

Pfeiffenberger, Hans, “Focusing on Social Constructs”

Briefing Document, NSF Workshop on April 26, 2010 “Changing the Conduct of Science in the Information Age”

Focusing on „Social Constructs“

Hans Pfeiffenberger,
Alfred Wegener Institut for Polar and Marine Research, Helmholtz Association
hans.pfeiffenberger@awi.de, www.awi.de/en, www.helmholtz.de/en

When we wonder how advancement of science came about, we may find as decisive the curiosity and openminded-ness not just of scientists but of society, including the willingness to spend money on science, and to provide it to specific people. On the other hand, scientists had to be confident that their achievements would be valued, intellectually and very practical as well. To keep both motivations in balance there had to be mechanisms to certify the quality of each incremental contribution and to make sure that each relevant piece of knowledge gained would contribute to the advancement of science as a whole.

We know that since the 17th century there has been an extremely successful mix of principles as well as their embodiment at the operational level: That each individual contribution has to be reproducible – peer review providing a proxy for this requirement in most cases - , and then re-usable – which is proven and acknowledged by citation of the work, when others build upon it. Around these “simple” constructs an ecosystem of self-organization of science and of service providers such as publishers and libraries evolved.

Let us acknowledge that this system is, to a non-negligible degree, based on trust. We trust that editors and reviewers maintain just the right amount of rigor in their task and that commercial entities and memory institutions together produce and maintain the records of science – all being overseen by “the” scientific community, represented, e.g., by learned societies and agencies and trusts (sic!) funding science.

How does this admirable system fare in the “information age”? Regarding the classical article it is being upheld – and fiercely so! BUT, in many disciplines or sub-disciplines the amount and import of information which is “off the records” of science, not available to peer reviewers, in many cases not even recorded in formal lab notebooks or laboratory information management systems, has increased dramatically. Whether it is data in all of its incarnations or software to implement models or data analysis: Its majority is not available, for all purposes of reproducibility or re-use by third parties.

This imperfect certification of results, as well as the incompleteness of the records of science as a whole, pose a significant danger: That the trust in the functioning and the results of science is being eroded. (The image of an iceberg [1] of unknown underwater extent comes to mind - a dangerous, colossal, beautiful challenge)

A new understanding of the way to conduct science in the information age needs to incorporate an appropriate recognition of making data available for reproducibility and re-use.

This has been addressed recently by learned societies and editorial boards in some (sub-) disciplines, e.g. [2], by requiring that underlying data or more details about

methods have to be supplied or published in parallel either before, when or immediately after an articles has been accepted.

It should be noted that

- in most cases there is no requirement (or possibility) for reviewers to look at these supplements during review
- it has been shown that mandates of this kind have frequently not been honored to a satisfactory extent [3]
- requiring data underlying specific articles may invite - in too many cases – delivery of (overlapping) fragments, but not of datasets re-useable as part of resource or reference data collections

Considering this and similar observations about disappointing adherence to weak or un-enforced Open Access mandates, one is lead to the alternative: Persuasive incentive.

Indeed, when we look beyond the review of articles describing conclusions from data, towards making data available for re-use, it will never be sufficient to rely solely on mandates, e.g., by funders requiring data management plans.

It will be necessary, more effective and - above all - consistent with the scientific method to expect and value the publication of data (and software) as potentially equivalent to articles about conclusions, methods, instrumentation, models, algorithms and whatever is considered a legitimate object of publication today.

In order to apply the concept of "Publishing" in its full meaning to data, we also recognize that it is not sufficient to put it online on some server (not to mention on a CD [4]) and to devise formats for the citation of data.

What is implicit in the concept of scientific publishing is the assessment and certification of quality, the provision of access to results and finally their preservation as "the scientific record". If these measures would be extended to data, strong incentive for sharing would clearly be present. How to provide certification will strongly depend on each (sub-) discipline and its practices. In some cases it may prove adequate to simply apply the well understood format and procedures of the scientific journal [5], which also provides an unmistakable signal to cite data in references. Elsewhere, the review may involve protocols or other collections of detailed documentation. This needs to be complemented by "brand named" data repositories or data libraries, which would be the other major source of trust.

[1] „Research Data: Unseen Opportunities An Awareness Toolkit“ commissioned by the Canadian Association of Research Libraries (CARL) (2009)
www.carlabrc.ca/about/working_groups/pdf/data_mgt_toolkit.pdf

[2] Whitlock MC, McPeck MA, Rausher MD, Rieseberg L, Moore AJ (2010), "Data Archiving", *American Naturalist* 175:145-146 DOI:10.1086/650340

[3] B. D. McCullough, „Open Access Economics Journals and the Market for Reproducible Economic Research“, *Economic Analysis & Policy*, Vol. 39 No. 1, March 2009

[4] Recommendation 7 , "Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft" - Vorschläge zur Sicherung guter wissenschaftlicher Praxis", / „Proposals for Safeguarding Good Scientific Practice“, DFG (1998)
www.dfg.de/en/research_funding/legal_conditions/good_scientific_practice/

[5] www.earth-system-science-data.net/general_information/about_this_journal.html