**Kerstin A. Lehnert**(1), **Leslie Hsu**(*1), **Tiffany A. Rivera**(2), **J. Douglas Walker**(3)

* Corresponding author: lhsu@ldeo.columbia.edu (1) Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA (2) Westminster College, Salt Lake City, UT 84105 (3) University of Kansas, Lawrence, KS 66045, USA

# IEDA EarthChem: Supporting the sample-based geochemistry community with data resources to accelerate scientific discovery

immediate

April 28, 2020

## 1 Abstract

Integrated sample-based geochemical measurements enable new scientific discoveries in the Earth sciences. However, integration of geochemical data is difficult because of the variety of sample types and measured properties, idiosyncratic analytical procedures, and the time commitment required for adequate documentation. To support geochemists in integrating and reusing geochemical data, EarthChem, part of IEDA (Integrated Earth Data Applications), develops and maintains a suite of data systems to serve the scientific community. The EarthChem Library focuses on dataset publication, accessibility, and linking with other sources. Topical synthesis databases (e.g., PetDB, SedDB, Geochron) integrate data from several sources and preserve metadata associated with analyzed samples. The EarthChem Portal optimizes data discovery and provides analysis tools. Contributing authors obtain citable DOI identifiers, usage reports of their data, and increased discoverability. The community benefits from open access to data leading to accelerated scientific discoveries. Growing citations of EarthChem systems demonstrate its success.

Three main points

1. Integrated geochemistry databases enable new science findings in diverse topics 1. EarthChem maintains a data file repository, synthesis databases, and a data portal 1. Disciplinary focus and user community input improve usability

## 2 How do integrated geochemical data systems enable new scientific discoveries?

Geochemical compilations of enormous numbers of dense, statistically significant measurements have driven large, global-scale scientific discoveries. Examples include studies on diversity in MORB composition (Gale et al., 2013, e.g.), global distributions of elements in the Earth's derma layers (Rauch, 2011) and global patterns of intraplate volcanism (Conrad et al., 2011). New analytical methodologies allow for increasing rates of data collection that should translate to more ground-breaking scientific discoveries. With this anticipated increase, it is not feasible for single scientists to compile "all available global data" from the existing literature. This inability highlights the need for data systems to provide support for data discovery, access, and analysis to investigators, who are otherwise left with a disorganized heap of un-usable data.

The IEDA (Integrated Earth Data Applications) EarthChem data facility (http://www.earthchem.org) develops and operates digital data collections focused on the geochemistry of rocks and sediments from a wide range of global geographic settings. EarthChem citations show that its use is extending far beyond its rock and sediment geochemistry origins (http://www.earthchem.org/citations). For example, EarthChem has been cited in diverse scientific studies such as prediction of natural base-flow stream water chemistry (Olson and Hawkins, 2012), a prototype of a web-based relational database for archaeological ceramics (Hein and

Kilikoglou, 2011), and strontium and oxygen isotope fingerprinting of green coffee beans and its potential to proof authenticity of coffee (Rodrigues et al., 2010).

The citations, both within geochemistry and petrology or extending to new innovative uses, demonstrate the utility of the databases to the scientific community. However, the utility comes only after much work to address the challenges and complexities of data and information standardization, lack of investigator contributions due to lack of time or willingness, time needed for organizing data extracted from the literature, and the development and maintenance of systems that are useful to and used by the community. Geochemistry is an example of a discipline in the "long tail" of data (Heidorn, 2008), where individual investigators and labs hold troves of data collected with one-of-a-kind newly developed techniques. This type of data has its own unique issues in data system development. Disciplinary expertise is extremely helpful for proper documentation of data and associated metadata for reuse. A recurring theme is how to balance quality control with the amount of documentation provided, while giving proper credit to the investigators who originally obtained the data.

In this contribution, we describe the origin and current capabilities of IEDA EarthChem resources for sample-based geochemical data, list the benefits of those resources for scientists, and highlight some of the derived scientific results. We describe the options available to investigators for submitting their data to the system and opportunities for scientific attribution. We show how EarthChem has addressed the challenges related to long-tail scientific data management and contributed to scientific output.



# 3 Background and history of sample-based geochemical databases

A sample-based data system stores observations that come from discrete samples, such as rocks, sediment, fluid, or other materials. Analytical measurements of the samples, descriptions of sampling location and techniques, analytical procedures of data collection, and pre-analysis sample preparation are stored in an integrated manner. Here, integration means alignment and standardization of vocabularies, sample names, and output. Multi-layered and interrelated pieces of information create additional challenges when compared to grid-based sensor data, (e.g. satellite, seismic, elevation) which may have better standardization and data formats. Sample-based databases often grow from single investigator interests and efforts, slowly gaining traction, data, and users, until they are morphed into an online, accessible system.

