Different aspects of valuating research data

Sonja Bezjak, Janez Štebe, ADP

Focus on Open Science, Citizen Science, Research Data, Access to Knowledge, and Beyond, 24 November 2017, Ljubljana

Content

1) Who we are and what we do
2) Open data
3) Data Publications
4) Criteria for Data Evaluation in SSDA
5) Open questions and dilemmas

Social Science Data Archives, University of Ljubljana

http://www.fdv.uni-lj.si/

- 1997
- National data repository for social sciences
- Depositors from all 4 (3 public) universities, private research centres, Statistical Office of Slovenia (8-10 research centres per year)
- 600 social science surveys
- cca. 700 users yearly (90 % education, 10 % scientific/research purpose)
- Member of CESSDA
- International projects (DwB, Foster, SEEDS, SaW, SERISS …)

ADP IS CELEBRATING 20 YEARS, 8TH OF DECEMBER 2017

Date: Dec. 8, 2017
Place: Ljubljana, Slovenia
Organizer: ADP - Arhiv družboslovnih podatkov

The ADP is celebrating its 20 years of operation. We will celebrate the anniversary on **Friday, 8 December 2017, from 9 am to 1 pm with two main events:**:

- 09:00 – 10:30 **Roundtable "Challenges of qualitative research in social sciences"**
- 10:45 – 12:00 **Official ceremony "20 years of ADP"**

Together with invited guests, we will **present the traveled path and indicate the future direction of development of ADP as an advanced national repository for the social sciences**, embedded in the European data infrastructure CESSDA ERIC.

We will round off the ceremony with **a party and a relaxed conversation with researchers, partners, and colleagues** . Book your date!

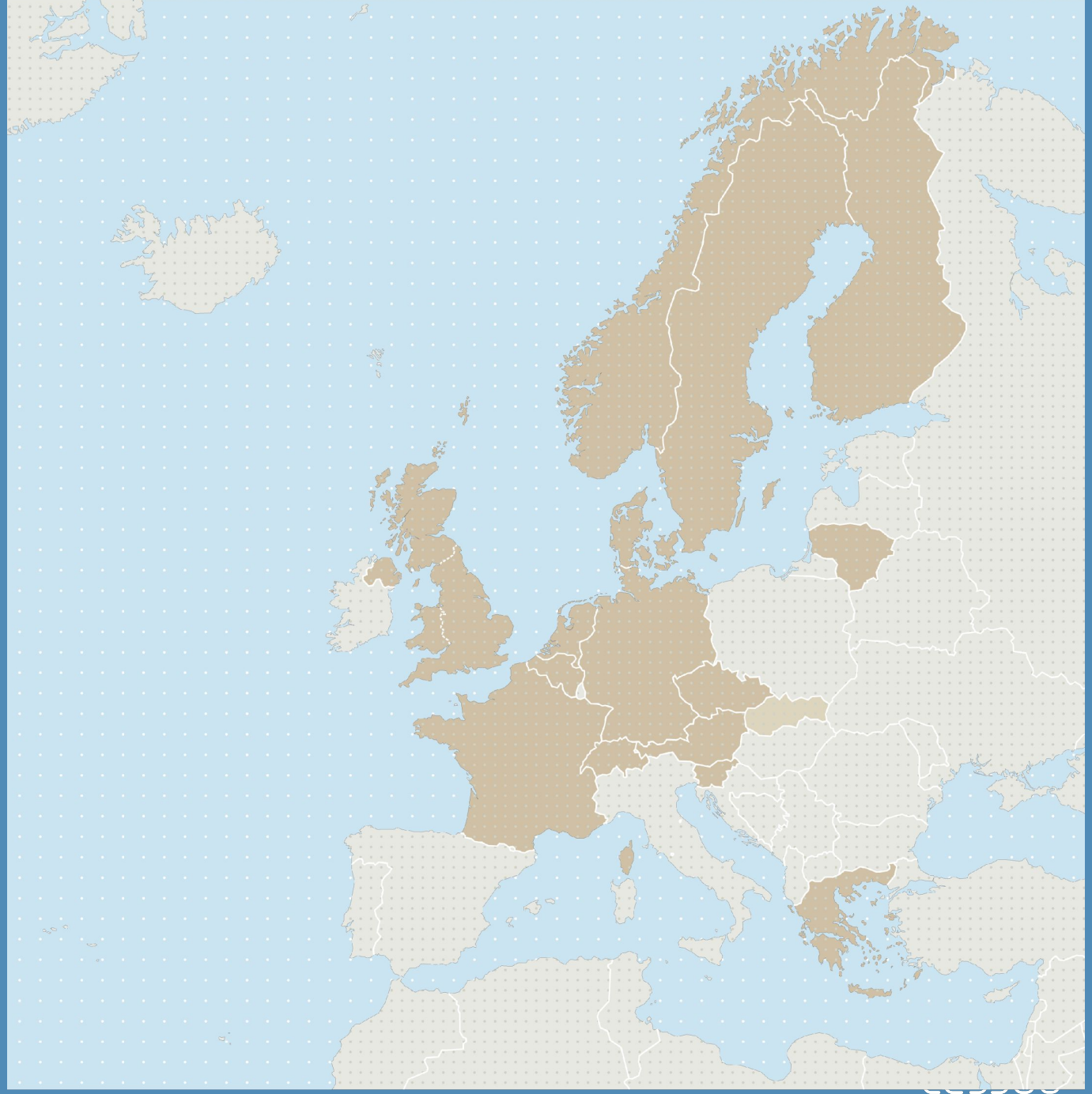
More detailed information about the event and the registration forumlar is available HERE.

