

Open Data, Data Management Plan & All That

Alberto Morpurgo

Department of Quantum Matter Physics & Group of Applied Physics

University of Geneva

*Working in the broad domain of Materials/Electronics
Leading a group of 10-12 young scientists*

*Since April 2017: Member of the Research Council
of the Swiss National Science Foundation, division II*

Disclaimer(s)

*What I will say represents **my personal opinion**.*

*Overall it is quite **representative** of the opinion of many colleagues in my research domain that I talked to during the last 6-12 months.*

***It does not reflect the official position of the SNF:**
in fact it is quite critical of it.*

*Probably I will make statements that are **NOT politically correct**:
my only intention is to convey a point of view so that it can be discussed.*

There is no established “truth” at this point.

The goal of this type of meetings is to openly bring up issues that should be seriously considered before we waste time/money/efforts

Open Data & Data Management

*The philosophy behind all of this in an ideal world
(or at least how it is currently presented/perceived)*

- *Data have great value:*

*It is the people who pay for research funded with public money:
People own the data and have the right to access the data freely*

- *The idea:*

*When the data are analyzed by someone other than the researcher
who collected them, they may lead to added value*

- *Therefore:*

*every researcher who receives public funds has to guarantee that
the data that he/she collects are accessible to the layman*

*(starting from data in publications, but in the ideal world this applies
to **ALL** data that are generated...)*

Slowly Now...

Let's look at this step by step

*It is the people who pay for research funded with public money:
People own the data and have the right to access the data freely*

Who dares challenging Democracy 4.0? But...Does it make sense?

Does it imply the most effective use of public money?

That is really what is the right of the people:
that we organize a system making the best use of available means

NO

It does not at all imply the most effective use of public money

There is no Free Lunch

Making data publically available according to **existing spirit/guidelines** of the open data policies (defined by the EU)

- 1) will cost a very significant amount **time, efforts, and money**
- 2) In many disciplines will have **marginal or no relevance**
- 3) Will have a negative impact on **young scientists**
(**where top young scientists will go** at an early stage of their career?)

*In general the cost will not cover any possible benefit
because this statement is too general and generic*

**Things function differently in different disciplines
It makes no sense to impose general regulations**

...without even mentioning all other problems

Technical...

every researcher who receives public funds has to guarantee that the data that he/she collects are accessible to the layman

For example:

In the real world, even if I would like to abide, I could not:

I cannot guarantee that my students document data as needed to make them accessible to an external third party

Rules should be applicable in practical situations

Existing regulations are written by technocrats: they have no idea of what “data” means and of how things functions in practice. How can they write meaningful rules? **Scientists have not been asked!**

*...without even mentioning all other problems
& less technical and more “conceptual”*

- *Do we want to give our data to our competitors?*

In many domains they will be the main users & research groups in USA/China/Japan will not reciprocate the favor

A PhD student spends year to produce data that are used by someone else elsewhere who takes all the benefits? It does not sound right

- *Open data policies separate data from those who produced them (you want to know that Barolo is Barolo when you “use” it)*

Going towards “Fake Data”?

What then? “Open data” makes no sense?

Imposing identical general rules for all fields of research

– from social science to medicine and physics –

makes no sense

There are cases in which “open data” is worth it

- Learn from existing examples

(in fields related to mine, for instance “crystallography”: the “.cis” file...)

- Identify the relevant domains
- Think of modalities
- Focus all support on those areas

Note that in Physics and possibly other technical sciences

- *“Open data” are implemented spontaneously* in domains where they are obviously useful/important

- *Step 1 to implement “Open Data”*: identify precisely which type of data in which domains and define a **standard data format**

In Summary about Open Data

- **General rules:**
 - * Administrators love them but **make no sense for open data**
 - * Ask **domain-by-domain** what is an effective policy
(and be ready to accept the answer)
 - * **Put efforts only in those domains where it makes sense**
- Current **SNF** websites on the topic rephrase **EU-generated policies**
 - * Not a surprise that **technocracy transpires all over**
- **Grand-vision and (pseudo-)planetary strategy**, but:
 - * **Not carefully** thought through on the technical side
 - * **Naïve** on the strategic side

(In the hands of persons who do not understand the implication of the rules they write)

Open Access & Open Data

- **Publications:**
 - cost of subscriptions to journals too high
 - role in evaluation of scientists and of their research
 - failing as a good vehicle to disseminate science

There is a **pressing problem** with “scientific publications”
Open Access is part of the equation to find a solution

- **Data:**
 - is there a problem to solve?
 - is there a pressing need?
 - “something to be gained”: has it been clearly identified?

There is **no real problem** with data

The Open Data policy is well on its way to create one

“Time is the most precious thing & there is less and less of it.

This is largely due to an
increasing volume of new tasks unrelated to research
often of questionable utility”

Imposing new constraints that sink time and remove flexibility
should be avoided at all costs

The opposite should be done: remove constraints

A key strength of Switzerland is that the country has common sense
It is vital that we keep it that way

Think first, then think again, then set the rules!

Questions?