data Publication will ensure that data will potentially be considered as a first-class research output*  
| Knowledge Exchange (2013).

For a dataset to “count” as a publication, it should follow a similar publication process to an article (Brase et al., 2009) and should be:

- » Properly documented with metadata;
- » Reviewed for quality;
- » Searchable and discoverable in catalogues (or databases);
- » Citable in publications.

The authors of a report from Knowledge Exchange (Knowledge Exchange, 2013) define this type of data publication as ‘Publishing with a capital P’ and compare it with ‘publishing with a small p, meaning that researchers publish their data files on a website somewhere. Publishing with a small “p” means that there are no guarantees that the data will be there after some time or that the files will not get corrupted.

### Five routes

There are different ways to publish your data. Your preference may depend on the existing practices 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 scientific journals recommend depositing in them. A data repository is a digital archive collecting, preserving and displaying datasets, related documentation, and metadata (OpenAIRE, 2017)

In the comparisons below we show five ways of publishing your data, together with their advantages and disadvantages.

#### Journal supplementary material service

##### Advantages

- » Most likely to comply with the journal or publisher’s requirements;
- » Data readily available alongside published findings.

##### Disadvantages

- » May be costly;
- » May claim copyright over the data;
- » May keep data behind a subscription wall;
- » Unlikely to offer a data repository’s functionality or long-term solution;
- » May not apply user-friendly or preservation formats;
- » More likely to accept subsets rather than complete datasets.

## **Institutional data repository**

### **Advantages**

- » Most likely to accept any data of value, especially if no suitable home can be found for it elsewhere, and to ensure that policy requirements for long-term access are met;
- » Researchers may trust such a repository more readily;
- » Possibly no charge for the data deposit;
- » May make your data visible via dissemination and promotion.

### **Disadvantages**

- » May not offer sustainable long-term access to your data collection;
- » Might not have sufficient expertise in data and metadata standards needed for long time preservation and access.

## **General purpose repository**

### **Advantages**

- » Most likely to offer useful search, navigation and visualisation functionality;
- » Reach a wider audience of potential users;
- » Accepts a wide range of data types;
- » Suitable for cross-disciplinary data.

### **Disadvantages**

- » Requires scrutiny of terms and conditions to ensure consistency with your funder, journal or institution's policies on cost recovery, copyright/IP, and long-term preservation;
- » No editorial control over quality of deposited materials;
- » In most cases, only simple metadata is available, which is usually not enough for reuse.

## **Domain specific data repository**

### **Advantages**

- » Offers specialist domain knowledge and data management expertise, e.g. to create a catalogue record and documentation;
- » Likely to accept complete datasets (and not only the part of the dataset on which a publication is based);
- » May make your data visible via dissemination and promotion.

### **Disadvantages**

- » Likely to be selective about what kind of data they accept.

## Trusted domain specific data repository

### Advantages

- » Offers specialist domain knowledge and data management expertise, e.g. to create a catalogue record and documentation;
- » More likely to accept complete datasets;
- » Provides preservation and curation to community standards, e.g. file formats migration;
- » Ability to control access of (sensitive) personal data;
- » May handle data re-use queries;
- » May make your data visible via dissemination and promotion.

### Disadvantages

- » Most likely to be selective about what kind of data they accept;
- » May charge for data publishing;
- » Requires advance planning of the effort needed to meet high standards for metadata and documentation.

## Choosing a data repository

*There are hundreds of repositories worldwide. Some cater a specific research domain, while others are general-purpose repositories. They may be called something other than a repository, for example, a data centre or an archive | Whyte (2015).*

If you decide to choose a data repository for publishing your data, which data repository should you choose? Sometimes the repository is already determined by your funder or another external party. But if the choice is yours to make, you may consider following the order of preference in the recommendations by OpenAIRE (2016b):

### » 1: 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.

### » 2: An institutional or recommended data repository

If a domain repository is not available, use an institutional research data repository. If such a repository is not available, you may follow the guidelines of your university or publisher. Some publishers provide lists of recommendations e.g., PLoS ONE (2014b) recommended repositories.

