






Time to change the data culture in geochemistry

Katy J. Chamberlain¹  , Kerstin A. Lehnert², Iona M. McIntosh³, Dan J. Morgan⁴  and Gerhard Wörner⁵

Geochemical data are vital for understanding Earth's past, present and future. However, currently only a fraction of geochemical data are findable, accessible, interoperable and reusable, limiting their use in the broadest range of scientific studies. There is an urgent need for international coordination of geochemical data and methods to unlock their full research potential.

Geochemical data are essential for understanding planetary and environmental processes, including the origin and evolution of Earth, the Solar System, and beyond. Increasingly complex analytical equipment has dramatically expanded the production and variety of geochemical data. However, only a fraction of these data are FAIR (findable, accessible, interoperable and reusable)¹, owing to a lack of international community-wide standards and best practices. The lack of a shared data culture seriously hinders data quality assessments, incorporation of geochemical results into multidisciplinary studies, and the (re)use of geochemical data in modelling and big data studies. Output geochemical data need to be corrected to defined standards, and detection limits are variable between laboratories and run times, introducing restrictions in comparisons between laboratories. Potential future discoveries are, therefore, hindered, necessitating an urgent community need to improve access to, and reusability of, geochemical data to unlock their full potential.

Data value and longevity

Every geochemical data point is valuable, each representing an investment of time, money and resources. Therefore, careful data description and curation are required to optimize return on investments, ensure long-term utility, and support transparent and reproducible scientific research. Data preservation is especially critical where samples are ephemeral and/or impossible to obtain again in the future, such as gas or water samples that can't be preserved.

Geochemical data are often fragmented and diverse in their character, for example with respect to material analyzed (rocks, minerals, soils, water), bulk or in situ measurements, concentrations and (isotope) ratios, compositional maps and more. These geochemical data are more than the final published values. They require comprehensive metadata² (TABLE 1) to maximize their post-publication usefulness. At a minimum, metadata should include information about sample provenance

and preparation, the laboratory, analysis date, instrument model, configuration and calibration, data quality, accuracy, reproducibility, and data reduction procedures (for example, normalization and fractionation corrections). In doing so, geochemical data would be reproducible, can be compared between labs, and recalculated as appropriate as analytical techniques develop and standards are refined.

The value of geochemical data is further augmented if data and metadata are consistently structured, so they are machine-readable and fully interoperable for computational and statistical approaches. Interoperability allows data in curated databases such as GEOROC and PetDB and others to be integrated into large global datasets that cover the entire Earth history. GEOROC and PetDB have used the EarthChemXML to automatically encode and integrate their data and metadata, but most databases today need to be manually mapped to the EarthChemXML schemes.

The problems

Data quality assessment is undermined by the non-standardized approach to geochemical data publication and methods. For example, sharing reference material and rejected analyses is not common practice across the community. However, it is impossible to evaluate and test reproducibility, uncertainty estimation, normalization procedures and similarities between different research groups without publishing the complete data and metadata. Ultimately, this practice compromises the use of geochemical data. In particular, it impedes integration with other multidisciplinary datasets, as non-geochemists are more likely to be unfamiliar with uncertainties and different methods, affecting data comparison and interpretation.

Databases such as GEOROC, PetDB and NAVDAT have improved access to rock geochemical data and interoperability among themselves. However, maintaining common metadata standards, updating vocabularies, and providing harmonized data evaluation tools remain

¹Environmental Sustainability Research Centre, University of Derby, Derby, UK.

²Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA.

³Solid Earth Geochemistry Research Group, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Japan.

⁴School of Earth and Environment, University of Leeds, Leeds, UK.

⁵Geowissenschaftliches Zentrum, Georg-August-Universität Göttingen, Göttingen, Germany.

[✉]e-mail: k.chamberlain@derby.ac.uk

<https://doi.org/10.1038/s43017-021-00237-w>

Table 1 | Minimum required geochemical metadata

Meta-data descriptor	Required meta-data				
<i>Geographic information</i>					
All samples	Sample number	Sample lat.–long.	Lat.–long. accuracy	Elevation	Elevation accuracy
<i>Sample type</i>					
All samples	Sampling time	Time zone	Collector or scientist	Archiver or curator	Sampling method
Rocks	Chemical characterization	Sediment grain size	-	-	-
Water and gas	Sampling container	Sampling treatment	Fluid type	Fluid classification	-
Fossils	Fossil group	Fossil species	Fossil preservation	-	-
Minerals	Mineral name	Occurrence type	-	-	-
<i>Laboratory information</i>					
Interoperability	Analyst and laboratory	Analytical code	Instrument used	Measured parameter	Technique description
<i>Analysis information</i>					
Reproducibility	Reference sample name	Uncertainty value	Uncertainty unit	-	-
Normalization	Reference sample name	Normalization value	Normalization unit	-	-
Fractionation	Reference sample name	Fractionation parameter	Fractionation value	Fractionation unit	-
Detection Limit	Detection limit	Detection unit	-	-	-
Blank	Blank type	Blank value	Blank unit	-	-

