

SNSF Open Research Data Policy and Data Management Plan (DMP)

Lionel Perini

Dealing with Open Data in Language Sciences

December 10, 2018

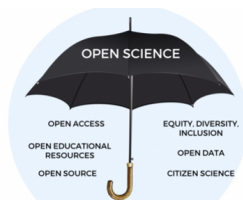


Outline

- 1 Context
- 2 Research Data Management and Open Data
- 3 SNSF Open Research Data Policy
- 4 FAIR Data Principles
- 5 Data Repositories
- 6 Next Steps

Open Science

“**Open Science** is about the way researchers **work, collaborate, interact, share resources and disseminate results**. A systemic change towards open science is driven by new technologies and data, the increasing demand in society to address the societal challenges of our times and the readiness of citizens to participate in research.” *Source: Amsterdam Call for Action*



What do we mean when we talk about Open Science?

Context

Research Data Management and Open Data
SNSF Open Research Data Policy
FAIR Data Principles
Data Repositories
Next Steps

Data Deluge



Reproducibility

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

INTRODUCTION: Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. Scientific claims should not gain credence because of the status or authority of their originator but by the replicability of their supporting evidence. Even research of exemplary quality may have irreproducible empirical findings because of random or systematic error.

RATIONALE: There is concern about the rate and predictors of reproducibility, but limited evidence. Potentially problematic practices include selective reporting, selective analysis, and insufficient specification of the conditions necessary or sufficient to obtain the results. Direct replication is the attempt to recreate the conditions believed sufficient for obtaining a pre-

viously observed finding and is the means of establishing reproducibility of a finding with new data. We conducted a large-scale, collaborative effort to obtain an initial estimate of the reproducibility of psychological science.

RESULTS: We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. There is no single standard for evaluating replication success. Here, we evaluated reproducibility using significance and *P* values, effect sizes, subjective assessments of replication teams, and meta-analysis of effect sizes. The mean effect size (r) of the replication effects ($M_r = 0.197$, $SD = 0.257$) was half the magnitude of the mean effect size of the original effects ($M_r = 0.403$, $SD = 0.188$), representing a

substantial decline. Ninety-seven percent of original studies had significant results ($P < .05$). Thirty-six percent of replications had significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

CONCLUSION: No single indicator sufficiently describes replication success, and the five indicators examined here are not the only ways to evaluate reproducibility. Nonetheless, collectively these results offer a clear conclusion: A large portion of replications produced weaker evidence for the original findings despite using materials provided by the original authors, review in advance for methodological fidelity, and high statistical power to detect the original effect sizes. Moreover, correlational evidence is consistent with the conclusion that variation in the strength of initial evidence (such as original *P* value) was more predictive

ON OUR WEB SITE

Read the full article at <http://dx.doi.org/10.1126/science.aac4716>

Good Scientific Practices

Sign In | Link your subscription | TNY Store

THE NEW YORKER

NEWS CULTURE BOOKS & FICTION SCIENCE & TECH BUSINESS HUMOR CARTOONS MAGAZINE VIDEO ARCHIVE **SUBSCRIBE** 🔍

APRIL 30, 2013

THE REINHART AND ROGOFF CONTROVERSY: A SUMMING UP

BY JOHN CASSIDY

SHARE TWITTER ST+ MAIL

In one of life's little ironies, last Friday's disappointing G.D.P. figures, which reflected a sharp fall in government spending, appeared on the same day that the economists Carmen Reinhart and Kenneth Rogoff published an [Op-Ed](#) in the *Times* defending their famous (now infamous) research that conservative politicians around the world had seized upon to justify penny-pinching policies. Addressing a new paper by three lesser lights of their profession from the University of Massachusetts, Amherst, which uncovered data omissions, questionable methods of weighting, and elementary coding errors in Reinhart and Rogoff's original work, and which went around the world like a viral video, the Harvard duo dismissed the entire brouhaha as "academic kerfuffle" that hadn't vitiated their main points.



