

The Scientific Drilling Database (SDDB)—Data from Deep Earth Monitoring and Sounding

by Jens Klump and Ronald Conze

doi:10.2204/iodp.sd.4.06.2007

Introduction

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, which provides the infrastructure for data capturing for many ICDP projects, has facilitated dissemination of data within project groups. However, unless published in journal papers or books the data themselves in most cases were not available outside of the respective projects (see Conze et al. 2007, p. 32 this issue). With the online Scientific Drilling Database (SDDB; <http://www.scientificdrilling.org>), ICDP and GeoForschungsZentrum Potsdam (GFZ), Germany created a platform for the public dissemination of drilling data.

Access to Monitoring and Sampling Data

Effectively publishing data requires that data are citeable and that the location of the referenced data is unique and retrievable in the long term. In the past, the internet was a problematic reference for data because URLs are short-lived; therefore, data publication on the internet needs a system of unique and persistent pointers to a citeable web publication (Lawrence et al., 2001; Klump et al., 2006). For their conventional publications many scientific publishers use Digital Object Identifiers (DOI) for web referencing. GFZ is a member of the project “Publication and Citation of Scientific and Technical Data” using DOI techniques (STD-DOI). In this project the German National Library for Science and Technology (TIB Hannover), together with GFZ, the Alfred Wegener Institute (AWI) in Bremerhaven, the University of Bremen, and the Max Planck Institute for Meteorology in Hamburg, set up a system to assign DOIs to data publications (Brase, 2004; Paskin, 2005). Since May 2006 TIB Hannover is the first DOI registration agency for scientific and technical data worldwide, and GFZ Potsdam is one of its publication agents. To emphasize that the uploaded data have become citeable publications, SDDB displays bibliographical citation data with every dataset and offers an automated export of the citation data into common bibliographical database software (e.g., “Endnote”, Fig. 1).

To make the exchange of data between databases easier, the database structure of SDDB is similar to the structure of

the PANGAEA® database (Diepenbroek et al., 2002) and to the Drilling Information System used in ICDP projects and on IODP mission specific platform expeditions (Conze et al., 2007).

Access to data in scientific databases is commonly 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, users rarely know the precise contents of a database, and yet few, if any, databases will offer datasets matching any search query. Database users therefore “browse” the contents. The design of the SDDB meets this challenge with its graphical user interface that contains dynamically generated catalogue listings and cross-links (e.g., between datasets and sample material, datasets and authors, datasets and parameters). The geographical and geological context of data is visualized in order to help the user to assess if the data are useful. At present visualization primarily shows the sampling positions in their geographical context. Virtual globes, such as Google Earth, are useful and intuitive tools for geographical visualization (Butler, 2006). To show the sampling locations, SDDB offers a file in Google Earth’s kml format for download. The kml-file can be automatically imported as “place marks” and viewed in Google Earth (Fig. 2). The “place marks” are interactive and act as links back to the SDDB. In a second step, it is planned to add more specific maps which will be displayed alongside the data or as separate maps in an online geographical information system.

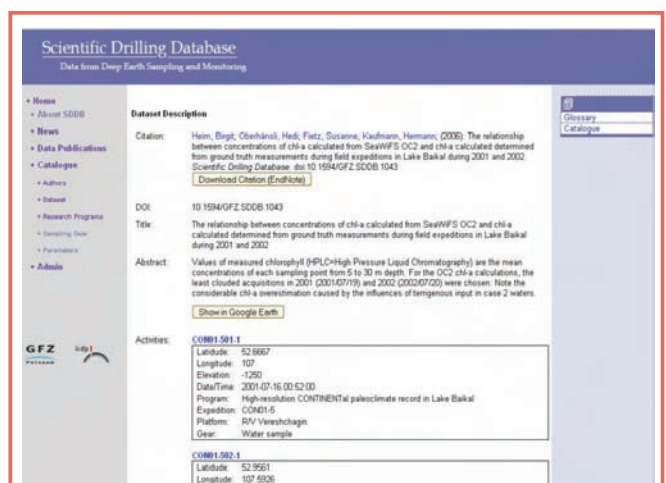


Figure 1. Screenshot of a data citation in SDDB. Note the buttons for download of citation into a reference manager and for the visualization of sampling locations in Google Earth. The Digital Object Identifier of this dataset is doi:10.1594/GFZ.SDDB.1043

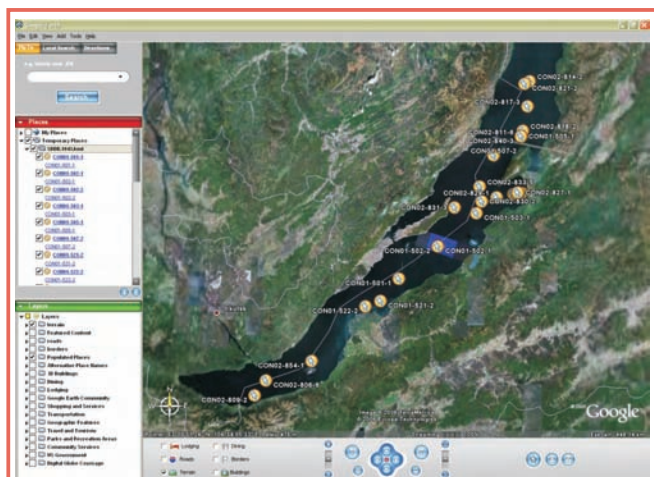


Figure 2. Sampling locations of doi:10.1594/GFZ.SDDB.1043 by Heim et al., 2005. displayed in the Google Earth virtual globe. The place marks at the sampling locations link back to an online description of the field activities at the respective locations as recorded in SDDB.

