



INTERNATIONAL ASSOCIATION OF SCIENTIFIC, TECHNICAL & MEDICAL PUBLISHERS

www.stm-assoc.org

To: The EU Commission,

DG-CNECT-C3

Re: Consultation on Open Research Data, July 2013

Date: 15 July 2013

Introduction

The International Association of STM Publishers ([STM](http://www.stm-assoc.org)) welcomes the initiative by the EU Commission on Open Research Data. Representing publishing organizations who work closely with the research community on a global scale, we are well aware of the topicality and timeliness of this issue. In the past decade, STM publishers were already closely involved in numerous new initiatives in the area of data publishing. For the future we offer our experience and knowledge to the upcoming EU initiatives in this area. In particular we look forward to continue to contribute further to a well functioning communication research infrastructure in which research publications and research data are well connected and interlinked.

We appreciate the invitation by the EU Commission and DG CNECT for our input on this issue and wish to address in particular your questions on:

1. Data Re-use
2. Accessibility of stored data
3. Culture of sharing

Integration of Data and Publications

We believe that for the findability, accessibility, understandability and re-usability of research data it is paramount that data and publications are closely inter-connected and integrated. Links between data and publications must exist in a sustainable way ensuring longevity of open research data.

We believe that it is important that authors of research publications include identifiers and links to available underlying source data and that reversely, data centers include links from the data to all relevant publications in which these data are mentioned. STM is encouraging its members to make such links visible and easily accessible.

Reliable and trustworthy Data Centers are an indispensable entity in this information chain – the areas currently most advanced in data sharing all have excellent, well known and community endorsed data archives that are subject-specific. We think of GenBank, World Protein dBase, Pangaea, Cambridge Crystallography Data Centre and so forth. All of these work closely together with publishers on the integration of data and publications. In other subject disciplines that are or are becoming more data-intensive, the establishment of such subject specific expert Data Centers would be a welcome opportunity to achieve better data sharing.

It is no secret that among researchers there is at times serious concern for sharing their data, certainly if done too early before proper processing and validation of the data has taken place. If and when privacy-issues are at stake, such as often can be the case in medical data, or other legal issues, including patenting and intellectual property issues, embargos on the openness of data may be necessary. And then there is the more psychological issue concerning fear for misuse that others may make of the data. Persistent linking with and good citation of datasets in publications is one of the best ways to ensure that data are being understood properly, including their possibilities and their restrictions for further re-use. Proper data citation, with links to the unique researcher identifier (see also about ORCID here below) is a way to ensure that credit is given where credit is due and that the origin of data remains traceable. It will create new incentives for researchers to make their data available.

Data Curation and Actionable Data

For understandability of data and for its re-use, it is of great value to present data underlying a research publication in an actionable way. For example via data viewers and analytical tools and software applications that enable the readers of the publications to investigate the data deeper from their own perspectives. STM publishers can demonstrate how such actionable data have been integrated into publications in several exciting new initiatives recently.

For the evolving future research infrastructure, we also like to make the point that data sharing and data curation deserve to be treated as two distinct elements. Both are essential for better re-use of data, but may need their own approach in the development of policies. Without proper curation, most data will end up not re-usable. Varying types of reuse will require different levels of curation.

Particular subject niches of science have made great strides forward in sophisticated and yet practical data curation over the past decade or so, but the massive volume of the current data deluge across so many areas of science creates its own new challenge in terms of effectiveness, efficiencies and costs of data curation. For workflow integrated processing of data, for quality assessment and control, and for longterm preservation and re-use purposes, new solutions will be needed. In this context, the curation of data is only in its first footsteps. Publishers are investigating solutions applying longstanding experience in large-scale curation and processing of digital texts and related digital objects. They are actively seeking partnerships with the research data community to create more synergies also in this area.

A Role for the EU.

For the new initiatives, including the Data Pilots that the EU Commission is planning as part of Horizon-2020, STM publishers would be keen to contribute and to play a supportive role in the aim

to achieve better data sharing practice. Elements that would benefit from a clear steer by the EU Commission are in our view the development of more best practice examples for the integration of data and publications, of common conventions for data citation, for peer review of data, for persistent identifiers, novel approaches for data curation and for robust data archiving infrastructures that support the scholarly communication eco-system. Establishment of interoperable metadata standards is perhaps one of the highest priorities at this moment. More broadly, policy development would be welcome around standards, infrastructure, best practice recommendations as well as the creation of best practice champions and, not to forget, funding guidelines.

In the research workflow cycle, the moment of publication of research results is always a milestone of great significance – and at the same time a moment that authors can be incentivized to make the availability of their validated data part of the presentation of their research results.

The community of STM Publishers is keen to play a role in this and we would wish to encourage the EU Commission to adopt a multi-stakeholder approach comprising all players in the communication chain for scholarly information to develop and establish common practice approaches that strengthen the integral communication chain for research data.

STM Initiatives and partnerships in the area of Research Data

In this context we like to memorize that the first [STM statement on Open Data](#) stems already from 2007 in the Brussels Declaration:

Raw research data should be made freely available to all researchers. Publishers encourage the public posting of the raw data outputs of research. Sets or sub-sets of data that are submitted with a paper to a journal should wherever possible be made freely accessible to other scholars.

A statement on the desirability of bi-directional links between researcher validated data and publications was made by STM in 2012, together with [DataCite](#) and [Crossref](#), see at this [link](#). This statement encourages data deposits at trustworthy data archives and bidirectional linking via visible links between data and publications.

The STM Community was and is also a leading player among the initiators of [ORCID](#), whose unique researcher identifiers have a pivotal role to play in attributing data sets to its originators. The launch of ORCID, for which the publishing community provided start-up financial support, also showcases what successful public-private partnerships can achieve in this respect.

Recently, STM has also taken a leading role within new initiatives of the Research Data Alliance and the [World Data System of ICSU](#), especially in the area of data publishing workflows, bibliometrics, services and sustainable cost models. See for more about this, under this [link](#).

STM was also a project partner in the EU-cofunded project ODE, Opportunities for Data Exchange, whose results have been published on this [website](#) after its successful conclusion in November 2012.

The STM Community

[STM](#) is an international association of over 120 scientific, technical, medical and scholarly publishers, collectively responsible for more than 60% of the global annual output of research articles, 55% of

the active research journals and the publication of tens of thousands of print and electronic books, reference works and databases. We are the only international trade association equally representing all types of STM publishers – large and small companies, not for profit organizations, learned societies, traditional, primary, secondary publishers and new entrants to global publishing. Members represent all types of business models, including Open Access.

EU-based publishers publish 49% of all research articles worldwide (STM's members originate approximately 2/3 thereof), employing 36,000 staff directly and another 10-20,000 indirectly, and make a £3b contribution to the EU's balance of trade. STM publishers have actively embraced the opportunities of the digital online environment, starting with journal content and other "native" digital products such as software, data and databases, as well as other digital tools. For more than ten years now, science and medical researchers, along with medical practitioners, have had ubiquitous access to online tools that include published information, links between references in the literature, data sets and software that can be manipulated by the user, and visual supplemental information such as video and three-dimensional illustrations that can be viewed from different perspectives by the user.

For questions and more information: Eefke Smit, smit@stm-assoc.org

STM received input to this written contribution from:

STM, FEP, Wiley, e-Life, Elsevier, ACS