



## INNOVATING AND NETWORKING GLOBAL GEOCHEMICAL DATA RESOURCES THROUGH ONEGEOCHEMISTRY

Alexander M. Prent<sup>1,2,3</sup>, Dominik C. Hezel<sup>4</sup>, Marthe Klöcking<sup>5</sup>, Lesley Wyborn<sup>6</sup>, Rebecca Farrington<sup>1</sup>, Kirsten Elger<sup>7</sup>, Lucia Profeta<sup>8</sup>, Angus L. Nixon<sup>9</sup>, and Kerstin Lehnert<sup>8</sup>

DOI: 10.2138/gselements.19.3.136

A vision of the not-too-distant future: Imagine yourself as a researcher who is making plans for fieldwork in the Andes Mountains. You are behind your computer and load a three-dimensional visualisation of the Earth and its geology to investigate the research already done in the southern Patagonian regions. You zoom in to the mountain range of interest and select various data layers to show samples, their chemical and isotopic compositions, as well as rock ages for further reference. One area shows particularly young ages and a single click brings up an image showing the thermal and chemical evolution of the rocks, bringing to life the events experienced by that part of the Earth. Each sample and data point have all the necessary information about uncertainties in the data, and the associated description of the analytical methods enables you to verify and trust the quality of data associated with (anomalous) points. Looking back at the visualisation, you change a few parameters and, in (close to) real-time, the modelled Earth changes to display the outcomes of the selected model scenario. The model recipe and selected data are exported by another click to a standard formatted file. These data are directly usable in your chosen application and offline device. Adding your own recently collected data to the model via drag and drop makes them show up in bright colours, contrasting with the prior known data and putting them in direct context. The visualised additional data alter the prior geological understanding and confirm your suspicions regarding what information is missing. Together with a customisable visualisation of model uncertainty, this information helps you to plan the collection of new samples. You are excited to go into the field to collect and subsequently analyse samples that you know will complement existing research in the area and vastly improve the current geological understanding of this region. With this preparation in hand, you are now able to apply for the necessary funding, proposing an efficient plan that minimises cost at a high likelihood of success.

### TOWARDS ACHIEVING A SHARED VISION

This scenario envisages the streamlined handling of digital geochemical data and interoperability from various sources in the near future. This is in stark contrast to the current situation where such models must be painstakingly built by individuals based on datasets scraped together from several decades of publications, having a wide range of formats, sources, and quality (Chamberlain et al. 2021). Discipline-specific databases such as GEOROC, PetDB, Astromat, or AusGeochem already help considerably to find and synthesise published geochemical data; however, to achieve the above scenario, additions to legacy and future generated data are required to make them fully interoperable, such as uncertainties and metadata to enable quality assessment. Tools

and simple workflows should be developed to add or combine missing data from, e.g., unpublished regional or institutional data stores that may be hidden there or have never been published at all.

Our envisioned scenario requires using newly obtained data directly in the same context as previously published data, enabling linkages of multiple components of geochemistry (which here always includes cosmochemistry). To establish these connections, it is essential for data collections to follow standardised procedures for reporting geochemical information, using agreed-upon terminology. For this, three major challenges need to be addressed: (i) *Standardisation of data collections and reporting*: this will enable researchers to compare and combine data from different sources; (ii) *Making data accessible online*: many data, particularly old datasets, are difficult to access and may not yet be digitally available; and (iii) *Enabling replication, verification of data quality and uncertainty*: this will allow decision making on how old and new data can be combined as a basis for new interpretations. These challenges are also known as making data Findable, Accessible, Interoperable, and Reusable for both machines and humans, in other words: FAIR (Wilkinson et al. 2016). While this has been partly achieved in related domains, geochemistry is just getting started. Together with the geochemical community, OneGeochemistry now needs to harness existing initiatives and better coordinate the various community groups around the world working towards making geochemical data FAIR.

### ONEGEOCHEMISTRY: ENABLING A SHARED VISION

The near-future vision and FAIR geochemical data can only be achieved in a global effort, similar to the “Editors Roundtable” in 2007 that brought together publishers and data repository providers to discuss implementation of consistent data publication practices, which resulted in a policy statement in 2009 (Goldstein et al. 2014).

The OneGeochemistry initiative seeks to create a global network of geochemical data infrastructure providers that facilitates and promotes discovery of, and access to, geochemical data through coordination and collaboration. It will serve as central hub for the collection and distribution of best practices for data reporting standards and their metadata, global definitions of geochemical vocabularies and ontologies, and will encourage the international community to adopt these. OneGeochemistry currently connects representatives from EarthChem, GEOROC, Astromat, MetBase, the Multi-scale Laboratories of the European Plate Observing System (EPOS), GFZ Data Services, NFDI4Earth, and the AuScope Geochemistry Network and has been appointed as “OneGeochemistry CODATA Working Group” in late 2022.