For more information visit: https://www.adp.fdv.uni-lj.si/dogodki/adp_praznovanje_20let/

Consortium of European Social Science Data Archives

- 15 CESSDA members
- Austria - AuSSDA
- Belgium - SOHDA
- Czech Republic - CSDA
- Denmark - DDA
- Finland - FSD
- France - PROGEDO/Réseau Quetelet
- Germany - GESIS
- Greece - So.Da.Net
- Lithuania - LiDA
- Netherlands - DANS
- Norway - NSD
- Slovenia - ADP
- Sweden - SND
- Switzerland - FORS
- UK - UKDS

- 1 observer
- Slovakia - SASD (observer)



CESSDA training group

- New on-line **Research Data Management module** for researchers will be launched on December 14 2017 in Ljubljana
- Prepared by CESSDA SP
- ADP leading a chapter „Archive & Publish“

cessda

Open data: idea → reality

Open data is data that is free to access, reuse, repurpose, and redistribute.

https://www.openaire.eu/opendatapilot

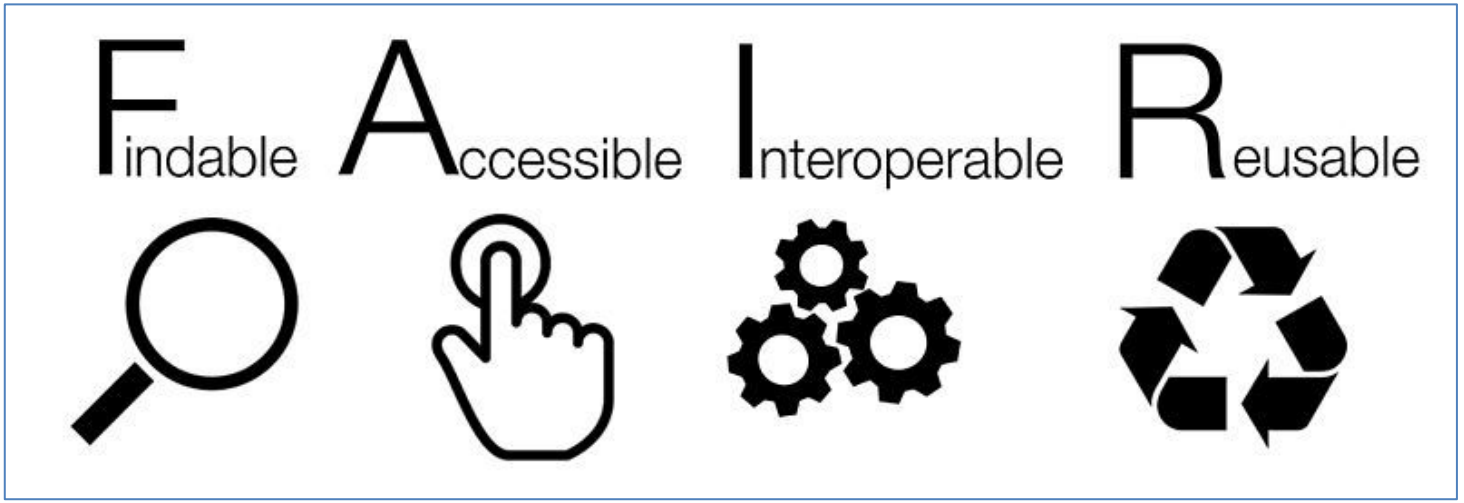
H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