One of the first online sample based geochemical databases was PetDB, the Petrological Database, formerly the Petrological Database of the Ocean Floor. The database was built on a sample-based data model (Lehnert et al., 2000), which served as a foundational structure for several disciplinary databases that developed in the following decade, including SedDB (Lehnert et al., 2005), GEOROC (Sarbas and Nohl, 2009), NAVDAT (Walker et al., 2006), and VentDB (Mottl, 2012). These databases combine data from numerous sources into a single relational synthesis database, allowing the rapid production of integrated datasets, and significantly reducing the time commitment that was previously necessary to manually compile the same data from the original sources.

The state of the art of geochemical data publication was laid out a decade ago by Staudigel et al. (2003) with the goal of initiating discussion of data formats and metadata in geochemistry at the "earliest stages of [geochemistry's] exploitation of Information Technology". Staudigel et al. (2003) highlight complexities within the organizational structure relating to standardization, conventions, lack of tabular data, and incomplete metadata. These issues have not disappeared, but management and mitigation have significantly improved and evolved. In the last decade, improvements such as governmental data policy statements [e.g. U.S. Office of Management and Budget Memo Open Data Policy—Managing Information as an Asset (M-13-13) [http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf], endorsement of best practices, and stricter rules regarding data reporting were implemented by editors, reviewers, professional societies, and funding agencies (e.g. CODATA Scientific Data Policy Statements). Editors from several peer-reviewed journals that publish manuscripts containing geochemical data agreed on minimum standards for documentation about data quality, sample information, and the format and accessibility, which was published as the Editors Roundtable document "Requirements for the Publication of Geochemical Data" (Goldstein et al., 2014). The recommendations have been implemented by some journals, but strict enforcement is not yet common.

Data management software that works directly with the laboratory equipment is one of the most efficient ways to overcome the hurdle of initiating data management. In addition to the development of suggested reporting norms for geochemical data, the Geochron (www.geochron.org) software works directly with mass spectrometers and reduction programs in order to retain the essential sample metadata (Bowring et al., 2011; Walker et al., 2011). The automated software improves the workflow and streamlines the metadata preservation process by bringing data directly from the machine to data management and visualization software on the computer. Software of this type has greatly increased the ability of scientists to collect, manage, and publish data that can be easily contributed to sample-based databases.

While these types of software programs provide ease and accessibility to the instrument users, the maintenance of the hosting database is commonly performed by a different entity. Because the hosting database is connecting data from various input sources, maintaining integrated synthesis databases is an arduous task that involves sustaining controlled vocabularies, obtaining data from authors, and tracking data in a way that captures the complex metadata relationships. Increasingly, investigators are seeking rapid publication of their data, along with the ability to search multiple disciplinary databases at once. In order to address these needs and provide useful search and discovery tools, EarthChem has built several complementary systems to support its user community.



# 4 Data publication, access, and citation: The EarthChem Library

Data publication, open access, and citation are the foci of the EarthChem Library (ECL). The ECL holds digital resources submitted by investigators, accompanied by a brief profile of informative and searchable documentation based on the DataCite standard (Starr and Gastl, 2011). Investigators submit their digital file, which becomes publically available and assigned a DOI (Digital Object Identifier). The ECL is committed to open access and long-term availability of datasets that might otherwise be behind a pay-wall, or inaccessible on a personal hard drive. Further, the ECL accepts all types of data, including images and software packages, and does not limit submissions to rock geochemistry. This feature allows the ECL to be open to a wide range of interdisciplinary scientists.

The ECL houses several compilations such as a mid-ocean ridge basalt compilation (Class and Lehnert, 2012), the VentDB hydrothermal vent chemistry compilation (Mottl, 2012), and the results of the geochemical IEDA Data Rescue Mini-Awards, which are compilations of career-spanning datasets. These include datasets from the Fiji, Izu, and Endeavour segment arcs (Gill, 2014) and Lunar geochemistry on samples from the Apollo missions (Delano, 2013a,b). Other types of resources include software like the Arc Basalt Simulator, an Excel macro that calculates chemical compositions of arc-derived basalts (Kimura, 2012). EarthChem dataset DOIs can be used to provide bi-directional linking from articles to supporting datasets, including those that will not fit in supplemental material size or format requirements.

