CESSDA

Archive and Publish your Data

Sonja Bezjak
Social Science Data Archives, University of Ljubljana

Train the Trainers event on RDM
12-13 April 2018, Ljubljana, Slovenia

cessda.eu @CESSDA_Data

Cite as: Bezyak, S. (2020). The CESSDA Data Management Expert Guide. Chapter 6. Archive & Publish [presentation]. Bergen: CESSDA ERIC.

Overview

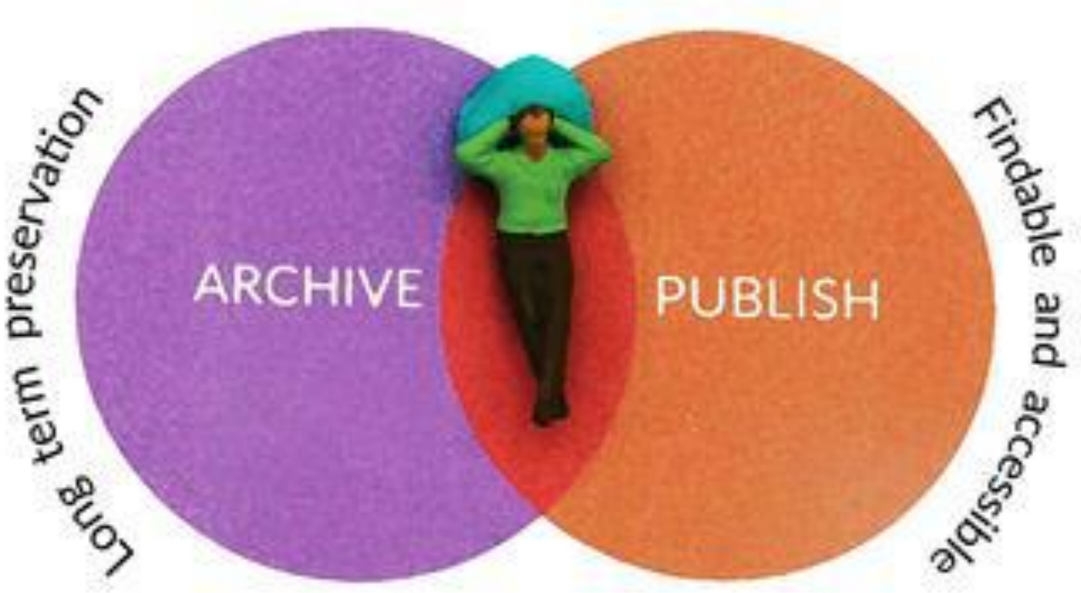
Archiving and publishing your data properly will enable both your future self as well as future others to get the most out of your data.

Areas to be covered:

- Reasons to archive & publish;
- Data publication;
- Different data publication routes;
- Domain specific data repository;
- Publishing with CESSDA archives;
- Promoting data publication.

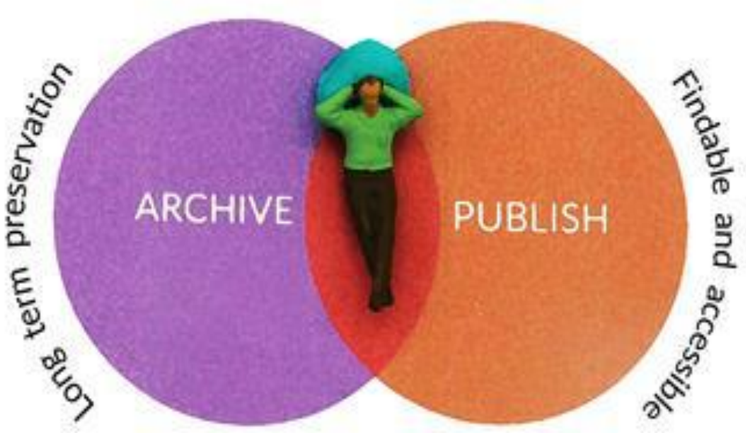
Towards archiving & publishing

ARCHIVING ? PUBLISHING ? SHARING ?



