



Data Management and Archiving - a Long Process

Petra Gebauer (1), Roland Bertelmann (2), Tim Hasler (3), Ingo Kirchner (1), Jens Klump (2), Nora Mettig (1), Wolfgang Peters-Kottig (3), Beate Rusch (3), and Damian Ulbricht (2)

(1) Free University Berlin, Institute for Meteorology, Berlin, Germany (petra.gebauer@met.fu-berlin.de), (2) German Research Centre for Geosciences GFZ, Potsdam, Germany, (3) Zuse Institute, Berlin, Germany

Implementing policies for research data management to the end of data archiving at university institutions takes a long time.

Even though, especially in geosciences, most of the scientists are familiar to analyze different sorts of data, to present statistical results and to write publications sometimes based on big data records, only some of them manage their data in a standardized manner. Much more often they have learned how to measure and to generate large volumes of data than to document these measurements and to preserve them for the future. Changing staff and limited funding make this work more difficult, but it is essential in a progressively developing digital and networked world.

Results from the project EWIG (Translates to: Developing workflow components for long-term archiving of research data in geosciences), funded by Deutsche Forschungsgemeinschaft, will help on these theme. Together with the project partners Deutsches GeoForschungsZentrum Potsdam and Konrad-Zuse-Zentrum für Informationstechnik Berlin a workflow to transfer continuously recorded data from a meteorological city monitoring network into a long-term archive was developed. This workflow includes quality assurance of the data as well as description of metadata and using tools to prepare data packages for long term archiving. It will be an exemplary model for other institutions working with similar data. The development of this workflow is closely intertwined with the educational curriculum at the Institut für Meteorologie. Designing modules to run quality checks for meteorological time series of data measured every minute and preparing metadata are tasks in actual bachelor theses. Students will also test the usability of the generated working environment.

Based on these experiences a practical guideline for integrating research data management in curricula will be one of the results of this project, for postgraduates as well as for younger students. Especially at the beginning of the scientific career it is necessary to become familiar with all issues concerning data management. The outcomes of EWIG are intended to be generic enough to be easily adopted by other institutions. University lectures in meteorology were started to teach future scientific generations right from the start how to deal with all sorts of different data in a transparent way.

The progress of the project EWIG can be followed on the web via ewig.gfz-potsdam.de