### » 3: 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 is not always guaranteed. Check the repository in question to find out.

### » 4: 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.



## Expert 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 deposit 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. This way, you can describe your datasets in more detail, which will increase their visibility and chances of being re-used. The data journal does not hold the datasets (they are in a data repository). See 'Promoting your data' for more information on this route.

### **Choose between self-archiving and expert help**

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.

## 6.4 Publishing with CESSDA archives

For high-quality data with a potential for reuse, we recommend you to assure long-term access by publishing your data with a trusted repository, like many of the CESSDA archives. CESSDA archives aim to make the research data accessible with as few restrictions as possible, while at the same time protecting (sensitive) personal data from inappropriate access.

### CESSDA archives per country

If you decide to publish your data to one of the CESSDA archives you will have to invest some time and effort to prepare your data. If research data management is a vital part of your work, then the majority of work has already been done on your way.

See the CESSDA website for an overview of national CESSDA archives:

<https://www.cessda.eu/About/Consortium>

### Added benefits of a CESSDA repository

As opposed to self-archiving your dataset, publishing your dataset at a CESSDA archive has a great advantage of having expert help within reach. CESSDA research data management experts can help you to increase the comprehensibility, visibility, findability, reusability, longevity and the overall quality of your datasets in numerous ways.

For a list of benefits and requirements, consult the online version of this guide:

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/6.-Archive-Publish/Publishing-with-CESSDA-archives>

### Do you want to dive in deeper?

For data licensing, data citation and data access we have prepared additional information in the following subchapters.

## 6.4.1 Citing your data

*Persistent identifiers ensure future access to unique published digital objects, such as a text or data set. Persistent identifiers are assigned to data sets by digital archives | American Sociological Review - Submission Guidelines (Sage Publishing, 2017).*

For data products to be uniquely identifiable and attributable to their data creators two types of identifiers are recommended:

### » **A persistent identifier (PID) to your dataset**

The publication of data sets is becoming more and more important as a citable contribution to research. To become citable, you need to make sure that your datasets gets a unique, persistent identifier. The Digital Object Identifier (DOI) is a well-known identifier in academia. Having a PID is an important aspect of making sure your data meets the F (Findability) and A (Accessibility) in FAIR data management.

### » **A persistent author identifier**

To make your research results even more connected you can create your personal persistent author identifier. The ORCID iD provides such a persistent digital identifier, distinguishing you from every other contributor and supporting automated linkages among all your professional activities. By creating and using an ORCID iD you will be able to present all of your - growing - work through one channel.

## Citing new data types

Citing rapidly changing data is also challenging. The Data Cite organization has published a recommendation regarding citing new data types. There is the possibility to cite the continuously updated dataset and only add an access date and time to the citation. However, this means that the citation does not result in access to the resource as cited when it was changed in the meantime. This limits reproducibility of the work that uses this form of citation. Another option is to cite a specific "snapshot" (i.e., a copy of the entire dataset made at a specific time) but this requires unique identifiers for each version/snapshot of data.

### Data Citation and impact

*Data citation is the practice of providing a reference to data in the same way as researchers routinely provide a bibliographic reference to other scholarly resources | Australian National Data Service (n.d).*

The impact of your research may be determined by a wide range of research outputs such as data sets, software, blog posts, presentations, tweets, etc. Being able to cite such research outputs is important for building a culture where all types of research outputs count. In the video (Research Data Netherlands, 2014) linked below data citation and the role of persistent identifiers is explained:

<https://www.youtube.com/watch?v=PgqtiY7oZ6k>

## Expert tips



### **1. Deposit your data in a data repository**

When you deposit your data in a (trusted) data repository, a persistent identifier to your data sets is often automatically assigned.

### **2. Register for an ORCID iD**

Registering for an ORCID iD is easy. Do it now (ORCID, n.d.)! Or first have a look at this video (Vanhaverbeke, 2017) in which other researchers state how having an ORCID iD benefits them.

### **3. Check how FAIR your data are**

Want to know how FAIR your data are? Have a look at the checklist by Jones and Grootveld (2017).