EUROPEAN COMMISSION
Directorate-General for Research & Innovation

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf



SangyaPundir. (2016). English: FAIR guiding principles for data resources. Retrieved from https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

Open data →

Benefits for research community:

- New research findings;
- Further analysis based on data related to publication;
- Removal of duplication of effort;
- Using research data for teaching purposes;
- and many more…

Štebe, Janez, Bezjak, Sonja, Vipavc Brvar, Irena (2015). Preparing research data for open access : guide for data producers. URN:NBN:SI:DOC-G0DPXMZ1 from https://www.dlib.si

2 important questions

1) What is a data publication?

2) What is archiving data?

„ data Publication" vs. „ data publication"

Publishing with a small „p" means:

- researchers publish their data files on a website somewhere,
- there are no guarantees that the data will be there after some time or that the files will not get corrupted.

Publishing with a capital „**P**" means:

- for a dataset to "count" as a publication, it should follow a similar publication process as an article

Costas, R., Meijer, I., Zahedi, Z. and Wouters, P. (2013). The Value of Research Data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report, available from www.knowledge-exchange.info/datametrics

Data *P*-ublication

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 should be:

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

Costas, R., Meijer, I., Zahedi, Z. and Wouters, P. (2013). The Value of Research Data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report, available from www.knowledge-exchange.info/datametrics

Where to publish research data?

- Journal supplementary material service
- Institutional data repository
- General purpose repository
- Domain specific data repository
- Trusted domain specific data repository

Recommendation:

- Use an external **data archive** or **repository already established for your research domain** to preserve the data according to recognised standards in your discipline.

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

How to publish research data?

Self-archiving

- without any help
- a quick and easy way to publish data
- check: Dryad, Figshare, Zenodo…

Archiving with trusted domain repository

- 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: CESSDA ERIC SP

For more information see soon: CESSDA Expert tour guide on Data Management (on-line module) to be launched in December 2017 on CESSDA ERIC web-site!

Quality criteria in CESSDA archives

„In several CESSDA archives, an expert will review the quality of your data by judging e.g.

- the content of the study,
- methodology,
- relevance,
- legal consistency and
- documentation of materials.

You can have a look at the European diversity in such quality checks by CESSDA archives if you open the accordion.“ (*CESSDA Expert tour guide on Data Management)*

For more information see soon: CESSDA Expert tour guide on Data Management (on-line module) to be launched in December 2017 on CESSDA ERIC web-site!

