Geophysical Research Abstracts
Vol. 14, EGU2012-8393, 2012
EGU General Assembly 2012

Geoscience data: Defining policies and workflow tools for long-term storage of continuously and temporarily collected data

P. Gebauer (1), I. Kirchner (1), W. Peters-Kottig (2), J. Klump (3), R. Bertelmann (4), B. Rusch (2), D. Ulbricht (3), and M. Wattenbach (3)

(1) Free University Berlin, Institute for Meteorology, Berlin, Germany (petra.gebauer@met.fu-berlin.de), (2) Konrad-Zuse-Centre for Information Technology, Berlin, Germany, (3) German Research Centre for Geosciences, Potsdam, Germany, (4) Library Wissenschaftspark Albert Einstein, Potsdam, Germany

The intention of the project EWIG (**De**veloping **w**orkflow components for long-term archiving of research data **i**n **g**eosciences) is to support geoscientists in transferring their data in a standardized way for storage in digital long-term archives.

In the pilot phase test data are provided by two participating research institutions, both producing large amounts of data. One of them - Institut für Meteorologie, Freie Universität Berlin - provides continuously meteorological data, these data are measured every minute at several stations throughout Berlin. The other - Deutsches **G**eo**F**orschungs**Z**entrum Potsdam - conducts geophysical field experiments and thus produces data temporarily. The digital long-term archive test system is provided by an infrastructure facility - Konrad-Zuse-Zentrum für Informationstechnik Berlin.

Both use cases require definite ways for the digital preservation workflows. Policies for the workflow independent from the working area will be defined.

At the beginning of the project the actual state of the art in science data preservation policies has to be identified, gaps should be detected and analyzed, so that missing workflow components can be designed. Contact to other institutions, having already policies for their data lifecycle, is necessary to get an overview of existing operating procedures and data management software tools.

Assuring the usability of the archived data is necessary during all stages of the project. All information essential for interpreting the data has to be available in a simple way. Usability tests of the archive will be performed together with domain scientists as well as students (bachelor, master, graduate). Questions about the quality of data access, documentation, metadata, etc. should be answered.

User feedback and the knowledge of the consortium will be used to compose a university lecture or seminar series on digital data curation, so that future generations of scientists become familiar with the handling of their data without any difficulty. Data curation should be self-evident for scientifically working people.

Developing policies, documentations and a lecture concept is intended to advance the re-use of archived research data by the scientific community.