### **4. Include persistent identifiers as a variable**

Include the persistent identifier to your dataset as a variable in your data file. For example, the database from the ISSP 2015 on Work Orientations (GESIS, n.d.) includes the following variable: name of the variable: DOI; variable label: "Digital Object Identifier". It has the same value for all the cases: doi:10.4232/1.12848. The link goes directly to the metadata in the GESIS data archive.

## 6.4.2 Licensing your data

If you publish your data in a data repository of your choice, a licence agreement will be applied to your data. A licence agreement is a legal arrangement between the creator/depositor of the data set and the data repository, signifying what a user is allowed to do with the data. Stating clear re-use rights is like having a warm 'Welcome' on the doormat of your dataset. It is an important aspect of making sure your data meet the R (Reusable) in FAIR data management.

To make re-use as likely as possible we advise you to choose a licence which:

- » Makes data available to the widest audience possible;
- » Makes the widest range of uses possible.

### About Creative Commons licences

The main attributes of using Creative Commons (2017) licences for the licensing of data, datasets, and databases (Korn and Oppenheim, 2011) are:

- » The ease of use of the licences;
- » The widespread adoption of the licences;
- » Their flexibility;
- » Their availability in human-readable and machine-readable forms allowing both researchers and computers to immediately know what they are allowed to do with your data;
- » The chance that your data are reused.

There are 7 licences for which the details are given in the table below (inspired by Foter, 2015):

Licence	Can I copy & redistribute the work?	Is it required to attribute the author?	Can I use the work commercially?	Am I allowed to adapt the work?	Can I change the licence when redistributing?
CC0	Y	N	Y	Y	Y
CC BY	Y	Y	Y	Y	Y
CC BY-SA	Y	Y	Y	Y	N
CC BY-ND	Y	Y	Y	N	Y
CC BY-NC	Y	Y	N	Y	Y
CC BY-NC-SA	Y	Y	N	Y	N
CC BY-NC-ND	Y	Y	N	N	Y

Do note that a CC licence cannot be revoked once it has been issued.

The licence you are allowed to apply may be determined or limited by the data repository of your choice. An example is given in the following box.

## Data licences at the Slovenian Social Science Data Archives

The Slovenian Social Science Data Archives (ADP, 2017b) allows you to choose between three types of Creative Commons licenses (ADP, 2017c):

- » CC-BY  
Users:
  - » Are free to share — copy and redistribute the material in any medium or format;
  - » Are free to adapt — remix, transform, and build upon the material;
  - » May use the data sets for any purpose, even commercially.
- » CC-BY-NC  
Users:
  - » Are free to share — copy and redistribute the material in any medium or format;
  - » Are free to adapt — remix, transform, and build upon the material;
  - » May not use the data sets for commercial purposes.

Both licenses have the condition of Attribution. A user must give appropriate credit (Creative Commons, n.d.a).

Recently, the ADP also gives the possibility to choose a Creative Commons Zero License (in short: CC0 or CC Zero Waiver (Creative Commons, n.d.b.)). With this licence, the depositor waives all rights to the data.

## Considerations in choosing a licence

If you only consider your own benefit, you might choose a licence for which attribution is required. What you may not realise is that when such data is blended with similarly licensed data this may lead to impracticalities of required attribution (Dodds, 2014) whenever the data is reused. To facilitate the release of datasets and databases into the public domain, Creative Commons created the CC0 licence.

CC0 is the only truly open Creative Commons licence. The copyright owner waives all its rights, including the database right and the right to be identified as the creator.

Although CC0 can be used to prevent attribution stacking, attribution can be important as a means of recognising both the source and the authority of the data. To acknowledge this right, the use of CC0 can include the publishing of non-binding suggestions for best practices in attribution.

There will be circumstances in which CC0 is inappropriate, due to specific risks that might arise for the licensor and perhaps subsequently also for any users. E.g. when:

- » Datasets containing (sensitive) personal information are deposited for which consent has not been cleared (see the chapter on protecting data);
- » Permission of the copyright holder has not been sought;
- » The rights holders are unknown or cannot be traced (orphan works).