Evaluation process in SSDA

- **A lot of effort in the evaluation to be able:**

→ to fulfill the mission

→ to be sustainable and cost effective

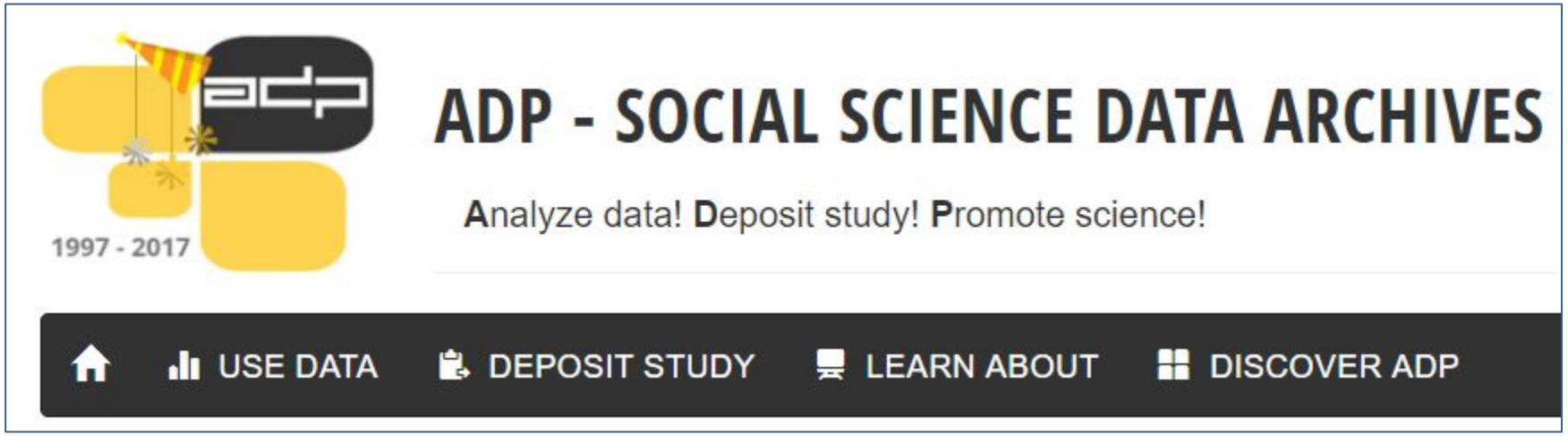
- **What we evaluate:**

⟶ researcher‘s work

Let's dive into…

Selection and appraisal criteria and workflow in:

https://www.ukdataservice.ac.uk/



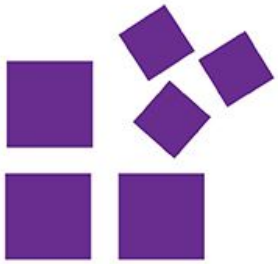
https://www.adp.fdv.uni-lj.si/

Selection and evaluation in the UK DS

- Collections Development Policy
- Collections Development Selection and Appraisal Criteria
- **UK Data Service Collection Development Appraisal Grid**

Special thanks to Dr Scott Summers, UK DS, who helped me to determine relevant criteria in UK DS

"Valuable resources that can be used and reused"



UK DS: Strategic Value

Scope of collections

The UK Data Service's Collection Development Policy (CDP) states that it acquires data to **meet three central purposes:**

- *Potential secondary use and analysis for research*:
- *Teaching and learning use*
- *Replication and validation of research*

UK DS Collections Development Policy, 19 January 2016, Version: 05.00, available at: https://www.ukdataservice.ac.uk/media/398725/cd227-collectionsdevelopmentpolicy.pdf

UK DS: Curation Categories for all data collections

- *CURCAT1: Data collections selected for l**ong-term** curation.*
- *CURCAT2: Data collections selected for **"short-term"** management.*
- *CURCAT3: Data collections selected for **'delivery' only,***
- *CURCAT4: Data collections selected for **"discovery"** only*.
- CURCAT5 relates to preservation-only which falls outside the scope of the UK Data Service Appraisal, and which is handled by the UK Data Archive. Data collections may be moved into higher or lower categories over time if the need arises.

UK DS Collections Development Policy, 19 January 2016, Version: 05.00, available at: https://www.ukdataservice.ac.uk/media/398725/cd227-collectionsdevelopmentpolicy.pdf

UK DS: Reference to the UK Data forum strategy

Selection and Appraisal Criteria

Criteria used to review materials / relating to **the value of the data**:

- *Relevance to the remit:* The content of the materials meet the ESRC's current strategies and are within the '**Scope of Collections**'. This includes being **high-quality, authoritative, reliable and up-todate** with '**good temporal and spatial coverage**', with **long and consistent time series** (as appropriate), and **strong opportunities for comparable research**
- Scientific or historical value
- New sources or types of data
- International value
- Uniqueness/risk of loss
- Usability/redistribution/operational benefits
- Replication data and resources

UK DS Collections Development Policy, 19 January 2016, Version: 05.00, available at: https://www.ukdataservice.ac.uk/media/398725/cd227-collectionsdevelopmentpolicy.pdf

UK DS: User need and analytical value

Data useful for supporting decision-making and policy formation
Data useful for international, longitudinal or comparative research (e.g. time series or harmonised)
Authoritative source of high quality data and / or widely cited
Microdata otherwise not available to the research community (e.g. disclosive)
Data directly requested by a critical number of users
Unique, unrepeatable data or data that are costly to reproduce
Are there related data already in the Collection?

