Publishers

As a result of the journal-based dissemination structure of research, publishers are key stakeholders in the open access ecosystem. They are, thus, in a unique position, in cooperation with the rest of stakeholders to contribute towards a culture of openly sharing research data of high quality, linked to the publications they support, and fit for re-use. The publisher ecosystem is diverse, comprising institutions which are very small non-profit scholarly led, university-based operations, and small entrepreneurial ventures, as well as giant multinational enterprises that are central to the market.

Whereas publishers have placed a strong focus on open access to publications and open access as a business model, their engagement with research data and open research data in particular is relatively recent. Publishers are interested in research data and open research data because they add value to their main products (publications) by enhancing the trustworthiness of the published research through the ability to verify it, which lies at the heart of ethical conduct of research. Publishers are also increasingly developing policies as a response to the pressure from funders' policies in relation to open access.

This recent attention to research data is leading publishers to exploit the possibilities of research data in new data-based products and services, such publishing data journals, extending peer-review to research data, and offering services to enhance data quality. The emergence of data journals should be linked to the effort of publishing data separately that allow essential parts of the scientific record to be made available in an intelligible form to the scientific community. Data journals are community peer-reviewed open access platforms for publishing, sharing and disseminating data that cover a wide range of disciplines. The papers published contain information on the acquisition, methods, and processing of specific data sets. The published papers are cross-linked with approved repositories, citing data sets that have been deposited in such repositories or data centres. The publication of data papers can be considered as a good practice example of data management as it includes an element of peer review to the dataset, maximizes the opportunities for data re-use and provides academic accreditation to researchers. As data papers are becoming distinct publishing products, a number of data journals are also supporting alternative metrics (altmet-

rics), thereby enhancing further data publication. Recent emphasis on open access to research data and data publications brings to the fore the scientific quality of research data and the significance of research data peer-review. A further related topic is the citation of research data. Apart from data peer-review, publishers may contribute to the standardization of research data by gradually introducing policies that are compliant with current best practices.

Publishers are also turning their attention to include content discovery and linking services, as well as services that focus on exploiting content with text and data mining (TDM) tools. Increasing attention on TDM is a direct result of researchers' need to explore large databases of content, data and publications. Despite the estimated economic opportunities TDM can bring, the perceived threat by publishers towards allowing fully unobstructed TDM to be performed in their content has resulted in restrictive measures that limit researchers' abilities for cutting-edge computer-aided research.

In developing policies for open access to research data, peer–review of research data, and products/services such as data journals, it is understandable that publishers are required to collaborate closely with other important stakeholders. Close collaboration with data centers and repositories (data managers) is necessary, since the latter are the primary content holders in the case of research data, and thus the destination to which the publications provide links to for access to research data. Data managers are the guarantors for the technical quality, security, curation and preservation of research data. As publications increasingly involve mixing and linking papers and data, collaboration is required in establishing principles for standards that will guarantee the long-term access to high quality data. Finally, close collaboration is required between the publishers and the scientific community, such as scholarly societies and journal editors, in developing those editorial principles that promote citation of research data through the development of disciplinary-specific standards alongside internationally accepted principles as well as data review processes. Collaboration between publishers and funders is also essential, in view of the development of products and services that align to funder requirements.

Recommendations

1. Gradually develop mandatory policies for open access to research data supporting publications

   Editorial policies should address issues like documentation, metadata and format of published data, licensing, and citation. Editorial policies should be enhanced further through data availability statements provided both during the article submission process and the peer-review process. Policies should provision measures in cases of non-compliance brought to light after publication (such as retracting the published article).

2. Collaborate with certified repositories and data centers to streamline data submission

   Publishers are encouraged to collaborate with repositories and data centers that meet accepted criteria regarding their trustworthiness. For disciplines without community endorsed data centers/repositories, publishers can assist researchers by providing guidance and assistance on appropriate institutional repositories or commercial data services may be designated for deposit and access.

3. Support data as a first-class scholarly output through the establishment of peer-review processes

   Establishing peer-review processes for research data is a measure that contributes to the further enhancement of products of high quality. Peer-review processes should specify the criteria used relating to the technical aspects and quality of research data (completeness and consistency of dataset, appropriate standards, software used), while their scientific quality is assessed by the research community through pre- and post-publication peer-review.

4. Develop policies requiring citations for research data

   Publishers should require that data accompanying their publications are citeable, and provide clear guidelines for data citation. Data cita-

tion should include DOIs, as well as licensing information (e.g. Creative Commons licenses), preferably machine actionable, that informs users about what they are able to do with research data.

5. Establish licensing policies that encourage the use of TDM

Editorial policies should clearly state the licenses (default and recommended) under which the data are published and re-used. Taking into consideration the significant economic benefits that can be derived from the use of TDM tools publishers are encouraged to adapt their policies to allow for an increases use of such techniques in research.