The OneGeochemistry initiative will promote, assist or coordinate global efforts in geochemical data standardisation, facilitate communication between international groups working towards this same goal, to lessen the duplication of efforts. This is to enable development of a distributed FAIR geochemical data framework that interoperates through standard web services following the OneGeology example.

### ARTICULATING GEOCHEMICAL DATA DIVERSITY

The field of analytical geochemistry is as fascinating as it is diverse. Geochemical research questions reach as far back as the beginning of our universe with the formation of the first elements, followed by the chemical and isotopic population of the periodic table and chart of nuclides during the evolution of galaxies and stars. The discipline studies planetary differentiation and the formation of planetary building blocks, the Earth–Moon system, plate tectonics, deep Earth, surface, oceanic and atmospheric processes. Geochemistry is important for addressing questions around the future of our ecosystem and natural resources across a plethora of research fields. Similarly, the number of geochemical methods, analytical instrumentation, and laboratories is

1 AuScope Ltd, Melbourne, Australia

2 John de Laeter Centre, Curtin University, Bentley, Australia

3 Faculty of Geosciences, Utrecht University, Utrecht, Netherlands

4 Institut für Geowissenschaften, Goethe-Universität Frankfurt, Frankfurt, Germany

5 Geoscience Centre, University of Göttingen, Göttingen, Germany

6 Research School of Earth Sciences, The Australian National University, Acton, Australia

7 GFZ German Research Centre for Geosciences, Potsdam, Germany

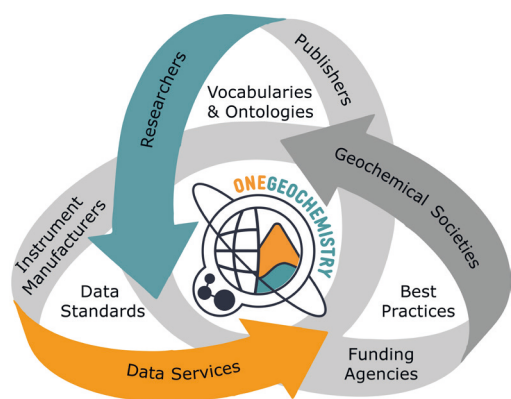
8 Lamont-Doherty Earth Observatory, Columbia University, Palisades, USA

9 University of Adelaide, Adelaide, Australia

diverse and has impressively multiplied during the past few decades, and continues to dynamically evolve. One of the largest challenges for OneGeochemistry is that most geochemical data are small and highly variable, collected by individual researchers or small teams across many disciplines. These so-called ‘long-tail data’ (Heidorn 2008), are especially challenging to harmonise and standardise, but represent a significant number of research results. With such diversity and cross-domain overlap, OneGeochemistry is focusing first and foremost on raising community awareness of the benefit of standardised reporting of chemical and isotopic data and their description. These are important steps for their full integration in the digital world.

## STANDARDISING GEOCHEMICAL DATA

Technological, sociological, and organisational leaps have been underpinned by shared community rules and understanding: standard containers revolutionised shipping, the internet protocol our modern world, and SI units the basic exchange of scientific knowledge. Standardising geochemical data reporting has a similar potential to revolutionise how the community accesses and interacts with geochemical data. Much can be learned from other communities such as chemistry, crystallography, seismology and oceanography that have established similar organisations to OneGeochemistry (Stall et al. 2020; Klöcking et al. 2023). As a community, we should strive to make geochemical data easily accessible to all and accelerate interdisciplinary data integration. OneGeochemistry has been established to connect the different stakeholders working with geochemical data to assist in driving this effort (Fig. 1).



**FIGURE 1** The OneGeochemistry initiative as a platform that aggregates and disseminates best practice recommendations from/to the diverse groups working with geochemical data.