UK Data Service Collection Development Appraisal Grid

UK DS: Usability and accessibility

Data have definable conditions within the access spectrum
Legal and ethical agreements do not prohibit re-use, e.g. consent to share/reuse data, IPR, copyright, DPA
Data has adequate supporting documentation/contextual information
Data are produced in, or can be converted at minimal or no cost to, a suitable format for reuse (without need for specialised software)
Do data exceed 5GB in size?
Third party host (e.g. API): is reliable and reputable
Third party host: is the data likely to change/be updated over time?
Third party host: suitable preservation processes are used to manage and curate data
Third party host: a persistent identifier can be made available (robust access)

UK Data Service Collection Development Appraisal Grid

UK DS: Criteria for not accepting data…

… even though the value is deemed to be high:

- Legal and ethical issues: there are insurmountable rights management issues e.g. consent, IPR, copyright and Data Protection issues which are unable to be satisfactorily resolved and where full use of the data would not be possible without infringing legislation;
- Lack of sufficient contextual materials to enable re-use;
- Old/unreadable formats that are difficult to convert or make usable, or impossible to recover.

UK DS Collections Development Policy, 19 January 2016, Version: 05.00, available at: https://www.ukdataservice.ac.uk/media/398725/cd227-collectionsdevelopmentpolicy.pdf

Selection and evaluation in the ADP

HOW TO DEPOSIT DATA?

RECORD ➜

PREPARE

➜ DEPOSIT

ADP: Collection development policy

The ADP is aiming for research **data:**

- that are of interest for social sciences' research;
- covering problems, connected with the Slovenian society;
- studies with theoretical, conceptual and methodological excellence;
- especially longitudinal and internationally comparable data that include data from Slovenia;
- exceptionally, studies outside social sciences', if they either structurally resemble research data from social sciences or lack other possibilities of preservation.

For ADP's Digital preservation policy see:
https://www.adp.fdv.uni-lj.si/eng/spoznaj/politika/

ADP: Criteria for evaluation

... which are followed when evaluating a study for deposit:

- **the richness of content** in terms of adequacy of conceptualisation, suitability, and the potential for use, and thematic broadening of the ADP catalogue,
- **the transparency and quality of the methodologies and completeness and validation of data and documentation** to enable further analyses,
- **the willingness** of the depositor, as **a copyright owner to deposit data** with ADP for dissemination.

For ADP's Digital preservation policy see: https://www.adp.fdv.uni-lj.si/eng/spoznaj/politika/

ADP: detailed questions

… to be answered by ADP staff before taking a decision on aquisition:

For ADP‘s Digital preservation policy see:
https://www.adp.fdv.uni-lj.si/eng/spoznaj/politika/

ADP: Workflow in a Pre-Aquisition phase

1. Head of the aqusition receives information about the new research
2. Digital archivist (DA) gets in touch with authors and gathers needed information on the research
3. Monthly a meeting of *Commission for aquisition* is held
4. DA reports to the *Commission for aquisition* about:

- Continuity of research
- Comparability of research
- Quality sampling and appropriate population
- Database size
- Geographic coverage
- Content of the study
- Other methodological criteria
- Relevance and quality of study materials

ADP: Workflow in a Pre-Aquisition phase

5. If yes -->

→ DA proposes whether to proceed or not into the aquisition phase

6. *Commission for aquisition* takes a decision on accepting or declining the research

Aquisition phase:

7. DA reports to the author about the decision and asks for further assistance if needed

8. If needed another meeting is planned with the author to resolve open questions

ADP: Quality check

- Detailed quality **checks** to ensure the **completeness and the understanding** of any deposited data and documentation;
- **Dataset dimension checks**: the number of cases and variables are checked against the documentation;
- **Metadata checks**: all variables should have variable labels and all categorical variables should have value labels / the dataset must be comprehensible in association with the documentation given to users;
- **Data validity checks**:

✓ All categorical variables must be checked for out-of-range values/wild codes;

✓ Possible interval variables must be checked for improbable or impossible value - dataset being checked for consistency with/against published results (e.g. report, journal article…);

✓ Research material being checked for legal consistency (e.g. GDPR, IPR).

