Call for a differentiated and careful approach

SCNAT Workshop on Open Data and Data Management, Bern, 29 October 2018
Indeed, you have invited the right expert! ...

Directive on the re-use of public sector information

From Wikipedia, the free encyclopedia

(Redirected from PSI Directive)

Directive 2003/98/EC on the re-use of public sector information, otherwise known as the PSI Directive,[2][3] is an EU directive that encourages EU member states to make as much public sector information available for re-use as possible. Previous to the creation of this directive this area was left to member states to regulate. This directive now provides a common legislative framework for this area.

The Directive is an attempt to remove barriers that hinder the re-use of public sector information throughout the Union.
Are we talking about...

Open access to published data?

or

Open access to all data enabling an independent reproducibility of the published data by third parties?

or

Open data to not yet analyzed and published data ("live-access")?

PSI per year:

1000 articles per year
...each 4 figures, 2 tables
...each 10-1000 x raw-data
Some arguments in favour of it

• Researchers have a **natural interest** that their findings (= published data) are used and serve as a basis for further knowledge gain.

• Cumbersome digitizing from figures or typing from tables of a publication are **no longer up to date**.

• Effort to place these data in a repository together with reference to the publication for the purpose of explaining the background (= meta-data), represents **minor additional effort for researchers**

However

• National/European regulations should also demand international regulations

• Taxpayers may expect first a national exploitation of the published data
Highly successful example - Protein Data Bank PDB

Total depositions >150’000
Downloads/year > 650’000’000

www.wwpdb.org

Research lab

Data Acquisition

Data processing

Structural Model and density map

Deposition of data (obligatory for publication)

Entry of a new structure

X-ray facilities or cryo EM

open access Data Bank PDB

Megabyte to Terabyte depending on Method

Megabyte

Kilo to Megabyte
In principle, this allows for meta-analysis and thus
- enforce more transparency
- fight against fraud, lack of (research) integrity
- impair bibliometric impact as the common most important indicator
- avoid duplication of (research) efforts
- (enable faster dissemination)

However
- Ethics, value systems and self-regulation have evolved over more than a century from the scientific community - beware of disruptive top-down regulations!
- Regulations in practice sometimes difficult to implement - enormous effort (preparation, annotation with metadata, archiving, accessibility, quality assurance)
- Switzerland is in an international context (! other standards outside the EU)
Extract from a recent report of the European Commission (2017):

“Researchers and research organisations (including both learned societies and research-performing institutions) are seen as crucial participants in any decisionmaking process mapping future Open Science implementation and related training, so as to ensure successful and effective uptake by the research community.”

This very sentence is puzzling, as it stands only as point #6 (out of 8) in the list of “key concerns around open science implementation for member states”!

source: «MLE on Open Science: Altmetrics and Rewards – Incentives and Rewards to Engage in Open Science Ativities», p.16, European Commission, November 2017
Astrophysical observation

4 parameters:
- absolute time
- absolute position
- wavelength range
- resolution

«simple» although potentially more complicated

• telescopes...
  construction and operation details...
• stars and universe
  «sample preparation» unknown...

“Observation” in materials science

>>4 parameters:
- sample
- instrument
- environment
- data collection
...

• samples are prepared under complex conditions: composition, processing, quality, dimensions, absolute orientation
• instrument setup: energy and momentum ranges, neutron/photon energies, resolution conditions
• samples are exposed to in-situ conditions: temperature, pressure, electric and magnetic fields, chemical environment, time scale
• data is collected with a wide range of strategies: orientation, acquisition rates

traditionally the publication itself and the references to earlier publications represent the meta-data to the data / versus / stand-alone data enriched with meta-data
Data Production at the Swiss Light Source PSI

Detector | Data Acquisition / Readout | Real Time Feedback | Offline Data Analysis
---|---|---|---
SLS 01 | SLS 01 | SLS 01 | SLS 01
... | ... | ... | ...
SLS 16 | SLS 16 | SLS 16 | SLS 16
S-FEL 01 | S-FEL 01 | S-FEL 01 | S-FEL 01

Max. 50 GB/s per detector
Max. 100 TByte per experiment
1-3 days
7-8 PByte
1-6 months
80 PByte for 5-10 years
Reciprocity?

- Effort to enrich data with metadata for independent reproducibility -> partly gigantic -> competitive disadvantage for CH researchers

- How to handle contradicting regulations within international collaborations with different funding bodies? PSI publications: 73% international, 33% >5 institutions

- Commercialisation of raw-data (or published data) by “more protectionist” countries. Are we fine with an “RoI” in a third country?

- Disadvantages by setting the example?

Switzerland is not an island. In the international competition CH researchers must not be discriminated against by hurried obedience.

Similar topic:

**Open Access**: CH ranked 1st with 39% (cp. mean 30%) followed by Croatia, Estonia, Belgium, Great Britain (36%), USA (36%), ..., France (32%), Germany (31%), ..., Japan (27%), China (17%).

-> With whom do we want to measure ourselves?

Source: European Commission based on Scopus
Conclusions

• Open access to *published data*
  Yes! Desirable and already being implemented.

• Open access to *all data* for the purpose of independent reproducibility by 3rd parties
  Difficult to implement in practice, requires HUGE effort

• Open access to enable *more efficient use of data*
  *implement suitable measures for suitable data*
  (adapted to individual research areas)

Consider in all three points that:

• Switzerland is not an island
• Research is international
• Must avoid competitive disadvantage for CH researchers and CH economy