

Open Data and Data Management – Issues and Challenges

A retrospective view on this SCNAT Workshop for Swiss stakeholders held on 29 October 2018 in Bern

Friedrich K. Thielemann (President Platform MAP & Workshop convener)

Steps leading to the Workshop

- **We Scientist Shape Science Congress, 26/27 January 2017**
working group, convenors D. Wyler (swissuniversities), F.-K. Thielemann (Platform MAP)
- **Plenum MAP, 10 November 2017**
with inputs from Eva Moser, Cornelia Sommer, SNSF; Romain Tessier, open source, Alberto Morpurgo, open data; Nick Thomas addresses funding gap in space mission operations due to data management
- **Round Table International, 14 November 2017**
the funding gap for data management in space missions is addressed once more
- **MAP Presidium, 21 February 2018**
decision taken to have an open community meeting with SNSF and other representatives on the open data topic in October, extending beyond MAP and across all research areas of SCNAT, organizing group created around Francesco Pepe
- **Extended Board of SCNAT, 23 March 2018**
project presented with strong support by all attendees

Steps leading to the Workshop (2)

- **Round Table International, 30 April 2018**
plan presented getting strong support by all attendees with the advice to have a balanced and constructive representation of viewpoints and approaches
- **Active search for presenters and other participants during spring and summer 2018**
*with the aid of F. Pauss, H.R. Ott, F. Pepe, F.-K. Thielemann, SCNAT;
A. Kalt, M. Egger, SNSF; C. Rossel EU Open Science Advisory Panel/SCNAT;
and an extended community of Swiss scientists*

All these steps led to the one day Open Data event on 29 October 2018 in Bern

Agenda of the Workshop

09:30 Arrival and registration

10:00 **Welcome and introductory words** **Marcel Tanner**, President SCNAT

10:05 **Views and plans from the EU side** **Christophe Rossel** (EU Open Science Policy Platform)

10:25 **Views from the Swiss NSF side** **Matthias Egger** (President Research Council SNSF)

10:45–12:40 **Views and experiences in Switzerland** (from research groups representing the natural sciences spectrum from small university labs to large research infrastructures)

10:45 **Joël Mesot**, Director Paul Scherrer Institute

11:00 **Florian Altermatt**, Evolutionary Biology, University of Zurich

11:15 **Marcel Mayor**, Chemistry Department, University of Basel

11:30–11:55 Coffee break

11:55 **Robert Jones**, IT Department, CERN

12:10 **Nick Thomas**, Space Sciences, University of Bern

12:25 **Ana Sesartic Petrus**, Research Data Management and Digital Curation, ETHZ Library

12:40 **Global discussion on presentations**

13:10–14:10 Lunch

Agenda of the Workshop (2)

14:10 **Repartition in 4 working groups** with their conveners

- Data Management Plans: Practice and Challenges (**Ana Sesartic Petrus**, ETHZ Library)
- Special aspects of small scale university labs (**Heinz Gäggeler**, Chemistry & Biochemistry, Bern)
- Special aspects of large research infrastructures (**Francesco Pepe**, Astron. Observatory, Geneva)
- International competition (**Ruth Durrer**, Theoretical Physics and Simulations, Geneva)

15:10 **Report by working group conveners**

15:50 **Panel with representatives from SERI, EU Open Science Policy Platform, swissuniversities, SNSF** Moderation: **Gerd Folkers**, President Swiss Science Council

Gregor Haefliger, State Secretary for Education, Research and Innovation SERI

Christophe Rossel, EU Open Science Policy Platform/SCNAT

Patrick Furrer, swissuniversities

Angelika Kalt, Swiss National Science Foundation SNSF

16:50 **Closing and next steps** **Friedrich K. Thielemann**

17:00 *End of workshop*

First impressions

- The workshop on Open Data and Data Management gathered **~120 participants**.
- It permitted to have presentations on the **views of funding agencies**, specifically of the European Union (EU) and the Swiss National Science Foundation (SNSF),
- followed by the **views of representatives across the science community** (from astronomy/space sciences, biology, chemistry, climate modelling, multi-disciplinary research institutions – like PSI, to particle physics),
- **more specific exchanges of ideas in four distinct working groups**, open to all participants (practice and challenges, small university labs, large research infrastructures, international competition),
- **followed by a panel discussion**, essentially culminating in the question how additional efforts related to OD and DMPs can and should be funded.
- **Globally, it was a very successful event** enabling the sharing of partially contrasting opinions on the topic by all involved parties about one year after the SNSF had introduced obligatory Data Management Plans. **The discussions were constructive and open minded.**