To facilitate submission to the ECL, IEDA supports scientists by providing suggested templates for common analytical techniques (e.g., major, trace, and isotope analysis) that guide the investigator towards the necessary information needed for data preservation within the ECL. The submitted files are reviewed for ECL standards by an IEDA geoscientist. This subject knowledge is an advantage of disciplinary data repositories and promotes trust from the community that published data are of high quality (Lehnert and Hsu, 2013). The ECL is an efficient way to make datasets publicly available and meet data release requirements from funding agencies. Whether linked to formal written publications or not, the reliable retrieval and improved citability of ECL data resources is important for the scientific community.

# 5 Data integration and standardization: Topical databases PetDB, SedDB, VentDB, Geochron, and NAVDAT

Topical synthesis databases provide integrated geochemical data for a specific topic of interest. Integration of geochemical data means that measurements, standards, and preparation are documented in a standard way, and delivered to the user in a single table from up to hundreds of original publications. Tabulation of such data by hand may take years, but synthesis databases have the benefit of years of operation. Examples of specific topics include ocean floor geochemistry PetDB, geochemistry of sediments SedDB, hydrothermal vent systems VentDB, xenoliths (Deep Lithosphere Data Set - now part of PetDB), Ar-Ar geochronology (Geochron), and North American volcanic rocks (NAVDAT) (**Table 1**).

Integrated data are used in compilations to examine larger scale geochemical trends. Examples of such compilations that have used PetDB include "Origin of a 'Southern Hemisphere' geochemical signature in the Arctic upper mantle" (Goldstein et al., 2008), and "Composition of the depleted mantle" (Salters and Stracke, 2004).

Each EarthChem topical database arose from a user community that determined standards, features, and scope of the desired data. Different data have different parameters that are important to the investigators when selecting and viewing the data. For example, in SedDB there is depth in core and the ability to accommodate different materials that are sampled such as wood and paleosol. In VentDB for vent fluids, there are vent flow types and end-member and raw data used to calculate the derived values. In Geochron, the system works directly with data reduction programs and data from the instruments that produce geochronology and thermochronology data.

The sample-based data is stored in a data model detailed in (Lehnert et al., 2000). Data managers enter analytical data and related sample metadata from the published literature into the topical database. Metadata (data about the data) include data quality and data type-specific information, such as the laboratory, sample preparation, and instrument used for analytical measurement. Standards and references used to translate raw measurements into chemical composition, which may be lab specific, are documented within the database when reported by the investigator in the original publication.

A database user creates an integrated dataset by first selecting samples from the database that meet particular constraints on geographic location, physiographic feature, rock classification or other grouping, then selecting the chemical parameters of interest. Parameters include major elements, trace elements, stable isotopes, radiogenic isotopes, rock mode, rare earth elements, and volatiles. Outputs can be further filtered by analytical method and compositional range.

Table 1: Table 1. Summary of major IEDA EarthChem systems

| System | URL | Description |
|---|---|---|
| EarthChem Library (ECL) | www.earthchem.org/library | Digital data repository, DOI assignment, long-term archiving |
| PetDB | www.earthchem.org/petdb | The Petrological Database: ocean floor rocks and xenoliths |
| SedDB | www.earthchem.org/seddb | Sediment geochemistry database |
| VentDB | www.earthchem.org/ventdb | Hydrothermal vent database |
| NAVDAT | www.navdat.org | North American Volcanic Database |
| Geochron | www.geochron.org | Management for geochronology |
| EarthChem Portal | www.earthchem.org/portal | Portal accessing PetDB, SedDB, GEOROC, GANSEKI, USGS |
| SESAR | www.geosamples.org | System for Earth Sample Registration, geosample registry to a |

# 6 Data discovery and analysis: The EarthChem Portal

The EarthChem Portal emphasizes data discovery and analysis by searching over several federated databases, returning integrated results, and providing data analysis tools for working with the data. The Portal provides an integrated dataset from topical synthesis partner databases GEOROC, USGS, NAVDAT, PetDB and SedDB (**Figure 1**). In addition, links to relevant data in the affiliate database GANSEKI (http://www.godac.jamstec.go.jp/ga are provided. The search and output procedure is similar to that from the topical databases but feature additional tools for data visualization and analysis.

