

STM is committed to supporting research data sharing

Sharing research data is crucial to advance science and research. It advances transparency, reproducibility, and provides additional opportunities for scientific discovery and collaboration. STM is committed to supporting data sharing to advance open scholarship.

Whether in the form of raw data, curated data, data collected in datasets or published in databases, research data has the potential to be a driving force for scientific discovery. Novel and promising techniques, fueled by the availability of data, could speed up discovery, improve analysis and make science more efficient. These include both “big data” science analytical tools and artificial intelligence. In addition to facilitating such applications, making research data FAIR (Findable, Accessible, Interoperable and Re-usable) plays a vital role in making science more transparent and improving the reproducibility of research.

Publishers are partners with scientists and researchers in our shared mission to advance research worldwide. At its heart, publishers seek to connect researchers, their research and the wider world. They continually innovate to add value into an increasingly digital and interconnected environment. To this end, publishers have founded crucial infrastructure initiatives such as [Crossref](#), and have both the experience and the expertise to further advance data sharing. Publishers are committed to supporting the ongoing transition to Open Science and scholarly communication. This includes providing a greater focus on the dissemination and use of research data. To this end, we are engaged in efforts to help researchers share, cite, and link data with all forms of research output as well as championing the introduction of data journals.

STM publishers have held a long-standing commitment to sharing data alongside publications, as seen in, amongst others, the [Brussels Declaration](#) and the [joint statement](#) between DataCite and STM publishers, which included best-practice recommendations to “make research data easier to find, link to, reuse and cite”. Publishers make important contributions to ensure data is FAIR, for example through the inclusion of data availability statements in articles, the development and roll-out of journal data policies, and the development of the SCHOLIX framework allowing for the sustainable establishment of links between articles and data. STM is an active member of various initiatives and organizations such as the RDA, is a [FAIRsFAIR Champion](#), and continues to play an active role in the development of standards and platforms. In light of its commitment to promoting the sharing, linking and citing of research data, STM declared 2020 to be the [Research Data Year](#). We expect this commitment to be the start of an ongoing several-year effort to advance research data sharing.

Although the requirements and practices greatly differ between various academic disciplines, STM publishers have so far adopted the following best practice principles:

- Journals should have Research Data Policies, stating the extent and manner in which published articles be accompanied by the underlying research data.

- Published articles should include Data Availability Statements, informing readers where the data associated with a paper is available, and under what conditions the data can be accessed.
- Research data should be stored in repositories that comply with community standards, such as the [CoreTrustSeal Trustworthy Data Repository Requirements](#).
- Articles and research data stored in repositories should be linked in a persistent and sustainable way, similar to the way articles are linked by means of citations using DOIs and persistent links
- Research Data should be cited according to standards appropriate to community norms

In our continuing efforts to further realize the potential of research data, STM continues to be guided by the following general principles:

- Publishers **support Open Science and the FAIR principles for the sharing of data.** We believe that data should be as open as possible, but as closed as necessary where privacy needs or proprietary interests are involved. Raw research data that are facts, and not intellectual property, should be made as widely available as possible without restriction. Publishers actively facilitate its dissemination, sharing, linking and citing on their own platforms and through partnerships with appropriate repositories, usually at no cost to the user. Developing, maintaining, curating and verifying databases for research data, on the other hand, often requires a considerable investment of human, technical and financial resources. To incentivize the investment needed, conditions have been defined in international laws and EU directives to foster the creation, maintenance, and ongoing development of these valuable assets to scientific advancement, including intellectual property law.
- We believe that the development of community standards and principles around research data can continue to be best achieved through a **collaborative, self-regulating effort** by participants in the scholarly ecosystem, such as funders, institutions, publishers and policy makers. If needed, these efforts should be subject-specific, as needs and requirements can differ between areas. At the same time, the development of regional-specific standards or principles should be avoided as this can impede international collaboration, which has always been a catalyst of scientific progress.
- Although significant progress has been made to make Research Data FAIR, more needs to be done to make data 'AI-ready'. The quality of data ingested by AI systems is paramount to the quality of AI output. This requires the development of tools to efficiently collect research data, the development of data standards, the structural application of metadata and contextual documentation, as well as the discoverability of research data in appropriate accessible repositories. To bring these to a mature level and thereby

make research data AI-ready, **significant investments are required** for which a sustainable solution should be found.

- An increased focus on making data more accessible and usable will not be successful without making corresponding **changes in the way we recognize, incentivize and evaluate researchers**. We believe that bringing this about requires a collaborative effort by funders, institutions, policy-makers and publishers, ensuring that making research data FAIR and AI ready is fully recognized as a scientific activity in accordance to its significance to scientific progress.