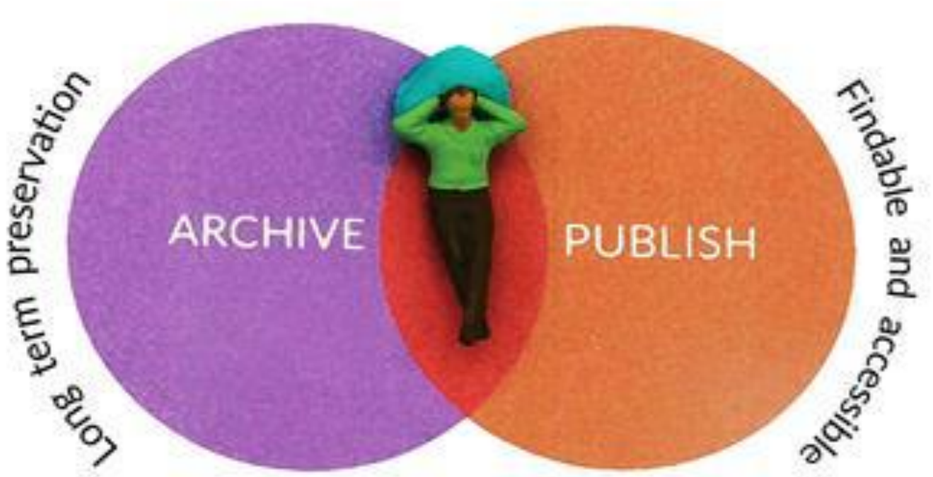
Archiving data

- Store your data in a suitable file format, with adequate documentation and keep your data safe on along term
- Make sure you can read and access the data later on
- Allow access to others for verification purposes



Publishing data

- Is the act of publicly disclosing the research data you've collected,
- Making them findable,
- Accessible and
- Reusable.



Why not excuses…

- We might want to use it in a paper;
- It is not very interesting;
- People may misinterpret the data;
- Poor quality;
- …

"Open Data Excuse" Bingo **#openDataExcuses**

I don't mind, but someone else might	Terrorists will use it	It's not very interesting	It's too big
People may misinterpret the data	Thieves will use it	There's no API	It's too complicated
What if we want to sell it later	We might want to use it in a paper	There's already a project to...	We will get too many enquiries
Data Protection	We'll get spam	Lawyers want a custom License	Poor Quality

For open data teams; print out a copy and put it on your office wall. Cross out each excuse people give you. There are no prizes, but you can tweet "bingo! #openDataExcuses" if you think it might make you feel better*.

* it won't

Generate your own bingo grids at http://data.dev8d.org/devbingo/

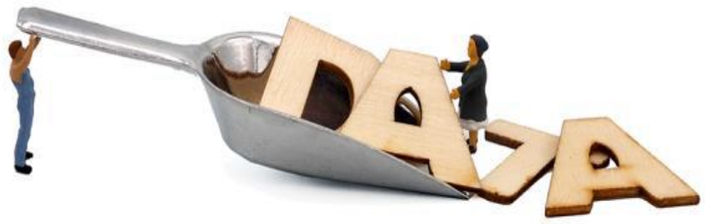
Why researchers should think about A&P?

Why researchers should think about A&P?

- **Career benefits**
  - increased visibility, reuse and citation and therefore recognition of scholarly work
- **Scientific progress**
  - enabling new collaborations, new data uses and links to the next generation of researchers
- **Norms**
  - openness of research data is at the heart of scientific ethics
- **External drivers**
  - Funders and publishers requirements

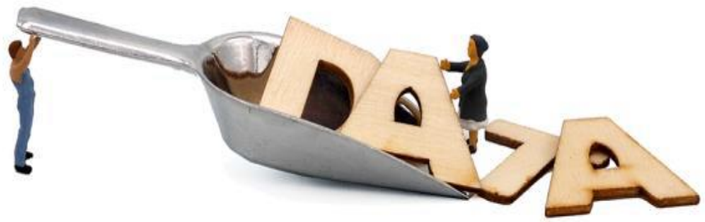
How to select data for A&P?

- **Does your data have potential value in terms**
  - of reuse,
  - national/international standing and quality,
  - importance for history,
  - uniqueness (the data contain non-repeatable observations),
  - originality, size, scale,
  - costs of data production or
  - innovative nature of the research?



How to select data for A&P?

- **Is your data set reusable?**
  - Can the data be read and used?
  - Are metadata available and sufficient to let future users understand your data?
  - Any legal objections which prevent the data from being published?



Expert tips

- **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.

What is a data publication?

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

(Brase et al., 2009)

What is a data publication?

- Publishing with a capital “P“
  - Properly documented with metadata;
  - Reviewed for quality;
  - Searchable and discoverable in catalogues (or databases);
  - Citable in articles.

(Brase et al., 2009)

- Publishing with a small “p“
  - There are no guarantees that the data will be there after some time or that the files will not get corrupted.

Which data repository to choose?

- Recommended by OpenAIRE(2016b):

1) A (trusted) domain repository already established for your research domain.

2) If a domain repository isn't available, use an institutional research data repository.

3) If none of the above is available, use a general purpose repository like Zenodo(n.d.), Figshare(n.d.) or Harvard Dataverse(2017)

4) Find your own at re3data.org: a registry of over 1500 research data repositories.