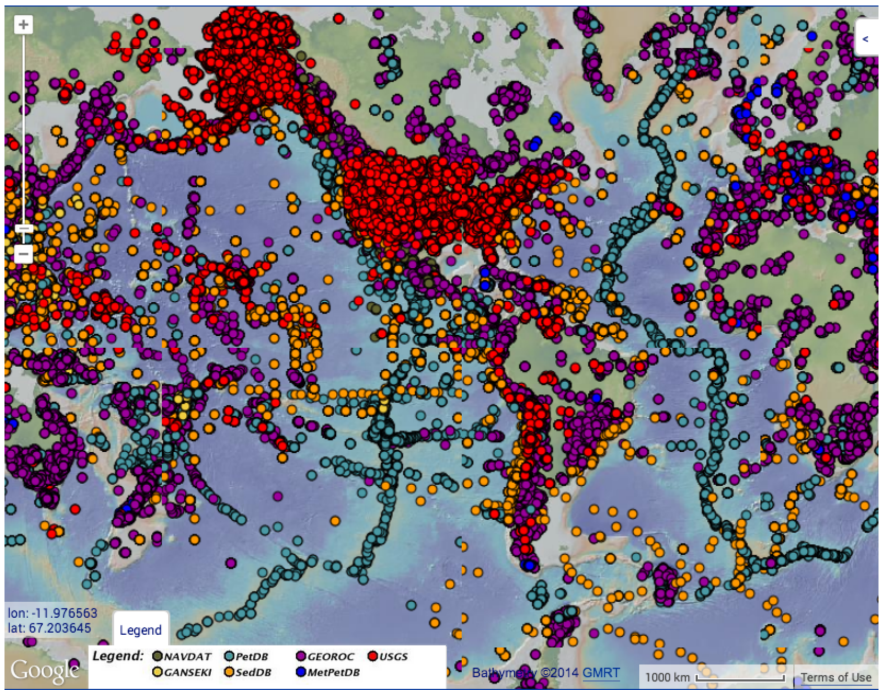


Figure 1: **Figure 1.** Map of sample locations from the federated databases of the EarthChem Portal, generated from the IEDA Data Browser (accessed 29 December 2014).

Recent examples of papers that cite the Portal include a publication on the lithospheric evolution of the Earth from a statistical review of ~70,000 samples (Keller and Schoene, 2012), and validation of a global lithological map using ~ 290,000 data points (Hartmann and Moosdorf, 2012). Applications of the Portal in the literature have been found in many fields beyond rock and sediment chemistry; citations exist in the field of ecology (Bataille et al., 2012), water chemistry (Olson and Hawkins, 2012), and climate change (Box et al., 2011).

Tools were developed to support investigators both as users and contributors of data. Logged-in users can save and rerun queries, which is particularly useful as new data are added to the database. A database of shared queries is stored in the Portal Glossary that can be reused by others. Samples may be selected by chemical composition in several ways, either by a numerical value (composition range of specific parameters), a classification assigned by the author, or a classification derived from the chemical composition. The difference between author-assigned and chemically-derived classifications is striking; one popular demonstration is to show that author-assigned classifications for rocks (e.g. "dacite") span a wide range of actual classifications when their chemical composition is plotted on a TAS (total alkali versus silica) diagram (i.e. the author-given classifications are incorrect). This makes the chemically-derived classification much more accurate if the composition is very important. **Figure 2** shows an example where the author assigned classification of dacite varies strongly from the chemically-derived classification. Once samples are selected, users can access built-in visualization tools including dynamic maps and Google Earth integration. TAS and Harker plots can be generated and edited quickly. Authors with data in the portal can use the citation function to see how often data from their own publications have been downloaded.