In these cases, licences that place 'some' restrictions upon the user, such as those with an "ND" (No derivatives) and/or "NC" (Non-Commercial) might be more appropriate.

## Tips for choosing a licence

### 1. Be sure who owns the data

Remember you can only archive and publish data you own (or if you have permission).

### 2. Use the licence selector

Choose an appropriate licence for your datasets with this licence selector (n.d.).

### 6.4.3 Access categories

Publishing data in a data repository does not automatically make them openly accessible. (Sensitive) personal data can still be protected by limiting access to the data. Access controls can permit control down to an individual file level, meaning that mixed levels of access control can be applied to a data collection.

Many data repositories operate a three-tiered approach to data access:

- » **Open access**

Data that can be accessed by any user whether they are registered or not. Data in this category should not contain personal information unless consent is given (see 'Informed consent').

- » **Access for registered users (safeguarded)**

Data that is accessible only to users who have registered with the archive. This data contains no direct identifiers but there may be a risk of disclosure through the linking of indirect identifiers.

- » **Restricted access**

Access is limited and can only be granted upon request. This access category is for the most sensitive data that may contain disclosive information. Restricted access requires the long-term commitment of the researcher or person responsible for the data to handle the upcoming permission requests.

- » **Embargo**

Besides offering the opportunity for restricted access 'for eternity' most data repositories allow you to place a temporary embargo on your data. During the embargo period, only the description of the dataset is published. The data themselves will become available in open access after a certain period of time.

Access conditions may differ slightly between data repositories. In the boxes below, two examples are given.

#### Access regulation of the Slovenian Social Science Data Archives (ADP)

At the Slovenian Social Science Data Archives (ADP) access to data and accompanying materials is determined in the Policy of Digital Preservation (ADP, 2017b). The types of access in ADP are the following:

- » **Open Access**

Users may freely access the catalogue of the ADP, study metadata and research data of a limited range of studies without registration. Nonetheless, the use of data and accompanying materials is limited by the legislation, the social sciences and institutional ethical standards and copyright.

- » **Standard Access**

Standard access includes the possibility to access most of the research data in the ADP. In order to obtain standard access, users need to fill in the Registration form to access materials in the Catalogue of the ADP. The users need to identify, define the terms of use of research data and comply with the General Provisions of and Terms of Use of the ADP. Research data that may be accessed by a standard registered user, are fully anonymized. These are the so-called Public Use Files (PUF).

- » **Special Conditions Access**

Some data sets are only accessible under special conditions. In order to gain access, a special permission from the original authors is needed. For example when:

- » **Data are not fully anonymized**

In this case, additional protection is required. Such files are called Scientific Use Files (SUF).

**» Embargo**

The authors place an embargo on access and decide that the datasets will only be available after a certain period of time, for example, after 6 months.

**» Limited availability**

The dataset is available only to the person/institution, ordering the study, or to the original authors.

If a user wants access to files in the Special Conditions Access section, not only a regular registration form (Standard Access) should be filled in but also an additional one which is called: "Application for access to materials on request". The Commission for the protection of Confidentiality carefully inspects such applications and decides on the possibility of access to the requested study data.

Types of possible special access are:

**» Access through a safe connection****» Access in a safe environment**

If the data are especially sensitive the user may be granted access to it only in a safe room of the data archive. These are the so-called Secure Use Files (ScUF) that the user may access only after signing a special contract, determining the rights and obligations of use of the requested research data.

**Access regulation at DANS, the Netherlands**

All research data at DANS are stored in and made available by its online repository EASY (DANS, 2017b). A licence agreement is always agreed between DANS and the depositor of the dataset: the person or organisation depositing a dataset in EASY who is normally the rights holder. One of the most important parts of this licence agreement is the access category by which the access to the dataset can be specified.

DANS supports the Open Access movement. This means that DANS encourages research data and publications to be made freely available as much as possible, without any restrictions. However, substantiated reasons exist why research data is not, or not immediately, freely accessible. This can be due to the presence of personal data or a temporary embargo on data due to an impending PhD thesis or other publication, contract obligations with third parties, etc. DANS, therefore, provides along with open access, the possibility of restricted access to research data.