A) Motivation for Open Data

- The Open Data topic has raised some enthusiasm leading in the EU framework to:
 1. **The FAIR data principles** (*Findable, Accessible, Interoperable, Reusable*)
 2. **The European Open Science Cloud** (EOSC) as a data repository
- The major driving ideas behind and arguments in favour when launching the Open Data initiative are:
 - **Publicly funded** research results should be **publicly available** (*but to which public?*)
 - **Verification/Reproducibility** should be made possible and falsification of data should be hindered/easily recognized
 - **Re-use of data by independent researchers** could lead to additional important findings ..., (*positive examples exist!*)

B) Questions, Challenges, Efforts

1. What is exactly meant by Open Data?

- a) *only published data appearing in figures and tables*
- b) *intermediate, pre-processed, partially analysed data*
- c) *all raw data*

2. To which communities should data be made open?

- a) *scientist from the specific research field*
- b) *scientists from other fields (transdisciplinary)*
- c) *students*
- d) *the general public*

This requires quite different procedures and a large variety in the efforts involved.

3. Which means are available to do so?

- a) *Which repositories?*
- b) *How can procedures be adapted to the needs of different research fields*
- c) *Has proprietary (commercial) software to be converted?*

B) Questions, Challenges, Efforts (2)

4. What are the limitations to Open Data?

- a) *embargo dates of research facilities
embargoes/delay until the original research group has finished a full analysis?*
- b) *personal data, Intellectual Property rights*
- c) *patents, proprietary software*
- d) *collaboration with industrial partners*
- e) **Are there limitations by sheer complexity and amount of data involved?**
- f) **Should “reciprocity” be enforced** for fair competition: *Providing data access only to researchers supported by funding agencies sharing uniform rules? Other agencies might limit openness to “sharing upon request” practice.*
- g) **Should the openness be controlled by a data request system?**
Like researchers producing data, shall archive scientists outline their research ideas and their motivation to obtain data? Who will evaluate this?

C) Preliminary Conclusions

1. Data Management Plans are a useful tool

- a) not only for the funding agencies (and possibly other researchers)*
- b) but also for use within the research groups to keep an overview and structure of obtained results*
- c) this enables later re-use and avoids loss of knowledge after graduate students have left*

2. To which extent should data be made public?

- a) published data appearing in figures and tables should be openly available*
- b) more extended public research data can lead to further progress in science (e.g. in space science, astronomy/telescopes, partially CERN, and clearly demonstrated in biology and medicine)*
- c) general applicability is however limited by several constraints (cf. previous slide). The scientific (sub)communities should be permitted to participate in outlining the appropriate procedures.*

C) Preliminary Conclusions (2)

3. Financial issues with respect to a full implementation of Open Data down to raw data

- a) Estimates from areas where this is done already lead to an increase of financial efforts of the order 15%.*
- b) The suggested 10 kCHF by the SNSF is thus probably not sufficient in all cases to cover the full costs.*
- c) While it is possible to apply for a higher funding from SNSF, the related budget should come from different sources than the research funding in order not to endanger research.*
- d) If possible, professionals rather than PhD students should do the data preparation work.*
- e) Can cost/benefit considerations be applied with respect to expected (low) use of raw data?*

D) Outlook

- The Workshop led to a **positive and constructive exchange of ideas, opinions, and facts.**
- The **community valued this possibility to listen from all sides.**
- **All inputs at the meeting are made public** on the SCNAT website.
- **The hope is that in the years to come** among SNSF and the different research communities **flexible and appropriate best practices will be found**, permitting to optimize the advantages of (A) while considering the constraints from (B) and (C).

Thanks go to all speakers and participants of the meeting!