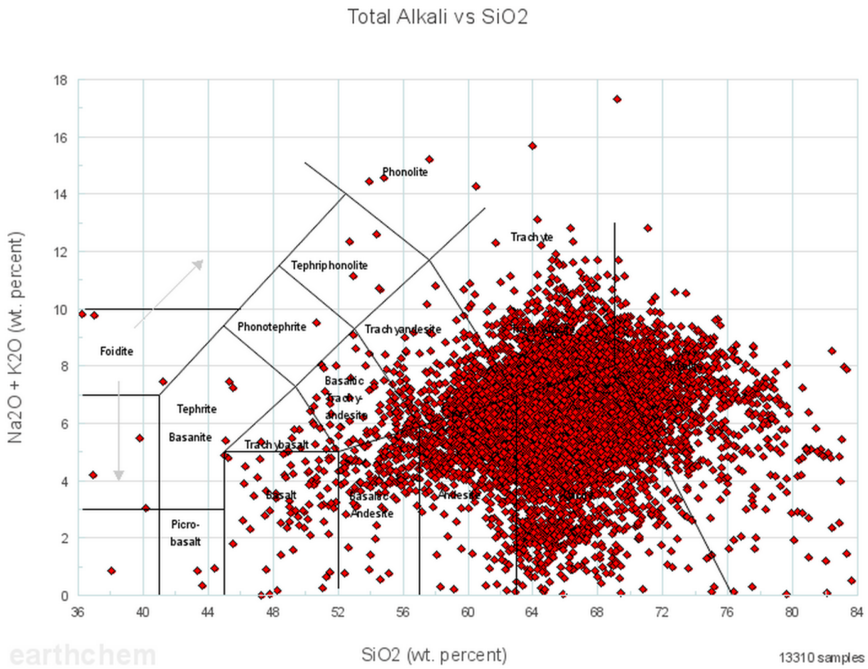


Figure 2: **Figure 2a.** Author-designated classification of the rock classification dacite (samples called "dacite" in the published work).

The Portal data results are linked with the LEPR (Library of Experimental Phase Relations) database (Hirschmann et al., 2008), allowing a comparison of field and laboratory derived rocks with matching or similar compositions. The portal can easily display sample results as grouped by a publication DOI, as used in direct linking from Science Direct online articles (Elsevier press release, accessed 25 January 2014).

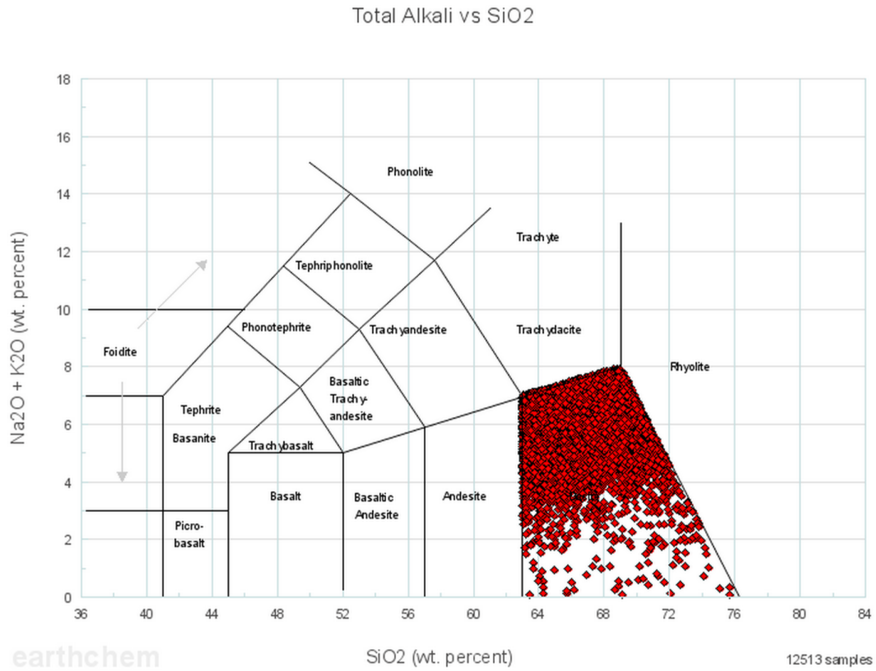


Figure 3: **Figure 2b.** TAS (Total Alkali versus Silica) diagram showing samples that match the chemically-determined definition of dacite (contrast with Figure 2a, what authors assign as "dacite" in their writing).

# 7 Benefits for Users and New Science Discoveries

The EarthChem Library, the topical synthesis databases, and the EarthChem Portal together form a suite of tools to support geochemists throughout their research workflow – from data discovery to analysis to publication. To be successful, data systems must preserve useful metadata, respond to community driven development of features, and have a user-friendly interface for accessing the data. An essential part of the success of IEDA EarthChem's data systems is the response to feedback and guidance provided by the user community (Lehnert and Hsu, 2013). The standards for data reporting were developed and refined by targeted user groups to ensure that the output result is useful for users. Frequent contact with the user community through workshops, international conference meeting booths, informational emails, social media, webinars, and a User Committee composed of faculty and specialist users ensure that feedback is continually gathered. Particularly in the long-tail field of sample-based analytical geochemistry, review by and iterative discussion with a domain scientist is important for safeguarding standards for archiving the data. Users report inaccuracies in the database, which are corrected and noted for others. Without community interaction, the EarthChem data systems would be susceptible to errors in both content and development of functionalities.

