0.1 SedDB –Next Generation Data Management for Marine Sediment Geochemistry

K. A. Lehnert (1), **S. L. Goldstein** (1), R. W. Murray (2), N. G. Pisias (3)

1. Lamont-Doherty Earth Observatory of Columbia University, USA (lehnert@ldeo.columbia.edu, steveg@ldeo.columbia.edu)

2. Department of Earth Sciences, Boston University, USA (rickm@bu.edu)

3. College of Oceanic & Atmospheric Sciences, Oregon State University, USA (pisias@coas.oregonstate.edu)

SedDB is a project to build a data management and information system for marine sediment geochemistry that will provide new and novel access to data of fundamental interest to a wide range of research topics, from paleoclimate reconstruction to fluxes between the Earth's surface and mantle. SedDB will be modeled closely after the highly successful web-based igneous rock databases PetDB, GEOROC, and NAVDAT [EARC] that give investigators easy and fast access to comprehensive **integrated** global data sets over the web, allowing them to extract subsets of these data in any size within minutes via interactive interfaces where investigators or students can customize their queries according to their research problems or course projects. The integration of data distinguishes these systems from *data archives* or *data catalogs* that offer access only to individual data sets such as data tables from publications.

SedDB will address the needs of communities studying marine sediment chemistry and convergent margin volcanism as defined by a community workshop held in 2004. It will contain the full range of analytical values for rock samples from major and trace element concentrations to radiogenic and stable isotope ratios, and data for all types of material such as biological components, leachates, and size fractions. It will include other relevant data, e.g. bulk density, which are essential for mass balance calculations and accumulation rates, and it will include a wide variety of supplementary data (metadata) that are relevant to describe both samples and analytical data for proper

data evaluation and analysis. These metadata also constitute the criteria by which the user can select and sort samples and chemical data while interacting with the database. The structure of the relational database, which was designed for PetDB and GEOROC and has since been applied for a number of other geochemical databases, fulfills all requirements for comprehensiveness of primary and secondary data [LEHN]. It accommodates any type of chemical data for all the types of material, and includes information about the sampling location (e.g. latitude-longitude, verbal description of locations, tectonic setting), age, lithology (e.g. rock type, texture, mineralogy, degree and type of alteration), source (sampling expedition, archive), information about analytical methods and procedures, errors and precision, correction procedures (e.g. mass fractionation corrections for radiogenic isotopes and normalization values) and standard values, about the material that was analyzed, and the bibliographical information for the reference of each analytical value, or information about the data provider.

SedDB's user interface will be developed based on the query tools designed and implemented in JSP (Java Server Pages) for the new PetDB interface. This interface allows users to choose various criteria such as latitude/longitude, geographical names, rock type, cruise, author, and publication, either separately or combined, to select samples from the database. Users select the material for which they want data (e.g. whole rock, mineral, leachate, size fraction, fossil type) and any combination of desired data types, either as groups such as 'Major Oxides', 'trace elements', 'Stable isotopes', 'Radiogenic Isotopes' or individually (e.g., P_2O_5, Zr, or $^{87}Sr/^{86}Sr$) from dynamically generated lists that show to the user only those data that are available for the selected group of samples and material. Data can be filtered by compositional constraints (e.g. select data for samples with <200 ppm Ba) and by analytical method. The final query result is a data table that can be downloaded in XLS format and viewed on web pages. Each analytical value in the web-displayed table is linked to a pop-up window that presents detailed information about the analytical procedure for this measurement, including precision, reference standard measurements, and fractionation correction.

A feature that is so far unique to PetDB and which is based on PetDB's implementation of unique sample identifiers, is the display on the sample info page of all analytical data available for a sample, even if published in different papers or submitted by various investigators. SedDB will make use of the new IGSN (International Geo Sample Number) to allow integration of analytical data from different sources within SedDB, and with data hosted in other data systems. The user interface will provide tools for extraction of data in the depth (time) and geographic domain. Strategies for dealing with absolute sample matches (i.e. all samples are from exactly the same depths) and for data extractions where the user will specify what is considered to be the "same depth" will be developed as well as interpolation tools that provide users the options

for matching data sets based on user specified conditions (e.g. samples within specified depth intervals, averaging strategies, interpolation options). Additional tools will allow translating depth data into ages, with ages from original age assignments made by the scientists who published the data set, or from database tools where depth-age models can be specified through the user interface or extracted from depth-age data sets.

Map interfaces to the SedDB data set will allow the user to plot queries on maps via ArcIMS using XML strings that encapsulate the data. Sample information will include a URL to the host database to allow users to access detailed sample data from the map. An alternate map interface will be available through GeoMapApp [GMAP], a Java? application with which users can explore marine geological and geophysical data with an interactive map.

A special emphasis will be placed on interoperability of SedDB with other emerging data and information systems (e.g. ODP/Janus [ODP], CHRONOS [CHRO], PaleoStrat {PALEO], MARGINS Data Management System [MGDMS]) to help build a broader Geoscience Cyberinfrastructure in which data can be seamlessly shared and linked.

We expect SedDB to have the same ***broad and profound impact*** on the application of the geochemical data set of marine sediments for science and education as PetDB, GEOROC, and NAVDAT have had on igneous geochemical data:

- SedDB will provide easy on-line access to an integrated data set allowing researchers, teachers, and students to explore both data and metadata, and any subset of the data customized to their specific research problem or teaching project.

- SedDB will preserve data that are currently in danger of being lost due to incomplete publication. This will provide new opportunities for more critical data and hypothesis evaluation and application of new data analysis methods to existing data, and it will reduce redundant efforts in data acquisition.

- SedDB will ensure proper documentation of metadata that are critical for proper evaluation of the data, and will guarantee their future validity and accessibility.

- SedDB will allow integration of the geochemical data with other data types, either on the basis of individual samples, by linking chemical analyses to other sample data, or on the basis of entire data sets, through interoperability with complementary data systems, thus facilitating new multi-disciplinary approaches in data interpretation.

The pilot phase will initially focus on collecting all of the available data for the Equatorial Pacific and the MARGINS Subduction Factory focus sites associated with the Izu-Bonin-Marianas and Central American arcs.

1 References

[CERV] *Cervato, C., Goldstein, S. L., Grossman, E. L., Lehnert, K., McArthur, J. M., ¿*Joint Discussion of Sedimentary Geochemistry Data Management Systems That Cross the Waterline", EOS Trans. AGU, 85, 44, 2004.

[CHRO] CHRONOS: An Interactive Network Of Data and Tools for Earth System History, http://www.chronos.org

[GMAP] GeoMapApp, http://www.geomapapp.org

[EARC] EarthChem: Advanced Data Management for Solid Earth geochemistry, http://www.earthchem.org

[LEHN] *Lehnert, K.A., Su, Y., Langmuir, C.H., Sarbas, B., Nohl, U*., "A global geochemical database structure for rocks", G^3, Volume 1, 2000.

[MGDMS] Marine Geoscience Data Management System, http://www.marine-geo.org

[ODP] Janus Web Database, http://www-odp.tamu.edu/database/

[PALEO] PaleoStrat: A Paleontological and Stratigraphic Information System, http://www.paleostrat.com

[SESAR] SESAR: Solid Earth Sample Registry, http://www.geosamples.org