Geophysical Research Abstracts
Vol. 19, EGU2017-4465, 2017
EGU General Assembly 2017

Components for the Global Digital Object Cloud

Helen Glaves (1), Hilary Hanahoe (2), Tobias Weigel (3), Larry Lannom (4), Peter Wittenburg (5), Dimitris Koureas (6), and Bridget Almas (7)

(1) British Geological Survey, Natural Environment Research Council, Keyworth, Nottingham, United Kindom (hmg@bgs.ac.uk), (2) Trust-IT Services / RDA (h.hanahoe@trust-itservices.com), (3) German Climate Computing Center, Department of Data Management, Hamburg, Germany (weigel@dkrz.de), (4) Corporation for National Research Initiatives, Reston, VA, United States (llannom@cnri.reston.va.us), (5) Max Planck Computing and Data Facility, Garching, Germany (Peter.Wittenburg@mpi.nl), (6) Natural History Museum London, United Kingdom (d.koureas@nhm.ac.uk), (7) Perseus Digital Library, Tufts University, Medford, MA, United States (bridget.almas@tufts.edu)

We are at a tipping point in the development of a common conceptual framework and set of tools and components which will revolutionize the management of scientific data. It is widely acknowledged that the current volumes and complexity of data now being collected, and the inevitable and enormous increase in that volume and complexity, have reached the point where action is required. Around 80% of the data generated is being lost after short time periods and a corresponding amount of time is being wasted by reseachers on routine data management tasks. At the same time, and largely in response to this perceived crisis, a number of principles (G8, RDA DFT, FAIR) for the management of scientific data have arisen and been widely endorsed. The danger now is that agreement will stop at the level of principles and that multiple non-interoperable domain and technology specific silos will continue to arise, all based on the abstract principles. If this happens, we will lose the opportunity to create a common set of low-level tools and components based on an agreed conceptual approach.

The Research Data Alliance (RDA) is now combining recommendations from its individual working and interest groups, such as suggestions for proper citation of dynamic data or how to assess the quality of repositories, to design configurations of core components (as specified by RDA and other initiatives such as W3C) and stimulate their implementation. Together with a few global communities such as climate modeling, biodiversity and material science, experts involved in RDA are developing a concept called Global Digital Object Cloud (GDOC) which has the potential to overcome the huge fragmentation which hampers efficient data management and re-use. It is compliant with the FAIR principles in so far as a) it puts Digital Objects (DOs) in its center, b) has all DOs assigned PIDs which are resolvable to useful state information, c) has all DOs associated with metadata, and d) has all DO bit sequences stored in trustworthy repositories. The presentation will give an overview of the types of components involved, the corresponding specifications of RDA, and the concept of the GDOC.