The EarthChem systems unquestionably provide increased access to scientific data and tools for the geochemistry community and to the broader global community of scientists, students, educators, and industry users (Lehnert, 2007, e.g.). Anyone with Internet access can download the data or use the tools. Over the past year, feedback from the PetDB system indicates that users who download datasets are ˜85% for research, ˜13% for education, and ˜2% for other reasons. In addition to bringing some data out behind publication pay-walls, EarthChem provides investigator support for publishing data that may not have otherwise been made publically available. This type of support is essential for scientists working in the long-tail, for whom there is little guidance.

Data included in EarthChem has an added value to the original published datasets in a number of ways, including additional sample metadata, sample name alignment, and sample classification by chemical composition. When chemical values that are relevant to a specific topical synthesis database are found in the literature without adequate documentation for inclusion in the integrated dataset, EarthChem data managers will obtain missing metadata from the author. Required information includes georeferenced sample locations, sampling technique, and laboratory preparation and equipment. Sample names, notoriously ambiguous for samples collected during research cruises (e.g., D1-01 for Dredge 1, Sample 1), are carefully aligned across all publications citing the same sample, and the topical synthesis databases provide a list of the sample aliases. This is particularly useful for dredge and core samples that realiquoted and subsampled several times, and may be given different names in publications by different authors (e.g., D1-01 and D1-1).

Data access and discoverability are enhanced by providing links to systems outside of EarthChem. These links include published manuscripts (digital and non-digital), the sample registry SESAR (System for Earth Sample Registration, www.geosamples.org), and the cruise catalog and datasets in the Marine Geoscience Data System (MGDS, www.marine-geo.org). Working with publishers, EarthChem provides services to access all available EarthChem data related to a paper, which can be accessed from the digital copy. SESAR provides detailed metadata for samples, based on a unique identifier called the IGSN (International Geo Sample Number). PetDB, SedDB, the Portal, the ECL and Geochron are linked by using that same IGSN identifier scheme.



Proper attribution for scientific results is a common concern of investigators whose data are integrated into larger databases. EarthChem promotes proper attribution to investigators for their scientific data in two ways. First, the ECL provides persistent, citable DOIs for each submitted dataset. This ensures that the individual author can be credited for their contributions to the larger database. A second way is the Portal tool that allows investigators to track the number of times data from a particular author has been downloaded into an integrated dataset. A third way is that all outputs come with a list of references at the bottom of the Excel sheet. This way, authors are automatically provided with the references for all of the data they just downloaded and can use that to populate their works cited list. The benefit of this is improved citation and reporting function for funding agencies, promotion cases, or other situations that require quantitative measure of scientific output. EarthChem also encourages the citation of the original sources of data in addition to PetDB itself.

The challenges for electronic data publication in geochemistry outlined by Staudigel et al. (2003) have not been completely eliminated, but new users and uses of electronic geochemical data have grown tremendously in the past decade, with new databases, web applications, and scientific results. Data systems that were created in the past must incorporate new technologies and standards to remain relevant. Examples include unique identifiers that have been developed for publications (DOI), samples (IGSN), and people (ResearcherID, ORCiD), which are being implemented into EarthChem systems.

We have shown how EarthChem has addressed the challenges related to long-tail scientific data management and has contributed to the scientific output both in its home geochemistry field but also other seemingly unrelated related disciplines. Disciplinary sciences are interlinked when solving grand challenges in science, especially Earth Science, and thus so are the challenges for efficiently producing and reusing scientific data. IEDA EarthChem and other disciplinary data systems will continue to grow and integrate to provide their user community with more powerful tools for scientific analysis and discovery.

# 8 Acknowledgments

This work was supported by NSF Grant 0950477. The EarthChem system was developed by a long list of developers and data managers over the past 15 years, whose cumulative contributions made this manuscript possible.

# References

Clement P. Bataille, Jason Laffoon, and Gabriel J. Bowen. Mapping multiple source effects on the strontium isotopic signatures of ecosystems from the circum-Caribbean region. *Ecosphere*, 3(12):art118, dec 2012. doi: 10.1890/es12-00155.1. URL http://dx.doi.org/10.1890/ES12-00155.1.

