

STM input to the ALLEA open consultation on the recommendation of the ALLEA E-humanities working group

The International Association of Scientific, Technical and Medical Publishers (STM) is the leading global trade association for academic and professional publishers. It has over 150 members in 21 countries who each year collectively publish nearly 66% of all journal articles and tens of thousands of monographs and reference works. STM members include learned societies, university presses, both subscription and open access publishers, new starts and established players who work collectively to ensure broad access to and use of the latest scientific and scholarly information. The majority of our members are small businesses and not-for-profit organizations, who represent tens of thousands of publishing employees, editors and authors, and other professionals across the world.

STM supports members in their mission to advance research worldwide. Together as partners in science and policy, we serve society by developing standards and technology to ensure research is of high quality, trustworthy and easy to access. STM promotes the contribution that publishers make to innovation, openness and the sharing of knowledge. For us, supporting the growth and sustainability of the research ecosystem means embracing change. As a common good, we provide data and analysis for all involved in the global activity of research.

It is with this perspective in mind that we like to share our input and offer our community's expertise and knowledge to the ALLEA e-Humanities recommendation for FAIR Data. We structured our input following the outline of the online consultation document:

Why open Science

STM applauds the efforts by ALLEA to propose and develop recommendations to foster the FAIR principle for 'data' in the humanities.

The FAIR (Findable, Accessible, Interoperable and Re-usable) principles as the ALLEA document notes were first introduced in the life, natural and technology disciplines and seem to be portrayed in this document as fully understood and operational. However, efforts continue in those above-mentioned disciplines to make them fully operational. It is particularly the 'interoperability' and 're-usability' elements which present further challenges and therefore even groups in the natural sciences are talking about 'supporting' the FAIR principles rather than being FAIR 'compliant' at the current time. STM feels that more education is needed to fully understand the impact of the FAIR principles and their application to the broad and diverse field of humanities, especially in terms what would be required to make different kinds of humanities data interoperable and reusable.

To 'make' data FAIR in the humanities will be a huge effort for all stakeholders but in particular for researchers. STM agrees with comments in the online submissions that further detailed considerations should be given to the costs/benefits to support FAIR in humanities as well as various

level of FAIRness, and the necessary additional financial support for managing and preservation of data.

Identify

Comments in the online version raise the issue of the diversity and complexity that are inherent to 'data' in the humanities. STM agrees with many of those comments that highlight this issue and also feels that the classification into three categories might not appropriately reflect the scope of diversity. Unstructured data such as photographs, are more demanding in their management and need to be also considered within the recommendations.

Data Management Plans (DMP)

Publishers want to enable all of their authors and journals to publish the best research, which includes achieving community best practices in the sharing and archiving of research data. Publishers also aim to facilitate compliance with research funder and institution requirements to share data that are described in DMPs. Publishers support this by developing standardised research data policies that can be easily implemented on a journal level. One crucial element of such a research data policy is the Data Availability Statement (DAS) which provides information where data supporting the results described in a published article can be found. This includes where applicable, hyperlinks to publicly archived datasets analysed or generated during a study. The DAS is journal specific and a variety of templates are offered by publishers describing under what conditions the data can be accessed.

The DAS should not be confused with the DMP which are generally not required by publishers contrary to the comment in the consultation.

Publishers can play an important role helping facilitating FAIRer data in the humanities whilst also encouraging wider cooperation across the academic community. Publishers are part of an [interest group \(IG\) within the Research Data Alliance](#) where different aspect of data policy standardisation and implementation are/were addressed.

As an additional resource to add to this chapter STM suggests the following:

More discussion of publication considerations for data can be found here: <https://cos.io/blog/landscape-open-data-policies/>

Deposit and Share

It seems that there are some misunderstandings relating to the notion of copyright and data as expressed in the proposed wording on the "*long established limitation of copyright by the motion ...*" in the first paragraph. Copyright protects an expression of an idea or of facts, but not the idea or facts themselves. Database protection extends to datasets, either originally in their arrangement or selection, or presentation, or to non-original databases where only the extraction of substantial parts is prohibited, not the extraction of an individual data point.

Data as 'individual facts' are in and of themselves not copyrightable, therefore sharing or retelling without copying substantial parts of the expression/representation, following the FAIR principles is legally possible and does not hinder the proliferation of Open Science. The publication of data should follow the rule "as open as possible as closed as necessary" which is also the overall guidance provided by the next EU framework program Horizon Europe.

There are a variety of license options available in the market, and researchers should have the freedom to choose what they feel is appropriate to their data We caution that here are other legal

restrictions on reproduction, storage and disclosure of data, fact and ideas that may apply to research on a case by case basis.

Trustworthy repositories and permanent identifiers

As this consultation is about how to make data FAIR, the findability of data can be improved by linking between publications and the data itself. [Scholix](#) is a shared infrastructure for the linking of research data and scholarly literature (bidirectional linking). It is an important piece to enable Open Science and helps the realisation of the FAIR principles. It is an excellent example of successful collaboration between publishers, open repositories and data archives. STM is constantly working to increase the use of Scholix within the STM publisher community.

The online consultation highlights the importance of persistent (named in consultation as permanent) identifiers (PIDs). Beside their relevance for long-term preservation those PIDs are also needed for data citation. We recognize that challenges associated with data publication vary among different disciplines.

STM would recommend to add more information on best practice for data citation into the text of this chapter (<https://www.force11.org/datacitationprinciples>) .

Legacy data

The online consultation highlights the potential of, and challenges with managing legacy data, but unfortunately omits key aspects. Legacy data, like any other data may be protected by copyright or rights in a database (see above comments on deposit and share) and therefore a scrutiny check is required before using them. STM recommends that this should be included within the text.

General comment

While many aspects and recommendations for FAIR data in the humanities are discussed in this document, the impact of data citation is not discussed in detail. Recommendations on best practice for data citations (before, during and after publication) might be helpful for researchers and will support a wider dissemination of research results in this area.

Contact:

STM – International Association of Scientific, Technical and Medical Publishers
Barbara Kalumenos, Director of Public Affairs
Kalumenos@stm-assoc.org