Expert tips

- Choose between self-archiving and expert help
  - While self-archiving is a quick and easy way to publish data, archiving with the help of an expert will enhance data quality.

Why publish with CESSDA archives?

- **(Trusted) domain specific data repositories**
  - For high-quality data with a potential for reuse, we recommend you to assure long-term access by publishing them with a trusted repository, like many of the CESSDA archives.

- **Advantage of having expert help within reach**
  - help you to increase the comprehensibility, visibility, findability, reusability, longevity and the overall quality of your datasets.

Why publish with CESSDA archives?

- **Accessible and protected when needed**
  - 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.

- **Comprehensibility**
  - CESSDA expert will advise you on what information is needed to understand your data. Ensuring that your metadata is as rich and complete as possible helps in making sure your data meet the F (Findability) and I (interoperability) in FAIR data management.

Why publish with CESSDA archives?

- **Find ability and visibility**
  - When you publish your data at a CESSDA archive your data become more visible through data citation, scientific credits, active promotion.

- **Accessibility and reusability**
  - With a combination of data licensing (see 'Data licenses') and access categories (see 'Access categories') CESSDA data archives can control the exact level of access and permitted reuse.

Why publish with CESSDA archives?

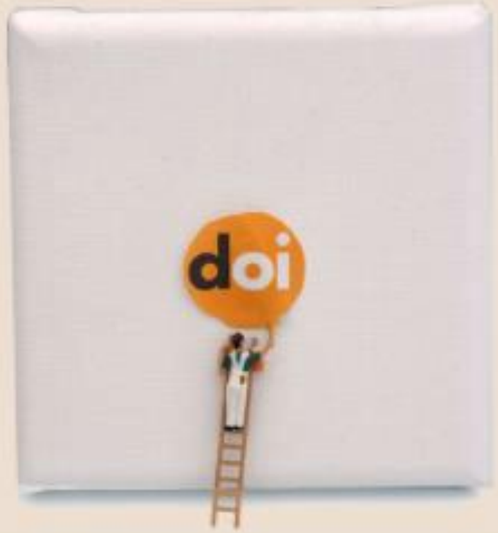
- **Longivity**
  - Experts at CESSDA archives add to the longevity of your datasets in the following ways: advice on the best file formats for long-term preservation; expertise and services to convert data to new formats; add value to the data, for instance by new functionality to query the data.
- **Quality**
  - In several CESSDA archives, an expert will review the quality your data by judging e.g. the content of the study, methodology, relevance, legal consistency and documentation of materials.

For in-depth information check...

Data licensing

Find out how applying a license to your data determines its reusability

Data citation



Find out how persistent identifiers allow your data to be found and cited

Access categories

Find out how which data access category is suitable for your data

How to make data more „visible“ …

- Choose open access
- Licence your data
- Always cite your data
- Publish in a data journal
- Teach with your data set
- Choose a data repository which promotes your data
- Grow your data‘s impact with altmetrics

Hands-on

- 1) Check if and how data is cited in the article
- 2) Prepare a Study description (ADP)

Cite the data - ADP

The citation is composed of:

- [Principal Investigators]. ([Version Year]). [Title]. Slovenia, Ljubljana: Univerza v Ljubljani = University of Ljubljana, Arhiv družboslovnih podatkov = Social Science Data Archives. [Study Number] Accessible at: [link]

Exam le: *Hafner - Fink, M. in Malešič, M. (2016). Slovensko javno mnenje 2015: Mednarodna raziskava Stališča o delu (ISSP 2015), Mednarodna raziskava Stališča o vlogi države (ISSP 2016), Ogledalo javnega mnenja in raziskava Stališča o varnosti [Data file]. Ljubljana: Univerza v Ljubljani, Arhiv družboslovnih podatkov. ADP - IDNo: SJM15. Accessible at:*

http://www.adp.fdv.uni-lj.si/opisi/sjm15/

Cite the data - GESIS

The citation is composed of:

- [Principal Investigators] ([Version Year]): [Title]. [Data Collector]. GESIS Data Archive, Cologne. [Study Number] Data file Version [Version Number], [DOI]

Example: *Schmitt, Hermann; Popa, Sebastian Adrian; Devinger, Felix (2015): European Parliament Election Study 2014, Voter Study, Supplementary Study. GESIS Data Archive, Cologne. ZA5161 Data file Version 1.0.0, doi:10.4232/1.5161*

Citing new data types

Citing rapidly changing data is challenging.

The Data Cite organization suggests 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.
- 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.

Concluding remarks

Archiving and publishing your data properly will enable both your future self as well as future others to get the most out of your data.

Questions

Sonja Bezjak
Social Science Data Archives
University of Ljubljana
Train the Trainers event on RDM
12-13 April 2018, Ljubljana, Slovenia

cessda.eu

@CESSDA_Data