



# Introduction to Data Management

---

## Archiving and discovering data

*Ricarda Braukmann (DANS)*

17th June 2021



# A closer look at the DMEG

**Data Management Expert Guide** ▾

- 1. Plan >
- 2. Organise & Document >
- 3. Process >
- 4. Store >
- 5. Protect >
- 6. Archive & Publish >
- 7. Discover >
- 8. Contributors >

Monday's presentation



# A closer look at the DMEG

<b>Data Management Expert Guide</b>	▼
1. Plan	>
2. Organise & Document	>
3. Process	>
4. Store	>
5. Protect	>
6. Archive & Publish	>
7. Discover	>
8. Contributors	>

This presentation



# Archive & Publish



# Archive & Publish



# Archive & Publish

- Why would you archive and publish your data?

## Career benefits

Scientific progress

Norms

External drivers

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](#).






# Archive & Publish

- Why would you archive and publish your data?

Career benefits   **Scientific progress**   Norms   External drivers


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.

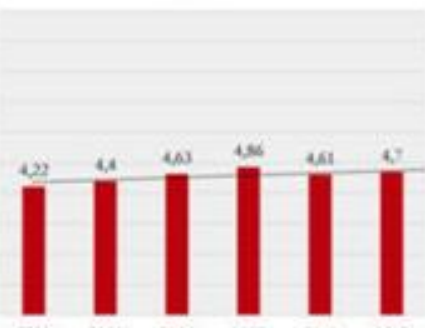
 **ESS ERIC**  
@ESS\_Survey Follow

A new report produced by @KU\_Leuven academics measures public opinion towards migration in Belgium using ESS data  
[bit.ly/2xzaTkR](http://bit.ly/2xzaTkR)

Lecture 1: Introduction, Overview, Why and How to Use ESS Data




*'Is Belgium made a worse or a better place to live by people moving to the country?'*






ESS round	Score
2002	4.22
2004	4.4
2006	4.63
2008	4.86
2010	4.61
2012	4.7

The European Social Survey (ESS) has been recognized to provide a high number of European countries since 2002. Figures for participation in each ESS round.

4:05 AM - 27 Sep 2017

4 Retweets 

  4 

# Archive & Publish

- Why would you archive and publish your data?

Career benefits

Scientific progress

**Norms**

External drivers

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).*



# Archive & Publish

- Why would you archive and publish your data?

Career benefits

Scientific progress

Norms

**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:



# Archive & Publish



- Selecting data for publication
  - Think about what (part of the) data you want to publish
  - Are you allowed to share your data?
  - Do you have consent for sharing from participants?

You can *archive* data without making it publicly available!



# Archive & Publish

- There are multiple ways to publish data

⊕ Journal supplementary material service

⊕ Institutional data repository

⊕ General purpose repository

⊕ Domain specific data repository

⊕ Trusted domain specific data repository



# Trustworthy Digital Repositories

- Is a **certified** archive
- Serves a **specific community** and offers **support**
- Stores your data **safely** and for the **long term**
- Offers **curation** (i.e. transforms file formats)
- Offers different **options of publishing** data

? You can *archive* data without making it publicly available!



# Trustworthy Digital Repositories



- Helps you make your data FAIR
  - You get a Persistent Identifier for your data
  - Your (meta)data can be found through the archive
  - You can specify license and re-use agreements