Best practices will allow users to re-use datasets and validate their quality. Numerous suggestions on how to best report geochemical data have already been published, but are not yet widely adopted (e.g., Schaen et al. 2021; Flowers et al. 2022; Wallace et al. 2022). Adoption of reporting best practices is about to change with the increasing requirement from journals to publish data in domain-specific repositories provided by, e.g., EarthChem, DIGIS hosted by GFZ Data Services, or Astromat. This improves access to these data, facilitates regional and global data analysis and interpretation (e.g., Keller and Schoene, 2012), and makes the data FAIR, as these repositories require and facilitate the use of best practices via templates and discipline specific data models e.g., EarthChem XML schema or EarthChem Data Submission Templates. OneGeochemistry aims to structure, publicise, and promulgate these existing efforts in a community-built, global system of FAIR geochemical data reporting. The aim is to agree on best practices and provide data and metadata in a machine-actionable form following the

10 simple rules for FAIR vocabularies (Cox et al. 2021). The collective, community-agreed data reporting best practices can then be used and implemented by researchers, repositories, publishers, service providers, laboratories, companies, etc.

## ONEGEOCHEMISTRY AS AN INTERNATIONAL ORGANISATION OF COMMUNITIES AND AGREEMENTS

Predominantly a voluntary effort, the OneGeochemistry initiative is currently run through a working group of the International Science Council's Committee on DATA (CODATA), providing the initiative an initial operational structure. Strong community engagement, inter-action, and feedback are essential for the success of the OneGeochemistry initiative for which it has already been endorsed by a number of international geo- and cosmochemical societies and unions: the Geochemical Society, the European Association of Geochemistry, the Association of Applied Geochemists, the International Association of Geochemistry, the Meteoritical Society, and the IUGS commission on Global Geochemical Baselines.

The current OneGeochemistry initiative members—largely similar to the authors of this article—are active in the community through the organisation of sessions and town hall meetings at international conferences such as the EGU General Assembly, Goldschmidt, and AGU Fall Meeting, where we encourage readers to engage with this initiative. OneGeochemistry is still very much developing and we encourage the community to get involved in deciding on its governance, formal and legal structure, help make decisions, and participate in the process of defining community-led guidelines for geochemical data handling and preservation. As part of this, you can make yourself electable for the board, and join or lead scientific or technical expert committees to advise on and help coordinate development of standard data reporting methods, method-specific vocabularies, as well as metadata profiles (e.g., Klöcking et al. 2023).

The OneGeochemistry initiative invites you, other researchers, data groups, and initiatives to help realise this vision. For more information and if you would like to join the initiative or highlight any existing best practice recommendations not yet included, please visit the OneGeochemistry website at [www.onegeochemistry.org](http://www.onegeochemistry.org)

## REFERENCES

- Chamberlain KJ, Lehnert KA, McIntosh IM, Morgan DJ, Wörner G (2021) Time to change the data culture in geochemistry. *Nature Reviews Earth & Environment* 2: 737–739, doi: 10.1038/s43017-021-00237-w
- Cox SJD, Gonzalez-Beltran AN, Magagna B, Marinescu, M-C (2021) Ten simple rules for making a vocabulary FAIR. *PLoS Computational Biology* 17: e1009041, doi: 10.1371/journal.pcbi.1009041
- Flowers RM and 9 coauthors (2022) (U-Th)/He chronology: part 1. Data, uncertainty, and reporting. *GSA Bulletin* 135: 104–136, doi: 10.1130/B36266.1
- Goldstein SL, Hofmann AW, Lehnert KA (2014) Requirements for the publication of geochemical data. *Interdisciplinary Earth Data Alliance (IEDA)*, doi: 10.1594/IEDA/100426
- Heidorn PB (2008) Shedding light on the dark data in the long tail of science. *Library Trends* 57: 280–299, doi: 10.1353/lib.0.0036
- Keller CB, Schoene B (2012) Statistical geochemistry reveals disruption in secular lithospheric evolution about 2.5 Gyr ago. *Nature* 485: 490–493, doi: 10.1038/nature11024
- Klöcking M and 34 coauthors (2023) Community recommendations for geochemical data, services and analytical capabilities in the 21st century. *Geochimica et Cosmochimica Acta* 351: 192–205, doi: 10.1016/j.gca.2023.04.024
- Schaen AJ and 40 coauthors (2021) Interpreting and reporting 40Ar/39Ar geochronologic data. *GSA Bulletin* 133: 461–487, doi: 10.1130/B35560.1
- Stall S, McEwen L, Wyborn L, Hoebelheirich N, Bruno I (2020) Growing the FAIR community at the intersection of the geosciences and pure and applied chemistry. *Data Intelligence* 2: 139–150, doi: 10.1162/dint\_a\_00036
- Wallace KL and 14 coauthors (2022) Community established best practice recommendations for tephra studies—from collection through analysis. *Scientific Data* 9: 447, doi: 10.1038/s41597-022-01515-y
- Wilkinson MD and 52 coauthors (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3: 160018, doi: 10.1038/sdata.2016.18