EASY offers two Open Access categories and one Restricted Access category. The access categories are:

**» Open Access** (CC0 Waiver, Creative Commons, n.d.b.)

The dataset is, without any restriction, made available to all EASY users, both registered and unregistered, in accordance with the conditions of the Creative Commons Zero Waiver.

**» Open Access for Registered Users**

The dataset is only made available to all registered EASY users. Any existing copyrights and/or database rights are respected.

**» Restricted Access**

The dataset is only made available to those registered users that have obtained permission from the rights holder.

Datasets containing personal data are mostly placed in the category of Restricted Access. Some datasets with personal data are made available in the Open Access categories. This is, however, only possible when explicit informed consent has been given by the persons involved. This is quite often the case with Oral History interviews (DANS, 2012). Besides from this open category, sensitive data can only be accessed by authorised users whose identities have been checked and who may be required to also sign



special, additional, conditions of use.

### **Open metadata for (sensitive) personal data**

Even if personal data cannot be published in open access, it is always possible to publish the metadata that belong to this dataset. Openly publishing metadata is, in fact, the only way to make such datasets discoverable.

Trusted data repositories are dedicated to increasing the discoverability of your data sets. Therefore, metadata are always freely accessible in any of the CESSDA archives. That means that:

1. No registration is needed for searching in the metadata;
2. No registration is needed for harvesting the metadata (e.g. by search engines).

Metadata of sensitive datasets should never contain confidential or identifying elements or characteristics, like names.

When someone finds a dataset under restricted access (most likely because they contain (sensitive) personal data), he or she can submit an access request to the rights holder. If this is granted, the dataset will be available for download by this user. Even then the use is restricted. The user is not allowed to make the personal data of this data set public and can only refer to the data in an anonymised way.

### **Access control strategy**

When choosing an access category, consider the following:

- » Does the data contain identifiable information?
- » Can the information in this data collection be linked with anything in another data collection which might lead to participant's identities being disclosed?
- » What did participants consent to?
- » If 'restricted access' is to be chosen who will manage the access to this request?

## 6.5 Promoting your data

How can you attract people to use your data and make them as impactful as possible? Consider promoting reuse of your data in one of the following ways:

### Choose open access

If you deposit your data in a data repository, choose open access. If researchers can easily access your data, it is more likely that your data will be re-used and have an impact on their work.

### License your data

Licensing your data is a prerequisite for data impact. If researchers are unclear about what they are allowed to do with your data, they might not use it at all (see 'Licensing your data').

### Always cite your data

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

How to cite data: an example

The following dataset which holds data on studying migrations patterns in the Summer Olympics between 1948 and 2012 covers approximately 40,000 athletes and contains information on the country they represented as well as their country of birth. According to the data repository, which holds this open access dataset, it should be cited as:

Reference: Jansen, J. (Erasmus University Rotterdam) (2017): Foreign-born Olympic athletes 1948 - 2012. DANS. <https://doi.org/10.17026/dans-2xf-pyqp>

Also, see the data citation paragraph in this tour guide.

### Publish in a data journal

Consider publishing an article in a peer-reviewed data journal. Data journals are designed to comprehensively document and publish deposited datasets and to facilitate their online exploration. Recommendations for such journals for social sciences and humanities are:

Research Data Journal for the Humanities and Social Sciences (RDJ, Brill, 2017);

Journal of Open Psychology Data (JOPD, Ubiquity Press, n.d.a);

Journal of Open Archaeology Data (JOAD, Ubiquity Press, n.d.b);

Open Health Data (Ubiquity Press, n.d.c.).

Tip Read this blogpost: Introducing the 'data paper' in the Research Data Journal for the Humanities and Social Sciences (Moody, 2017).

### Teach with your dataset

Consider preparing a lecture using your datasets (or that of others) and prepare video tutorials on how to use the dataset.

### Datasets for training purposes: easySHARE

The Survey of Health, Ageing, and Retirement in Europe (SHARE-ERIC, 2017a) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks. Surveys are organised bi-annually since 2004. SHARE currently covers 27 European countries and Israel (SHARE-ERIC, 2017b).