**F**indable

**A**ccessible

**I**nteroperable

**R**eusable





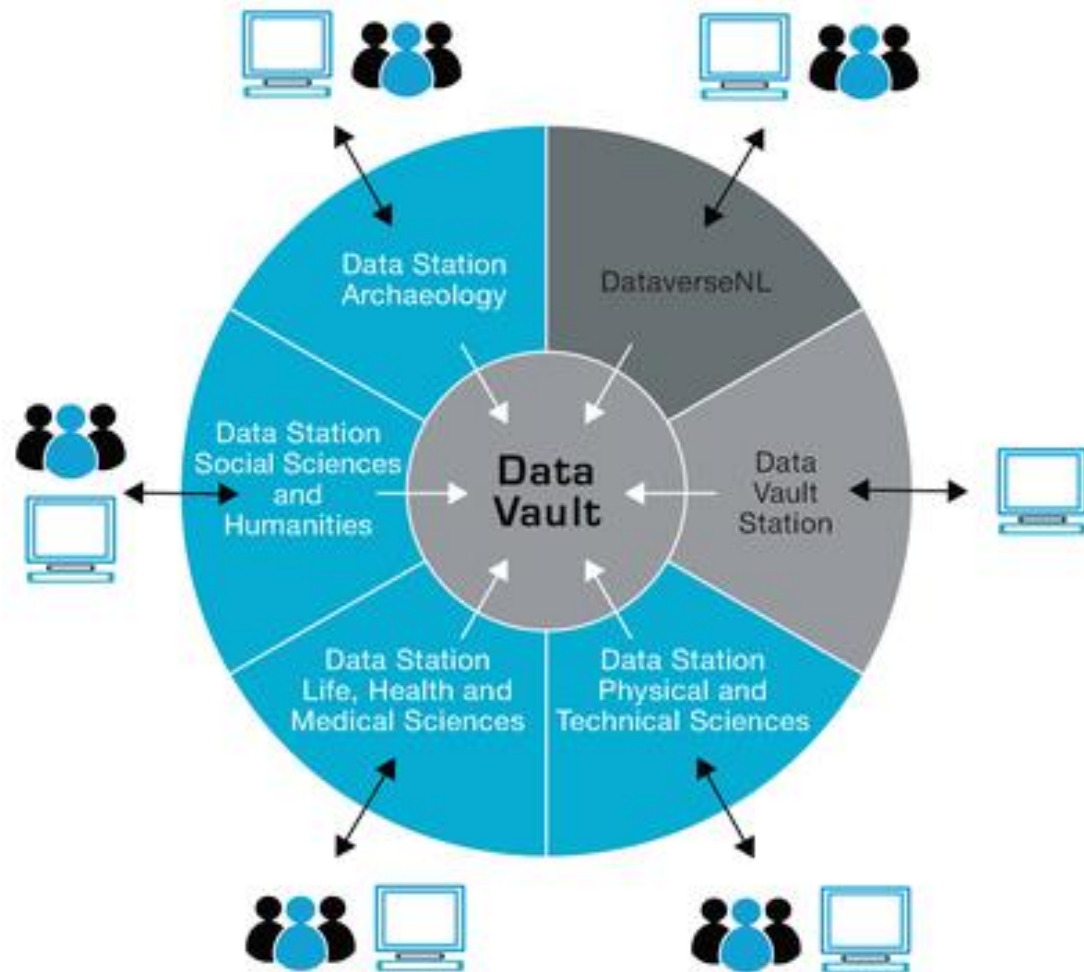
# Finding an archive for your data

- Ask about your institute's policy
- Find the CESSDA archive in your country
- Find other trustworthy digital repository in re3data



# DANS

- Trustworthy Digital Repository
- Service Provider of CESSDA



■ Members  
■ Partners



# Archiving personal data

- Special guidelines for data containing personal information
  - Follow GDPR
  - Gain informed consent for sharing data!
  - Minimize personal information!
  - Anonymize [?] No personal data!
  - Pseudonymize [?] Still personal data!
- It might not be possible to store highly sensitive data



# Archiving personal data

- Processing agreement

## Processing agreement EASY

When you deposit personal data in the meaning of the GDPR, you, or your organisation, are a data controller and DANS is a data processor under the GDPR. You and DANS must conclude a processing agreement. DANS offers you a **standard processing agreement** (.pdf). A processing agreement does not apply if your dataset only contains personal data to account for the dataset, such as the 'Creator' (one of the metadata fields) or citations.