The Data Upload Assistant

The effort that goes into preparing data for publication has deterred many potential data contributors from sharing their data through scientific databases. The data upload process by the SDDB data upload assistant, a browser-based user interface, must therefore be as simple as possible. The upload process is divided into four steps. In the first step, all necessary metadata are collected that will be required in the course of the upload process. In the second step, the title of the dataset, a short description, selects the number of authors and the data file for upload. If the data are organized as a table, the author selects the principal investigator for each data column in the third step, followed by the parameter measured, the type of sample material, and the method used for this measurement. In the fourth and final step, the author selects the license (i.e. restricted, open access, creative commons) and copyrights under which the data are to be published and the publication date.

Future Developments

The development of the SDDB is ongoing. The focus in the near-term will be on further improving the data upload process, drawing from the experience gained so far. The development of the user interface will also be guided by feedback from the scientific community. In addition, the SDDB will be equipped with interfaces to exchange data and metadata with other scientific information portals, such as IODP's Scientific Earth Drilling Information System (SEDIS; <http://sedis.iodp.org>), the World Data Center (WDC) MARE/PANGAEA (<http://www.wdc-mare.org/>), the Sediment Geochemistry Database SedDB (<http://www.seddb.org>), the petrological database of the ocean floor PetDB (<http://www.petdb.org>), the stratigraphic database CHRONOS (<http://www.chronos.org>), and others, as well as to provide a metadata harvesting interface that follows Open Archives Initiative standards (see www.openarchives.org).

In the longer term, the goal is to develop a close integration of geological sample information, the data derived from measurements on these samples, and the published studies in which the data are interpreted.

Acknowledgements

The project "Publication and Citation of Scientific and Technical Data" is supported by the German Science Foundation, Libraries and Information Systems (DFG-LIS).

References

- Brase, J., 2004. Using digital library techniques - Registration of scientific primary data. *Lect. Notes Comp. Sci.*, 3232:488–494.
- Butler, D., 2006. Virtual globes: The web-wide world. *Nature*, 439 (7078):776–778. doi:10.1038/439776a.
- Conze, R., Wallrabe-Adams, H.-J., Graham, C., and Krysiak, F., 2007. Joint data management in ICDP and IODP expeditions. *Sci. Drill.*, 4:32–34. doi:10.2204/iodp.sd.4.07.2007.
- Diepenbroek, M., Grobe, H., Reinke, M., Schindler, U., Schlitzer, R., Sieger, R., and Wefer, G., 2002. PANGAEA - an information system for environmental sciences. *Comp. Geosci.*, 28 (10):1201–1210, doi:10.1016/S0098-3004(02)00039-0.
- Heim, B., Oberhaensli, H., Fietz, S., and Kaufmann, H., 2005. Variation in Lake Baikal's phytoplankton distribution and fluvial input assessed by SeaWiFS satellite data. *Global Planet. Change*, 46(1–4):9–27. doi:10.1016/j.gloplacha.2004.11.011
- Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., Lautenschlager, M., Schindler, U., Sens, I., and Wächter, J., 2006. Data publication in the Open Access Initiative. *Data Sci. J.*, 5:79–83. doi:10.2481/dsj.5.79.
- Lawrence, S., Coetzee, F., Glover, E., Pennock, D., Flake, G., Nielsen, F., Krovetz, R., Kruger, A., and Giles, L., 2001. Persistence of Web references in scientific research. *IEEE Computer*, 34 (2):26–31.
- Paskin, N., 2005. Digital object identifiers for scientific data. *Data Sci. J.*, 4:12–20. doi:10.2481/dsj.4.12.

Authors

Jens Klump, GeoForschungsZentrum Potsdam, Data Centre, Telegrafenberg A3, D-14473 Potsdam, Germany, e-mail: jens.klump@gfz-potsdam.de

Ronald Conze, GeoForschungsZentrum Potsdam, ICDP Operational Support Group, Telegrafenberg A34, D-14473 Potsdam, Germany.

Related Web Links

- <http://www.scientificdrilling.org>
<http://www.std-doi.de>
<http://sedis.iodp.org>
<http://www.wdc-mare.org>
<http://www.seddb.org>
<http://www.petdb.org>
<http://www.chronos.org>
<http://www.openarchives.org>