J. F. Bowring, N. M. McLean, and S. A. Bowring. Engineering cyber infrastructure for U-Pb geochronology: Tripoli and U-Pb_Redux. *Geochemistry Geophysics, Geosystems*, 12(6):n/a–n/a, jun 2011. doi: 10.1029/2010gc003479. URL http://dx.doi.org/10.1029/2010GC003479.

M.R. Box, M.D. Krom, R.A. Cliff, M. Bar-Matthews, A. Almogi-Labin, A. Ayalon, and M. Paterne. Response of the Nile and its catchment to millennial-scale climatic change since the LGM from Sr isotopes and major elements of East Mediterranean sediments. *Quaternary Science Reviews*, 30(3-4):431–442, feb 2011. doi: 10.1016/j.quascirev.2010.12.005. URL http://dx.doi.org/10.1016/j.quascirev.2010.12.005.

Conny Class and Kerstin Lehnert. PetDB Expert MORB (Mid-Ocean Ridge Basalt) Compilation, 2012. URL http://dx.doi.org/10.1594/IEDA/100060.

Clinton P. Conrad, Todd A. Bianco, Eugene I. Smith, and Paul Wessel. Patterns of intraplate volcanism controlled by asthenospheric shear. *Nature Geosci*, 4(5):317–321, mar 2011. doi: 10.1038/ngeo1111. URL http://dx.doi.org/10.1038/ngeo1111.

John W. Delano. Electron microprobe analyses of Apollo 15 green and yellow glasses from Lunar sample 15426,72, 2013a. URL http://dx.doi.org/10.1594/IEDA/100410.

John W. Delano. Electron microprobe analyses of Apollo 17 volcanic orange glasses from Lunar sample 74220,128, 2013b. URL http://dx.doi.org/10.1594/IEDA/100411.

Allison Gale, Colleen A. Dalton, Charles H. Langmuir, Yongjun Su, and Jean-Guy Schilling. The mean composition of ocean ridge basalts. *Geochemistry Geophysics, Geosystems*, 14(3):489–518, mar 2013. doi: 10.1029/2012gc004334. URL http://dx.doi.org/10.1029/2012GC004334.

James B. Gill. Fiji Tonalites, 2014. URL http://dx.doi.org/10.1594/IEDA/100418.

Steven L. Goldstein, Gad Soffer, Charles H. Langmuir, Kerstin A. Lehnert, David W. Graham, and Peter J. Michael. Origin of a 'Southern Hemisphere' geochemical signature in the Arctic upper mantle. *Nature*, 453 (7191):89–93, may 2008. doi: 10.1038/nature06919. URL http://dx.doi.org/10.1038/nature06919.

Steven L. Goldstein, Albrecht W. Hofmann, and Kerstin A. Lehnert. Requirements for the Publication of Geochemical Data, 2014. URL http://dx.doi.org/10.1594/IEDA/100426.

Jens Hartmann and Nils Moosdorf. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry Geophysics, Geosystems*, 13(12):n/a–n/a, dec 2012. doi: 10.1029/2012gc004370. URL http://dx.doi.org/10.1029/2012GC004370.

P. Bryan Heidorn. Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends*, 57(2): 280–299, 2008. doi: 10.1353/lib.0.0036. URL http://dx.doi.org/10.1353/lib.0.0036.

A. Hein and V. Kilikoglou. ceraDAT-PROTOTYPE OF A WEB-BASED RELATIONAL DATABASE FOR ARCHAEOLOGICAL CERAMICS. *Archaeometry*, 54(2):230–243, jul 2011. doi: 10.1111/j.1475-4754.2011.00618.x. URL http://dx.doi.org/10.1111/j.1475-4754.2011.00618.x.

M. M. Hirschmann, M. S. Ghiorso, F. A. Davis, S. M. Gordon, S. Mukherjee, T. L. Grove, M. Krawczynski, E. Medard, and C. B. Till. Library of Experimental Phase Relations (LEPR): A database and Web portal for experimental magmatic phase equilibria data. *Geochemistry Geophysics, Geosystems*, 9(3):n/a–n/a, mar 2008. doi: 10.1029/2007gc001894. URL http://dx.doi.org/10.1029/2007GC001894.

C. Brenhin Keller and Blair Schoene. Statistical geochemistry reveals disruption in secular lithospheric evolution about 2.5 Gyr ago. *Nature*, 485(7399):490–493, may 2012. doi: 10.1038/nature11024. URL http://dx.doi.org/10.1038/nature11024.

Jun-Ichi Kimura. Arc Basalt Simulator (ABS) version 3, 2012. URL http://dx.doi.org/10.1594/IEDA/100045.