# Archiving personal data

## Reusability and the GDPR: Archiving your data with DANS



Practical information on sharing personal data from research

Emilie Kraaikamp – Advisor for legal affairs

@DANSKNAW

*Want to know more?*

*A short video about the regulations at DANS is available on our [YouTube](#)*





# Access and Licenses



License	Copy & redistribute?	Attribution required?	Commercial use?	Modify & adapt?	Change license?
CC0	Y	N	Y	Y	Y
CC BY	Y	Y	Y	Y	Y
CC BY-SA	Y	Y	Y	Y	Y
CC BY-ND	Y	Y	Y	N	N
CC BY-NC	Y	Y	N	Y	Y
CC BY-NC-SA	Y	Y	N	Y	Y
CC BY-NC-ND	Y	Y	N	N	Y

- Access categories

- Choosing a license

- **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 and Licenses



License	Copy & redistribute?	Attribution required?	Commercial use?	Modify & adapt?	Change license?
CC0	Y	N	Y	Y	Y
CC BY	Y	Y	Y	Y	Y
CC BY-SA	Y	Y	Y	Y	Y
CC BY-ND	Y	Y	Y	N	N
CC BY-NC	Y	Y	N	Y	Y
CC BY-NC-SA	Y	Y	N	Y	Y
CC BY-NC-ND	Y	Y	N	N	Y

- Access categories

- Choosing a license

❓ DMEG provides guidance

- **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 and Licenses



License	Copy & redistribute?	Attribution required?	Commercial use?	Modify & adapt?	Change license?
CC0	Y	N	Y	Y	Y
CC BY	Y	Y	Y	Y	Y
CC BY-SA	Y	Y	Y	Y	Y
CC BY-ND	Y	Y	Y	N	N
CC BY-NC	Y	Y	N	N	Y
CC BY-NC-SA	Y	Y	N	Y	Y
CC BY-NC-ND	Y	Y	N	N	Y

- Access categories
- Choosing a license

- **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.

*At DANS Personal data always needs to be deposited as restricted access.*



# Promoting your data

- Once archived, cite your data and get credits for it!

- ⊕ Choose open access
- ⊕ License your data
- ⊕ Always cite your data
- ⊕ Publish in a data journal
- ⊕ Teach with your dataset
- ⊕ Choose a data repository which promotes your data
- ⊕ Grow your data's impact with altmetrics





# Discover





# Reusing Data



## Benefits of reusing data

### Reuse data and save costs and time



Using existing data is a cost- and time-saving way to carry out your own research. You can use data in the same way as the researchers before you, or you can look for new perspectives and use the data differently.

### Compare results or make replication studies



Adopting previously used elements of research design allows you to compare your results across time and internationally and allows you to make replication studies. Many collaborative projects such as the [International Social Survey Programm](#) (ISSP, n.d.) or the [European Social Survey](#) (ESS, n.d.) make their data publicly available and rely on a culture of data sharing and open access.