The SHARE database is easily accessible to the entire research community; data from the SHARE Waves 1 to 6 are available since 2017. A longitudinal data set "easySHARE" (SHARE-ERIC, 2017c) has been

created especially for training purposes. It contains only selected variables merged into a single data file. It is more user-friendly than the complete set of SHARE panel data.

### **Choose a data repository which promotes your data**

You can promote your own data. In addition, you can choose a data repository which promotes data for you.

For examples of how CESSDA archives promote your data, see the online version of this guide:

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/6.-Archive-Publish/Promoting-your-data>

### **Grow your data's impact with altmetrics**

Altmetrics, or 'alternate metrics', are alternative parameters which measure the impact of your research. More and more, research data and software code are shared in data repositories and quoted in publications. More and more repositories attach a DOI - a persistent identifier - to such datasets, allowing to count how often a dataset:

- » Has been cited;
- » Has been viewed or downloaded;
- » Has been stored in online literature management systems;
- » Is listed in online news media or social media.

After you have uploaded your dataset to a data repository which adds DOIs, consider to:

- » Write a blog post or an article about your data publication;
- » Tweet about it;
- » Write about it on Facebook.
- » Etc.

Do not forget to always cite your dataset, everywhere (in your publications, in blog posts, in social media, etc.). It is the only way to keep track of how your data is used, viewed, liked. See 'Data citation' for information on how to cite your data.

## **Tracking data publications**

Data publications can be tracked by (Ball and Duke, 2015):

Citation-based metrics	Altmetrics-based metrics
<ul style="list-style-type: none"> <li>» DataSearch (Elsevier, 2017) Searches data repositories, including figures/tables and has a preview option so you can judge whether the data are useful.</li> <li>» DataCite (n.d.) Searches datasets that have been given a DOI.</li> <li>» Data Citation Index (Clarivate Analytics, 2017a). (Web of Science - licensed database with paid access only) Searches the metadata of the datasets for connected data repositories (Clarivate Analytics, 2017b).</li> </ul>	<ul style="list-style-type: none"> <li>» ImpactStory (n.d.)</li> <li>» PlumX (Plum Analytics, 2015)</li> <li>» Altmetric (n.d.)</li> <li>» Data repositories (Downloads and views counts)</li> </ul>

## 6.6 Adapt your DMP: part 6



This is the sixth 'Adapt your DMP' section in this tour guide. To adapt your DMP, consider the following elements and corresponding questions:

### Deposit your data

- » Will the data you produce and/or used in the project be usable by third parties, in particular after the end of the project?
- » Which data and associated metadata, documentation and code will be deposited?
- » What methods or software tools are needed to access the data?
- » Is documentation about the software needed to access the data included?
- » Is it possible to include the relevant software (e.g. in open source code)?
- » What data quality assurance processes will you apply?
- » Will the application of a persistent identifier to your data be ensured?

### Deposit timing and duration

- » When will your data be made available for re-use? Is there an embargo period?
- » How long does the data need to be retained? For how long should the data remain reusable?

### Access category

- » How will the data be made available? What access category will you choose?
- » When thinking about access categories consider the following:
- » What did the participants consent to?
- » Does the data contain anything sensitive?
- » Can the information in this data collection be linked with anything in another data collection which might lead to participant's identities being disclosed?
- » If 'restricted access' is to be chosen who will manage the access to this request?

### Data licensing

- » How will your data be licensed to permit the widest re-use possible?
- » Have you considered which kind of licence is appropriate for sharing your data and what, if any, restrictions there might be on re-use?
- » If you are purchasing or re-using someone else's data sources have you considered how that data might be shareable, for example negotiating a new licence with the original supplier?

For easy reference, we have put together a list of DMP-questions for all chapters in this tour guide. You can view and download the checklist as pdf (CESSDA, 2018a) or editable form (CESSDA, 2018b), and keep them as a reference while you are studying the contents of this guide.

## Sources and further reading

Please see the online version of this guide.