K Lehnert. The PetDB data collection: Impact on science. In *2007 GSA Denver Annual Meeting*, 2007. URL https://gsa.confex.com/gsa/2007AM/finalprogram/abstract_131184.htm.

K. Lehnert and L. Hsu. Data Publication: The Role of Community-Based, Disciplinary Repositories. In *EGU General Assembly Conference Abstracts*, volume 15 of *EGU General Assembly Conference Abstracts*, page 13218, apr 2013.

K. Lehnert, Y. Su, C. H. Langmuir, B. Sarbas, and U. Nohl. A global geochemical database structure for rocks. *Geochemistry Geophysics, Geosystems*, 1(5):n/a–n/a, may 2000. doi: 10.1029/1999gc000026. URL http://dx.doi.org/10.1029/1999GC000026.

J. M. Mottl. Explanatory Notes and Master Chemical Item Spreadsheet for the VentDB Data Collections housed in the Geochemical Resource Library, 2012. URL http://dx.doi.org/10.1594/IEDA/100207.

John R. Olson and Charles P. Hawkins. Predicting natural base-flow stream water chemistry in the western United States. *Water Resour. Res.*, 48(2):n/a–n/a, feb 2012. doi: 10.1029/2011wr011088. URL http://dx.doi.org/10.1029/2011WR011088.

Jason N. Rauch. Global distributions of Fe Al, Cu, and Zn contained in Earth's derma layers. *Journal of Geochemical Exploration*, 110(2):193–201, aug 2011. doi: 10.1016/j.gexplo.2011.05.008. URL http://dx.doi.org/10.1016/j.gexplo.2011.05.008.

Carla Rodrigues, Cristina Máguas, and Thomas Prohaska. Strontium and oxygen isotope fingerprinting of green coffee beans and its potential to proof authenticity of coffee. *European Food Research and Technology*, 232(2):361–373, sep 2010. doi: 10.1007/s00217-010-1362-z. URL http://dx.doi.org/10.1007/s00217-010-1362-z.

Vincent J. M. Salters and Andreas Stracke. Composition of the depleted mantle. *Geochemistry Geophysics, Geosystems*, 5(5):n/a–n/a, may 2004. doi: 10.1029/2003gc000597. URL http://dx.doi.org/10.1029/2003GC000597.

B. Sarbas and U. Nohl. The GEOROC database - A decade of online geochemistry. In *Geochimica et Cosmochimica Acta Volume 73, Issue 13, Supplement*, 2009. doi: 10.1016/j.gca.2009.05.015. URL http://dx.doi.org/10.1016/j.gca.2009.05.015.

Joan Starr and Angela Gastl. isCitedBy: A Metadata Scheme for DataCite. *D-Lib Magazine*, 17(1/2), jan 2011. doi: 10.1045/january2011-starr. URL http://dx.doi.org/10.1045/january2011-starr.

Hubert Staudigel, John Helly, Anthony A. P. Koppers, Henry F. Shaw, William F. McDonough, Albrecht W. Hofmann, Charles H. Langmuir, Kerstin Lehnert, Baerbel Sarbas, Louis A. Derry, and Alan Zindler. Electronic data publication in geochemistry. *Geochemistry Geophysics, Geosystems*, 4(3):n/a–n/a, mar 2003. doi: 10.1029/2002gc000314. URL http://dx.doi.org/10.1029/2002GC000314.

J. D. Walker, T.D. Bowers, R.A. Black, A.F. Glazner, G. Lang Farmer, and R. W. Carlson. A geochemical database for western North American volcanic and intrusive rocks (NAVDAT). *GSA Special Paper 397 Geoinformatics: Data to Knowledge*, 397:61–71, 2006. doi: 10.1130/2006.2397(05). URL http://dx.doi.org/10.1130/2006.2397(05).

J. D. Walker, J. A. Ash, J. F. Bowring, S. A. Bowring, A.L. Deino, R. Kislitsyn, A. A. Koppers, N. M. Mclean, and K. A. Lehnert. Collaboration of EarthChem and EARTHTIME to Develop a Geochronology and Thermochronology Database. In *Abstract IN23C-1463 presented at 2011 Fall Meeting, AGU, San*

*Francisco, Calif., 5-9 Dec*, 2011. URL http://http://abstractsearch.agu.org/meetings/2011/FM/sections/IN/sessions/IN23C/abstracts/IN23C-1463.html.