



## **Data Publication through the Scientific Drilling Database**

**J. Klump**, R. Conze and J. Wächter

GeoForschungsZentrum Potsdam (jens.klump@gfz-potsdam.de/Fax +49-331-2881703)

Projects in the International Scientific Continental Drilling Program (ICDP) produce large amounts of data. Since the start of ICDP, data sharing has played an important part in ICDP projects and the ICDP Operational Support Group has facilitated dissemination of data within project groups. Some of these data later became the basis of scientific publications, while others remained unpublished. However, in most cases the data sets themselves were not available outside of the respective projects. With the online Scientific Drilling Database (SDDDB, <http://www.scientificdrilling.org>) ICDP and GeoForschungsZentrum (GFZ) Potsdam created a platform for the dissemination of data from drilling projects. The thematic focus of SDDDB is on data that were used as the basis of a publication, but some published datasets are data publications in their own right.

To publish data requires that data are citeable. This means, a mechanism is needed that ensures that the location of the referenced data on the internet can be resolved at any time. In the past, this was a problematic issue because URLs are short-lived, many becoming invalid after only a few months. Data publication on the internet therefore needs a system of reliable pointers to a web publication to make these publications citeable. To achieve this persistence of identifiers for their conventional publications many scientific publishers use Digital Object Identifiers (DOI). GFZ Potsdam is a member of the project "Publication and Citation of Scientific Data" (STD-DOI), which is funded by the German Science Foundation. In this project the German National Library for Science and Technology (TIB Hannover), together with GFZ Potsdam, Alfred Wegener Institute (AWI) Bremerhaven, University of Bremen and the Max Planck Institute for Meteorology in Hamburg set up a system to assign DOIs to data publications.

In many online scientific databases, the access to data is through some kind of search interface, which may consist of a simple field for the entry of keywords or may offer more elaborate search criteria. However, in most cases database users do not know the precise contents of a database, yet no database is so perfectly comprehensive to offer datasets matching any search query. Database users would therefore rather "browse" the contents like in a catalogue. In the design of SDDB we therefore took a very different approach to other databases and designed the graphical user interface to contain dynamically generated catalogue listings and cross-links, e.g. between datasets and sample material, datasets and authors, datasets and parameters, etc., to allow the database user to browse through the SDDB contents.

Any sample taken in the field is taken from a geographical and geological context. Not every database user will be familiar with the locality from where the sample was taken. Visualization of this sampling context helps to assess whether the offered datasets are useful to the particular question asked, or which subset of data to choose. At present this geographical visualization primarily shows the sampling positions in their geographical context. Virtual globes, such as Google Earth, are useful and intuitive tools for geographical visualization. To show the sampling locations in Google Earth, SDDB offers a kml-file for download with every dataset. The kml-file can be automatically imported as 'place marks' and viewed in Google Earth. The "place marks" are interactive and act as links back to SDDB. In a second step we plan to add more specific maps which will be displayed alongside the data or as separate maps in an online geographical information system.