MOST POPULAR

1. Why You Hate Google's New Logo
BY SARAH LARSON
2. Six Weeks' Paid Leave Opposed By People With Thirty-Three Weeks' Paid Leave
BY ANDY BOROOWITZ

Research Data Management and Open Data



Research Data Management (RDM)

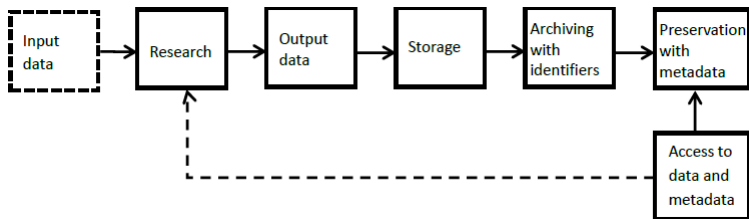
Definition (Digital Curation Centre)

RDM concerns the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results

Prerequisites

- Elaboration of institutional policy guidelines
- Technical solutions have to be set up
- Definition of responsibilities between the different stakeholders

Data Lifecycle



Source: Research Council Policy on Open Access to Research Data, RCN

Data Lifecycle

- *Research data* can be separated into two main categories, input and output data
- *Data storage* - any data stored on disk is backed up to secondary or alternative storage (institutional IT services for back-up services)
- *Data archiving* - institutions are most of the time not able to afford long-term maintenance of data \Rightarrow data repository/infrastructure
- The *data preservation* phase consists in enhancing data quality for a long-term quality and enriching them with metadata
- The objective of *data curation* is to connect the first use of data to the second use

Data Management Plan (DMP)

Definition

A research data management plan describes the organization, treatment and dissemination of data during the whole data lifecycle

Objectives and uses

- Long-term perspective concerning the use of data
- Clarification of the process - who has access to the data, where will data be shared and stored, data ownership issues
- Often required by funding agencies together with the implementation of a data policy

Summary

- A research data management plan is an excellent way to implement best practices in terms of data management in the scientific community - first reflection about how organizing the data
- Research Data Management: time-consuming exercise, interaction of different stakeholders, administrative burden, requires infrastructures and skills
- Data sharing/opening data is only a part of the research data management process

Open Research Data

- Publicly funded research data can be considered as public goods and therefore accessible for the scientific community but also for the society
- Poor data sharing culture in some research communities - Open data still in its infancy
- Most of researchers agree with the statement that they would use other researchers' datasets but very few make their data openly available
- Data sharing in an intelligent and usable manner

Definition

Definition (Open Knowledge Foundation)

Open data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose

Situation

- Strong differences between research communities in terms of data sharing practices
- No established systems (\neq Open Access to publications)
- “As open as possible, as closed as necessary”

Ethics and Consent

When research involves obtaining data from people, the following points should be accounted for:

- Personal data (data which relate to a living individual who can be identified from those data)
- Confidential data (data agreed to be kept confidential, e.g. health or income data)
- Sensitive personal data (data on a person's race, ethnic origin or political opinion)
- Consent forms should account for data reuse

Open Government Data

- Public sector information made available to the public as open data is termed “Open Government Data”
- “Public bodies produce and commission huge quantities of data and information. By making their datasets available, public institutions become more transparent and accountable to citizens.” (OECD)
- Strengthens transparency, efficiency and innovation
- opendata.swiss

Pros

- Increasing the transparency and quality of research
- Improving reproducibility of results and reducing scientific misconduct
- Reuse of data
- Leading to new collaborations between data users and data creators
- Rational use of resources available
- Accountability for providers of public services
- Availability of important resources for education and training

Benefits of Open Science



CC-BY Danny Kingsley & Sarah Brown

Impediments

- Diffusing your work is very attractive but sharing your data may not be the case \Rightarrow Lack of incentives (\Rightarrow DOI for data citation)
- Lost of competitive advantage if data should be opened
- Time-consuming exercise
- Lack of knowledge on how and where to share/archive data
- Lack of standards and vocabularies, which reduce data interoperability
- Infrastructures and financial issues

SNSF Open Research Data Policy



Background and Aims

- September 2015 - The SNSF discussed the foundations of Open Research Data strategies during an international workshop
- 2016 - Discussions at the Presiding Board of the Research Council and the Administrative Offices
- October 2017 - Implementation of the DMPs in the project funding scheme
- DMPs are now required in most of SNSF funding instruments
- New data policy but data sharing requirements are not new

Background and Aims

Policy Statement

The SNSF values research data sharing as a fundamental contribution to the impact, transparency and reproducibility of scientific research. In addition to being carefully curated and stored, the SNSF believes research data should be shared as openly as possible.

Funding Regulations Article 47b

Data collected with the aid of an SNSF grant must be made available to other researchers and integrated into recognized scientific data pools.

Background and Aims

Data Definition

Research data is collected, observed or generated factual material that is commonly accepted in the scientific community as necessary to document and validate research findings.

