Data Managers

The term data manager refers to those stakeholders within the open access ecosystem that are charged with the management of the scientific and/or cultural digital output in data: data centres, which are mainly government financed operations for making datasets available, libraries, archives and memory institutions that maintain collections of content. Some of them have developed strength in relevant technological infrastructures for the storage, curation and long-term preservation of digital data, while others are still lagging behind.

Data centres come in different forms and sizes and often emerge from a disciplinary community. Their most basic function relates to the storing of research datasets for a defined community and making them accessible for other researchers to discover and use. This entails two roles: firstly, ensuring that data is discoverable as well as ensuring the tools allowing other researchers to find and access them and, secondly, providing support services for researchers who need to get their data and metadata into shape prior to deposit.

Libraries traditionally provide access to resources and publications through subscription. With reference to research data, and under the pressure of research funders' mandates, they are gradually becoming involved in data curation, while being the primary training and information locus on this topic for the researchers, offering awareness-raising and advocacy services. Despite their eagerness to acquire an important role in the transition to an open access research culture, libraries currently fall short both in terms of their current practices, as well as in terms of meeting the demands of researchers and users in relation to the provision of data management and support services.

Irrespective of their character, data managers should address open access to research data as an important development towards open science and develop services to support the needs of their patrons. These services can be defined on the basis of their mission and context, and by establishing extensive collaborations with the research community and other important stakeholders in the scholarly communication ecosystem (research institutions, publishers, funders), as well as relying on current best practices and resources.

The costs of data management and curation services are a further issue that data managers should address. Data management costs are incurred by the acquisition, ingestion and access to data, personnel wages, training costs for researchers and (data) librarians, the technical infrastructure and out-reach programmes. Reliance exclusively on project funding is nonetheless problematic, as it does not guarantee long-term funds and, thus, operations. Consequently, developing sustainable funding models on the principles of diversifying sources of income and establishing collaborations should be addressed with particular care.

A further important contribution of data managers is towards the prolif-eration of high quality research data, i.e. securing the technical quality of research data. Data needs to be presented in standardized formats and ac-companied by appropriate metadata; if these conditions are not met data are hard to work with and require additional time and financial resources to make them accessible and usable. Several repositories and data centres have developed quality assurance measures and offer a range of services to evaluate the technical quality of data sets. These include providing pro-cess documentation, completeness/consistency checks, training on data management and sharing, file format validation, metadata checks, storage integrity verification and tools for annotating the quality information. In addition, numerous libraries and data centers have been experiment-ing with new mechanisms to enhance data quality through platforms for discussing data sets or offering tools for alternative metrics (altmetrics).

Data managers also have a role in the selection of data for long-term pres-ervation and retention. The gap between short-term access and long-term preservation to research data needs to be addressed, and emphasis needs to be placed on long-term preservation. The value of data is assessed both in terms of its technical as well as of its scientific quality.

Aside from the quality of the research data, the quality of services offered by data centres and repositories is becoming a cutting edge issue. Further-more, research funders and publishers are putting additional pressure by inquiring deposit in certified and accredited repositories, in an effort to secure the reusability and long-term preservation of research data. In such context, obtaining accreditation or certification to appropriate standards is a way for ensuring both the quality of data repositories and of the quality assurance process.

Finally, data managers have a significant role in providing training to researchers for meeting technical quality standards with their data sets, as well as in developing disciplinary standards. Data centers with expertise in data curation have an important role in enhancing the skills of research library staff in data managements, data quality and developing data services.

Recommendations

1. **Assess their position within the open access ecosystem in view of developing collaborative infrastructures and services**

   Research libraries and data centers are encouraged to evaluate their overall capacity and positioning within the open access ecosystem and assess the types of services to be provided, in collaboration with other stakeholders (research institutions, publishers and funders) and the research community.

2. **Develop sustainable business models to ensure long-term service provision**

   Planning for sources of income should be addressed efficiently and, as much as possible at the outset of service development, while the strategy should be reviewed at regular intervals. Acquiring income may require the diversification of income resources and the layering of the services offered, whereby some services incur charges for the users.

3. **Establish mechanisms for data quality that ensure re-use and long-term preservation through collaborative work**

   To ensure data quality for re-use and long-term preservation data managers are equipped with a range of quality assurance and control strategies. While they are expected to take the leading role in close collaboration with research communities (scholarly societies, research institutions and researchers) in establishing citation standards, their collaboration with funders, publishers and journal editors is central in ensuring the enforcement of relevant policies.

4. **Acquire certification/accreditation to guarantee high quality services in the long term**

   Establishing quality assurance mechanisms is important not only for the trustworthiness of research data but for the data centres hosting them. Data centres are thus encouraged to seek appropriate certification and accreditation guaranteeing the quality of their services, such as the Data Seal of Approval and/or other appropriate ISO certification.

5. **Support data management through the development of training programmes for researchers and librarians/ technical staff**

   Libraries should be minimally able to deliver training courses on DMP of general or discipline-specific nature to serve the particular needs of their research communities and librarians as well as more specialized topics like intellectual property rights, licensing, re-use of research data and ethical issues.