Before the data is being published, author gets final evaluation

ADP's Study Clasification

1	»Study outside of the geographic or thematic framework of the ADP«	Includes studies that do not fall under the broader thematic framework of social sciences, namely sociology, law, educational science, economics and humanities. Also, this group includes studies that geographically cover smaller or remote areas that are analytically not so interesting to our users.
2	»Ad hoc study of limited quality«	Includes studies that do not meet the minimum methodological criteria. These are, for example, seminar work studies made by undergraduate students that cover a very narrow field of research and collect data on small occasional samples.
3	»Study with a limited problematic or theoretical framework thematically covered elsewhere«	Includes studies that do not have a well-developed research problem and/or theoretical framework. These are smaller studies, whose content is already covered by other, more extensive and relevant studies.
4	»Study with a limited conceptual framework, less methodologically developed«	Includes studies with a limited conceptual framework. These are studies that cover smaller research areas, having typical methodological deficiencies.
5	»Study that is interesting as an attempt to conceptualize a new research area, using ad hoc samples«	Includes studies that represent a novelty in a particular research field and could later on fill in a research gap. However, at the same time they have certain methodological deficiencies, such as the use of casual samples instead of random sampling.
6	»Study with a limited problematic or theoretical framework and a narrower applicability to a number of practical problems; methodologically and contextually improved«	Includes studies that do not have a well-developed research problem and theoretical framework and cover a smaller research field. Such studies are less useful for advanced research in a chosen field. On the other hand, they meet the requirements of methodological relevance and contentment.
7	»Theoretically or practically relevant studies limited to a less general population; less influential«	Includes studies that meet the majority of the criteria of scientific excellence. On the other hand, they may be limited to a less general population or are not methodologically perfect or cover a smaller research area.
8	»Theoretically or practically important studies which fill the research gap or have wide applicability for many practical problems and have a long-term value to science«	Includes studies that meet virtually all the criteria of scientific excellence. As a rule, this group includes methodologically relevant studies that have useful value for science, however they are not part of series or cover a slightly smaller research field.
9	»The highest rank, comparative or continual studies, a relevant population, methodologically appropriate«	Includes only the best studies that meet all of the stated criteria of scientific excellence. These are studies, which are usually part of series, they are conducted on the highest methodological level and are of great importance for science.

https://www.adp.fdv.uni-lj.si/eng/deli/merila/klasifikacija/

ADP: Scientific data publication

Studies that are given grades 7, 8 or 9 are given the status of a scientific publication, according to the criteria of the Slovenian Research Agency that is as type H. Final scientific data collection or corpus (2.20) from the list of the agency (BIBLIO-D), 30 points.

ADP's Study Clasification:

7	»Theoretically or practically relevant studies limited to a less general population; less influential«	Includes studies that meet the majority of the criteria of scientific excellence. On the other hand, they may be limited to a less general population or are not methodologically perfect or cover a smaller research area.
8	»Theoretically or practically important studies which fill the research gap or have wide applicability for many practical problems and have a long-term value to science«	Includes studies that meet virtually all the criteria of scientific excellence. As a rule, this group includes methodologically relevant studies that have useful value for science, however they are not part of series or cover a slightly smaller research field.
9	»The highest rank, comparative or continual studies, a relevant population, methodologically appropriate«	Includes only the best studies that meet all of the stated criteria of scientific excellence. These are studies, which are usually part of series, they are conducted on the highest methodological level and are of great importance for science.

https://www.adp.fdv.uni-lj.si/eng/deli/merila/klasifikacija/

ADP: after using research data from ADP catalogue

Cite the data!

You are obliged to respect general provisions and conditions of use that you agreed to at registration even after using the research data from the Catalogue of the ADP.

