3TU.Datacentrum serves data-curation needs of geodesy, hydrology, and climate science in the Netherlands

A. Princic, J.P. Rombouts, and M.M.P. de Smaele

(a.princic@tudelft.nl) Delft University of Technology, TU Delft Library, the Netherlands

3TU.Datacentrum (datacentrum.3tu.nl) is a portal for science- and technology research data in the Netherlands. It serves data-curation needs of scientists in different disciplines of engineering and science varying from hydrology, water management, electron microscopy, to remote sensing. This data center provides long-term archive and permanent access to research data, it enables sharing and reuse of research data, and provides tools for citation of research datasets. The versatility of disciplines and versatility of needs is a challenge of the data center. The solutions offered differ in the functionality offered to the user.

For a complex collection of simple datasets consisting of ongoing hydrological measurements from several project partners, using several instruments, on several locations, over a long period of time, there was a need for central storage in order to share the data and enable easy combination of data sets. The measurements are not repeatable and are also valuable for multidisciplinary research, which justified investment in the preservation and 'publishing' of the data.

For a simple collection of complex datasets consisting of measurements of the meteorological radar, obtaining high resolution temporal and spatial precipitation data including those non-detectable by standard meteorological radar like drizzle, there was a need to store the data due to volume. These data come from a single instrument and from a single organization. Sharing with other scientists was difficult, which justified investment in the preservation and 'publishing' of the data.

3TU.Datacentrum offered solutions for both cases. Standardized solution to both collections is the data model: datasets, instruments, locations and time frames are all objects on their own, with own metadata and interconnected through rdf relationships. These relations are exploited in the user interface for navigation and inclusion, e.g. of instrument and location metadata in the display of datasets. They are also used to generate ORE Resource Maps. Common solution to both collections is also the file format NetCDF. This is an open standard, primarily used for sets of multidimensional arrays with metadata included, enabling interoperability and easy use. The NetCDF format can be used by common tools and programming libraries.

Differences in solutions to both collections are in the functionality offered to the user. For simple datasets a Fedora repository is used where datasets of any format are allowed: alongside the NetCDF an xml version (ncml) is stored that can be converted to other formats like Excel. For complex datasets of IDRA an OPeNDAP framework is used, which allows querying within and across datasets. The complex IDRA dataset contains raw data files, one or several processed data files, a quick-look file, and a user-friendly diagram displayed alongside the other metadata. Research data of different collections, be it simple or complex, and from different science disciplines such as hydrology, geodesy and climate research can be safely stored in 3TU.Datacentrum and become available for future use by the scientific community.