Access and preservation of digital research content: Linked open data services - A research library perspective

Angelina Kraft, Irina Sens, Peter Löwe, and Britta Dreyer
TIB German National Library of Science and Technology, Hannover, Germany (angelina.kraft@tib.eu)

Globally resolvable, persistent digital identifiers have become an essential tool to enable unambiguous links between published research results and their underlying digital resources. In addition, this unambiguous identification allows citation. In an ideal research world, any scientific content should be citable and the coherent content, as well as the citation itself, should be persistent. However, today's scientists do not just produce traditional research papers – they produce comprehensive digital collections of objects which, alongside digital texts, include digital resources such as research data, audiovisual media, digital lab journals, images, statistics and software code. Researchers start to look for services which allow management of these digital resources with minimum time investment. In light of this, we show how the German National Library of Science and Technology (TIB) develops supportive frameworks to accompany the life cycle of scientific knowledge generation and transfer. This includes technical infrastructures for

• indexing, cataloguing, digital preservation, DOI names and licencing for text and digital objects (the TIB DOI registration, active since 2004) and

• a digital repository for the deposition and provision of accessible, traceable and citeable research data (RADAR).

One particular problem for the management of data originating from (collaborating) research infrastructures is their dynamic nature in terms of growth, access rights and quality. On a global scale, systems for access and preservation are in place for the big data domains (e.g. environmental sciences, space, climate). However, the stewardship for disciplines without a tradition of data sharing, including the fields of the so-called long tail, remains uncertain. The RADAR - Research Data Repository - project establishes a generic end-point data repository, which can be used in a collaborative way. RADAR enables clients to upload, edit, structure and describe their (collaborative) data in an organizational workspace. In such a workspace, administrators and curators can manage access and editorial rights before the data enters the preservation and optional publication phase. RADAR applies different PID strategies for closed vs. open data. For closed datasets, RADAR uses handles as identifiers and offers format-independent data preservation between 5 and 15 years, which can also be prolonged. By default, preserved data are only available to the respective data curators, which may selectively grant other researches access to preserved data. For open datasets, RADAR provides a Digital Object Identifier (DOI) to enable researchers to clearly reference and reuse data and to guarantee data accessibility. RADAR offers the publication service of research data together with format-independent data preservation for an unlimited time period. Each published dataset can be enriched with discipline-specific metadata and an optional embargo period can be specified.

With these two services, RADAR aims to meet demands from a broad range of specialized research disciplines: To provide a secure, citable data storage and citability for researchers which need to retain restricted access to data on one hand, and an e-infrastructure which allows for research data to be stored, found, managed, annotated, cited, curated and published in a digital platform available 24/7, on the other.