- **Citation:** you are **obliged to reference** the data, their original author and the ADP in all subsequent publications and presentations, based on data (for example in graphs, tables, referencing data within text etc.). The used data files need to be listed also in the **list of used resources at the end** of your publication (Article 5).

https://www.adp.fdv.uni-lj.si/eng/uporabi/po_uporabi/

Tracking data publications / Services being developed

Citation-based metrics

- DataSearch (Elsevier)
  Searches data repositories, including figures/tables and has as preview option so you can judge whether the data are useful.
- DataCite
  Searches datasets that have been given a DOI.
- Data Citation Index
  (Web of Science - licensed database with paid access only)
  Searches the metadata of the datasets for connected data repositories.

Altmetrics-based metrics

- ImpactStory
- PlumX
- Altmetric
- Data repositories (Downloads and views counts)

Ball, A. & Duke, M., 2015. How to Track the Impact of Research Data with Metrics. DCC How-to Guides. Edinburgh: Digital Curation Centre. Retrieved from: http://www.dcc.ac.uk/resources/how-guides

Perverse Incentives

TABLE 1. GROWING PERVERSE INCENTIVES IN ACADEMIA

Incentive	*Intended effect*	*Actual effect*
"Researchers rewarded for increased number of publications."	"Improve research productivity," provide a means of evaluating performance.	"Avalanche of" substandard, "incremental papers"; poor methods and increase in false discovery rates leading to a "natural selection of bad science" (Smaldino and Mcelreath, 2016); reduced quality of peer review
"Researchers rewarded for increased number of citations."	Reward quality work that influences others.	Extended reference lists to inflate citations; reviewers request citation of their work through peer review
"Researchers rewarded for increased grant funding."	"Ensure that research programs are funded, promote growth, generate overhead."	Increased time writing proposals and less time gathering and thinking about data. Overselling positive results and downplay of negative results.
Increase PhD student productivity	Higher school ranking and more prestige of program.	Lower standards and create oversupply of PhDs. Postdocs often required for entry-level academic positions, and PhDs hired for work MS students used to do.
Reduced teaching load for research-active faculty	Necessary to pursue additional competitive grants.	Increased demand for untenured, adjunct faculty to teach classes.
"Teachers rewarded for increased student evaluation scores."	"Improved accountability; ensure customer satisfaction."	Reduced course work, grade inflation.
"Teachers rewarded for increased student test scores."	"Improve teacher effectiveness."	"Teaching to the tests; emphasis on short-term learning."
"Departments rewarded for increasing U.S. News ranking."	"Stronger departments."	Extensive efforts to reverse engineer, game, and cheat rankings.
"Departments rewarded for increasing numbers of BS, MS, and PhD degrees granted."	"Promote efficiency; stop students from being trapped in degree programs; impress the state legislature."	"Class sizes increase; entrance requirements" decrease; reduce graduation requirements.
"Departments rewarded for increasing student credit/contact hours (SCH)."	"The university's teaching mission is fulfilled."	"SCH-maximization games are played": duplication of classes, competition for service courses.

Modified from Regehr (pers. comm., 2015) with permission.

„Fair" evaluation in the research community

European Commission (2017):

„In developing a system to evaluate and recognise engagement in Open Science, the full spectrum of OS activities must be taken into account. These include open access to publications, **open data**, open peer review, research integrity, citizen science and stakeholder engagement.

In general, evaluating a researcher cannot be reduced to a number because their merits and achievements are a complex set of different variables, difficult to be summarised by a single figure. A better approach is through multi-dimensional criteria evaluation, taking into consideration what is expected from a researcher and what is relevant for his/her career/recruitment."

Evaluation of Research Careers fully acknowledging Open Science Practices; Rewards, incentives and/or recognition for researchers practicing Open Science, European Commission (2017):
https://cdn1.euraxess.org/sites/default/files/policy_library/os-rewards-wgreport-final_integrated_0.pdf

Evaluating research data →

Looking for the right, fair way towards high quality research…

Contact

Thank you!!!

University of Ljubljana

Faculty of Social Sciences

Social Science Data Archives

Kardeljeva ploščad 5

1000 Ljubljana

SI - Slovenia

www.adp.fdv.uni-lj.si

arhiv.podatkov@fdv.uni-lj.si

Arhiv.Druzboslovnih.Podatkov

@ArhivPodatkov

Univerza v Ljubljani

seriss SYNERGIES FOR EUROPE'S RESEARCH INFRASTRUCTURES IN THE SOCIAL SCIENCES

cessda saw

seeds

cessda