

Standards for the Conduct of Science in the Information Age

Martin Fenner, Gobbledygook

April 17, 2010

Assuming our airports are again open next weekend, I will be attending a meeting organized by the NSF (National Science Foundation) and EuroHORCS (European Heads of Research Councils) on Changing the Conduct of Science in the Information Age in Washington on April 26. We have been asked to submit a one page white paper in advance of the meeting.¹ I decided to focus on the importance of standards, obviously leaving out many other important technological and social aspects. But defining and adhering to standards will enable or enhance a number of very interesting ways to conduct and report science in the information age.

Improving the conduct of science through digital technology requires standards for linking to and formatting scholarly resources. These standards should be coordinated by independent organizations that are not restricted to geographic areas or particular research domains.

Data access

Digital Object Identifiers (DOIs) are the primary system to link to digital content. The International DataCite initiative is the DOI registration agency for scientific primary data. Although there are many uses of DOIs for primary research data (e.g. PANGAEA, earth system research), many systems still use different identifiers.

Research funders and journals working in specific domains should collaborate on standards and best practices for primary research datasets, and journal publishers should encourage or even require linking to research datasets from publications. Successful examples include GenBank (genetic sequences) and MIAME (microarray gene expression).

Knowledge access

DOIs have become the standard identifier for electronic scholarly publications and are managed by the CrossRef registration agency. Journal articles, databases and websites linking to scholarly publications should use DOIs whenever possible



Figure 1: Flickr photo by Joye~.

instead of internal identifiers such as the PubMed ID or direct links to publisher webpages. Publishers should implement citation styles that use the DOI instead of volume, issue and page numbers.

The NLM DTD is the standard format used by PubMed Central and many scholarly publishers to produce content for reading in the HTML, PDF or ePub formats. The Article Authoring Add-in for Microsoft Office Word and Lemon8-XML allow researchers to produce content in the NLM DTD format. The workflow of writing, reviewing and publishing scientific papers should be based completely on the NLM DTD and tools for collaborative writing, journal submission and peer review should be build around that format.

Attribution

The recently announced² Open Researcher and Contributor ID (ORCID) is one of many initiatives for a unique researcher identifier, but has probably the broadest support among institutions, publishers and research organizations. ORCID will be managed by an independent non-profit organization, and will allow the exchange of profiles with other researcher identifier systems such as those used by Scopus, RePEc, or Inspire.

The information in the author profile may be initially provided by an institution, society or publisher, but should eventually be claimed by the individual researcher because of privacy concerns and because automated author disambiguation is never 100% accurate. Attribution should include all aspects of scholarly activity, including curation of primary research datasets and peer review.

The Public Library of Science (PLOS) article-level metrics make available comprehensive information (citations, downloads, social bookmarks, comments, etc.) with every published article. This system should be linked to author identifiers and developed into a standard for scholarly resources. Other scholarly publishers and databases for primary research data should then adopt these metrics.

fn1. Cameron Neylon's draft white paper is [here](#).

fn2. Interestingly both DataCite and ORCID were first announced December 1, 2009 at two independent events in London ([press releases here](#) and [here](#)).