### Reuse verified elements of research design



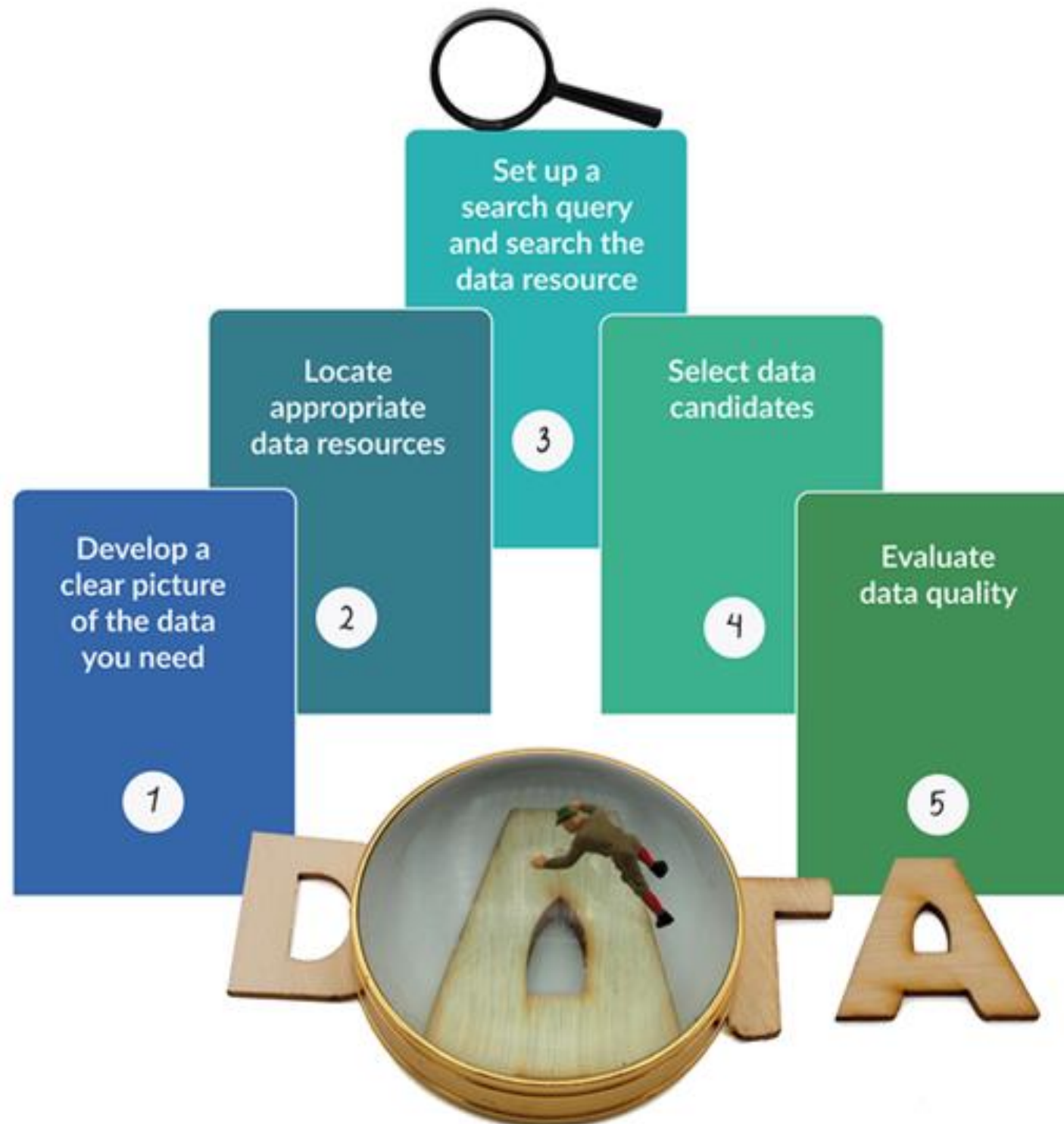
Existing databases and their metadata allow you to check the measurement instruments and other elements of study design that have been tested in prior research. You could use such verified elements of research design in your own data collection.

### Enhance data quality and foster innovation



Discovering existing data helps you to adopt existing research standards, embed your research into a contemporary state of knowledge and make your study more innovative.

# Discovering Data



# Discovering Data

Data Archives as data resources

Important (trusted) domain repositories are:

- ⊕ CESSDA Archives
- ⊕ Out-of-CESSDA European social sciences data archives
- ⊕ Selected non-European data archives
- ⊕ Other important data repositories





# Discovering Data

## Data resources for ageing



Key European data resources for research related to ageing and its effects on individuals and society.

## International comparisons



Interested in research data for international comparison? Have a look at our directory of international surveys.

## Other curated data sources



CESSDA prepares data discovery materials, selections of data resources and organises data discovery events.



+ section on social media data



# ODISSEI Portal

- One place to find social science datasets
  - Metadata from various providers is aggregated
  - Information is enriched using semantic technologies
  - Information is linked in a graph that can be queried





# Access, use and cite data

- To reuse data you need the information a data depositor provided to the archive
  - Access conditions
  - License information
  - Citation information





# Thank you for your attention!

---

*[ricarda.braukmann@dans.knaw.nl](mailto:ricarda.braukmann@dans.knaw.nl)*

This workshop is organized by ODISSEI, CESSDA-ERIC and DANS.

