Research Data as Integral Part of Scholarly Publishing

Michael Diepenbroek

Data-intensive science is building on data sharing and integration of data from various sources. This – besides efficient usability of data - requires first and foremost a new transparency of scientific work. Recent surveys have shown the willingness of researchers to share their data but also their concerns. This clearly indicates that a cultural change in science is needed. This change is fostered through the idea of data publishing as a clear incentive for scientists to share their data. Only in the past few years have scientists began calling for data "citation" and referring to data "publication" rather than data "sharing" and "availability". Data publication can be similar to the conventional publication of articles in journals that includes online submission, quality checks, peer-review, editorial decisions, and an equivalent of 'page proofs'. In fact, data storage in authoritative open access databases is getting increasingly important or is already mandatory for the acceptance of peer-reviewed publications in specific research fields as e.g. molecular sciences or ecology. Data publication services need to be integrated into the traditional science publication process. This requires collaborations with science publishers and services to link articles and data on the editorial and for discovery of data. The impact on citation rates could be shown in bibliometric studies on science articles having supplementary data. Eventually, an index for science data would be needed similar to the Data Usage Index (DUI) or the newly propagated index for scientific data by Thomson Reuters. Data publishing is strongly supported by the ICSU World Data System (WDS), the Global Biodiversity Information Facility (GBIF), the European Commission, and further stakeholders.