



Data-centric Science: New challenges for long-term archives and data publishers

Martina Stockhause and Michael Lautenschlager
WDC Climate / DKRZ, Hamburg, Germany (stockhause@dkrz.de)

In the recent years the publication of data has become more and more common. Data and metadata for a single project are often disseminated by multiple data centers in federated data infrastructures. In the same time data is shared earlier to enable collaboration within research projects.

The research data environment has become more heterogeneous and the data more dynamic.

Only few data or metadata repositories are long-term archives (LTAs) with WDS/DSA certificates, complying to Force 11's 'Joint Declaration of Data Citation Principles'. Therefore for long-term usage of these data and information, a small number of LTAs have the task to preserve these pieces of information. They replicate, connect, quality assure, harmonize, archive, and curate these different types of data from multiple data centers with different operation procedures and data standards.

Consortia or federations of certified LTAs are needed to meet the challenges of big data storage and citations. Data publishers play a central role in storing, preserving, and disseminating scientific information. Portals of these federations of LTAs or data registration agencies like DataCite might even become the portals of the future for scientific knowledge discovery.

The example CMIP6 is used to illustrate this future perspective of the role of LTAs/data publishers.