Approach

The SNSF favours a bottom-up approach. It provides best practice guidelines and gives each scientific community sufficient flexibility in defining and applying its own standards. In particular, the best way of managing and sharing data depends on the research field.

Implementation

The SNSF expects all its funded researchers

- to store the research data they have worked on and produced during the course of their research work,
- to share these data with other researchers, unless they are bound by legal, ethical, copyright, confidentiality or other clauses, and
- to deposit their data and metadata onto existing public repositories in formats that anyone can find, access and reuse without restriction.

Data Management Plan

- Researchers have to include a data management plan (DMP) in their funding application for most of the funding schemes: DMP is a formal requirement
- DMPs are not part of the review process (no access for external reviewers).
- At project submission, DMPs are considered as drafts.
- Applicants can explain in the DMP if there are any issues linked to data sharing

Content of the DMP

1 Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2 Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?
- 2.3 How will you handle copyright and Intellectual Property Rights issues?

3 Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4 Data sharing and reuse

- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.



DMP on mySNF

Applications and Projects

Grant application 27















- 1. Personal data
 - Responsible applicant
 - Other applicants
 - Applicants' employment
 - Project partners
- 2. Application data
 - Basic data I
 - Basic data II
 - Use-inspired project
 - Re-submission
 - Continuation of
 - Link to other SNSF projects
 - Further requested and available funds (not from the SNSF)
 - University or research institution
 - Requested funding
 - Data management plan (DMP)**
 - Research requiring authorisation or notification
 - Exclusion of external reviewers
 - General remarks on the project
- 3. Annexed documents (upload)

1. Data collection and documentation

  **1.1 What data will you collect, observe, generate or reuse?**



Questions you might want to consider:



- What type, format and volume of data will you collect, observe, generate or reuse?
- Which existing data (yours or third-party) will you reuse?

B /              

...

Briefly describe the data you will collect, observe or generate. Also mention any existing data that will be (re)used. The descriptions should include the type, format and content of each dataset. Furthermore, provide an estimation of the volume of the generated data sets. (This relates to the FAIR Data Principles F2, I3, R1 & R1.2.)

  **1.2 How will the data be collected, observed or generated?**

  **1.3 What documentation and metadata will you provide with the data?**

Data Life Cycle

- A “plausible” DMP is a condition for the release of the funds.
- DMPs are editable. Researchers have the possibility to update the content of their DMP at any time during the funding period of the research project.
- Once SNSF funding has ended and the final scientific report has been approved, the DMP cannot be modified anymore.
- The DMP is shared on P3 (SNSFs public database) at the end of a project.

Data Sharing

Data sharing policy

- SNSF expects data of a publication to be shared.
- Data needs to be shared as soon as possible on data repository, but at the latest at the time of publication of the respective scientific output.
- Additional data can be shared if the researcher wishes to do so.

Data repositories

Data repositories need to be digital and conform to the FAIR data principles.

Cost contributions

- A maximal cost contribution of CHF 10'000 per grant is installed for **data uploading costs** and related **data preparation costs** (prior to and for upload only).
- Data uploading costs are not paid if the data repository is commercial.
- Cost contribution can be exceeded, if justified.

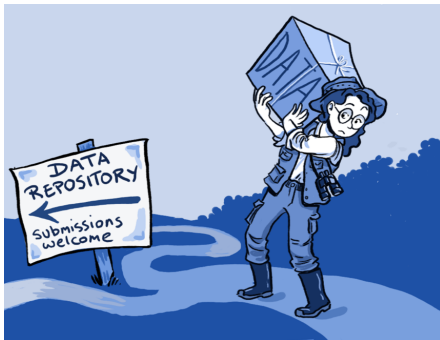
FAIR Data Principles



FAIR Principles - SNSF Checklist

- Globally unique and persistent identifiers for data (e.g. DOI, ARK)?
- Upload of descriptive and project-specific metadata?
- Data license (e.g. CC0, CC BY, etc.)?
- Metadata publicly accessible even in the case of datasets with restricted access?
- Structured and machine-readable metadata?
- Long-term preservation plan for data?

Data Repositories



Definition

- A data repository is an infrastructure allowing to deposit and/or to store your research data.
- Data repositories can be multi-disciplinary, institutional or field-specific.
- Some institutions deliver “seal of approvals” /certifications for data repositories, e.g. CoreTrustSeal
- FAIR Data Principles apply to data/metadata but data repositories should comply with these principles.

Registry of Data Repositories



Source: re3data.org

Examples of Data Repositories



Context in Social Sciences

- Important data repositories are available for the social sciences, especially in the United States (ICPSR), UK (UK Data Archives) and Germany (GESIS).
- In Switzerland, FORS is the main data repository for social sciences.
- Scientific journals which require data underlying the publication are sometimes associated with data repositories:
 - American Political Science Review (APSR) ⇒ Harvard Repository
 - European Economic Review (EER) ⇒ Mendeley

Context in Humanities

There is a lack of adapted data repositories in humanities:

- Heterogeneity in data formats and standardized vocabularies (ontologies), which reduces data interoperability
- Difficulty to store technically all data formats at the same place
- Data infrastructures are often the result of project-specific solutions

However, alternative solutions are proposed through distributed data architecture (e.g. CLARIN) or the semantic web (Open Linked Data) where data and metadata from different sources can be linked together (e.g. DaSCH).

Next Steps

- First experiences in the different SNSF funding schemes and monitoring (feedback to institutions)
- Collaboration with Science Europe and European Open Science Cloud (EOSC) Initiative
- Collaboration with swissuniversities: concept for data repositories in Switzerland
- Multi-annual plan SNSF: concept of SNSF actions for fostering data repositories

At the European Level - Science Europe



At the European Level - Science Europe

- Domain Data Protocols (DDPs) as a pragmatic solution to ensure proper implementation of individual DMPs
- Based on this general Framework, scientific communities are encouraged and enabled to set up protocols according to their specific needs (protocols as template for DMPs)
- Representatives of DARIAH, CESSDA and CLARIN are members of the Science Europe Working Group (WG)

Landscape Analysis

News room

› News Archive

Dossiers

Research magazine
Horizons

Media

SNSF Newsletter

Programmes Newsletter

Social media

RSS

Research database P3

NRP

NCCR

Call for tenders for a landscape and cost analysis of data repositories

Submission deadline

02/jul/2018

The SNSF, together with the programme «Scientific information» (P-5) of swissuniversities, launches a call to develop a concept that analyzes the current need for data repositories, shows the existing types and describes different funding scenarios. The submission deadline is 2 July 2018.

The report should deliver a multi-layered overview of national and international data repositories, their features, ways of operation and financial flows. Conclusions should contribute to possible solutions for supporting the establishment, repurposing or expansion of data repositories and contribute to their visibility.

The call is public and aimed at institutions specialising in issues related to research data management and data storage. The submission deadline is 2 July 2018.

SNSF Website Open Research Data

- Profile
- Organisation
- Evaluation procedures
- Partners
- Research policies
 - › Animal testing
 - › Gender equality
 - › International Collaboration
 - › Language policy
 - › Open Access to Publications
 - › **Open Research Data**
 - › Promotion of young researchers
 - › Scientific integrity
 - › Use-inspired basic research
 - › Statements and factsheets
- Jobs & mandates
- Contact
- Publications
- Events
- SNSF Blog

Open Research Data



Research data should be freely accessible to everyone – for scientists as well as for the general public.

The SNSF agrees with this principle. Since October 2017, researchers have to include a data management plan (DMP) in their funding application for most of the funding schemes. At the same time, the SNSF expects that data generated by funded projects are publicly accessible in digital databases provided there are no legal, ethical, copyright or other issues.

Please consult the webpages of the different funding schemes to see whether a DMP is required when submitting an application.

[SNSF policy on Open Research Data](#) ▾

[Guidelines and Regulations](#) ▾

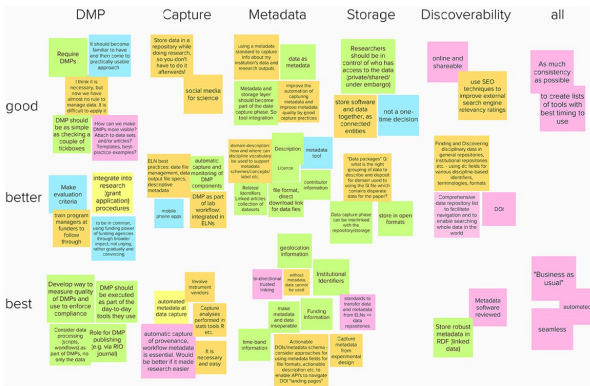
[FAIR Data Principles for Research Data Management](#) ▾

Contact

Open Research Data
E-mail: oro@snsf.ch

Context
 Research Data Management and Open Data
 SNSF Open Research Data Policy
 FAIR Data Principles
 Data Repositories
 Next Steps

It is simple, isn't it?



Questions?



Contact: ord@snf.ch