Minimum required metadata^a that should be submitted alongside all geochemical datasets, irrespective of whether samples are rock, water or gas.

a challenge. Many other data curation systems and databases exist at national geological surveys and other institutions (for example, GeoKniga, USGS National Geochemical Database, National Geochemical Survey of Australia) but lack interoperability. Even large international repositories such as the [EarthChem Library](#) have to follow their own standards, because a globally endorsed and governed standard of data does not yet exist. Moreover, data publishing and metadata requirements are variably enforced between different funders and journals, and are not yet embedded into community culture³. It commonly falls to individual researchers to archive their new data in repositories.

FAIR solutions

Fortunately, there is growing recognition that a FAIR, robust approach to handling geochemical data is sorely needed. Since 2014 (REF.⁴), minimum metadata requirements for publishing geochemical data have been defined (TABLE 1). Funders and publishers have started to implement policies that require researchers and authors to deposit data in public repositories, making data open, citable and available⁵. Further progress should be made to ensure that such policies and requirements are widely enforced by the various publishers and funding bodies, and that data are deposited in FAIR repositories such as the [EarthChem Library](#).

New projects and initiatives are emerging at national and global levels to advance, integrate and share geochemical data standards and promote best practices. For example, Australia's [AuScope Geochemistry Network](#), the [European Plate Observing System \(EPOS\)](#), and laboratories participating in NASA's asteroid sample-return mission ([OSIRIS-REx](#)), are all promoting the sustainable and universal use and re-use of geochemical data.

In addition, the Digital Geochemistry Infrastructure project ([DIGIS initiative](#)) aims to modernize the widely used GEOROC database, (which hosts 1.8 million published analyses of igneous rocks) incorporating FAIR principles and improving metadata structure and user accessibility. Streamlined connections with other geochemical databases, such as [EarthChem](#) and [AstroMat](#), will further improve data interoperability.

In response to open-access policies and science demands, ever-increasing numbers of geochemical databases are emerging at national, programmatic, and subdomain levels. As such, there is an urgent need for the geochemistry community to coordinate these efforts and collaboratively define the required data and metadata standards, and best practices, that will enable worldwide interoperability and re-use of geochemical data and methods. The international [OneGeochemistry](#) initiative is responding to this need, as it aims to create a global geochemical data network that facilitates and promotes discovery of, and access to, geochemical data by unifying diverse data strategies developed across the globe^{6,7}.

Changing data culture in geochemistry

Changing data culture requires broad adoption of new rules, workflows, and values. Open and comprehensive data sharing needs to be mandated and enforced by funders and publishers. All data should be shared via repositories that follow the FAIR principles, and every dataset should be assigned a globally unique identifier (such as a DOI), where every physical sample should have an IGSN (International Geo Sample Number) to link the original sample material unambiguously to data and publications. Repositories should ensure compliance with domain-relevant community standards that describe the analytical procedure, sample provenance,

and data quality. Adaptation to a FAIR data culture will require the potentially uncomfortable recognition that fully illuminating data and methods could identify mistakes or oversights. Researchers must establish the attitude that this identification is a positive outcome that will ultimately advance scientific progress.

Researchers will need guidance and resources to adopt and adapt to new practices. Innovative digital tools are needed to support automated data and metadata capture in the laboratory (for example using electronic laboratory notebooks, ELNs) and to facilitate data cleaning during database incorporation. Manufacturers of analytical instrumentation need to provide open interfaces to data and metadata generated by their software to be linked to ELNs. Institutions and funders need to allocate funds to support the effort of proper data management.

A fundamental requirement for geochemical data culture change is that researchers get credit for sharing their data. Universities and research institutes must recognize the value of open data curation and help grow awareness and appreciation of the FAIR data principles. Teaching of data handling and open science principles should be embedded alongside the earliest introduction to geochemical data during undergraduate study. Meanwhile, data and open science sessions at geoscience conferences must be promoted on center stage instead of being scheduled in competition with discipline-specific sessions.

Lastly, a sustained global geochemical data infrastructure, like the OneGeochemistry initiative, is urgently needed to facilitate the use and maximize the impact of data to ensure scientific progress. International coordination is essential to ensure geochemical database interoperability and consistent analytical quality and metadata standards. Only then will the full potential of multidisciplinary progress be unlocked to address

fundamental research questions regarding Earth's past and uncertain future.

1. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
2. Staudigel, H. et al. Electronic data publication in geochemistry. *Geochem. Geophys. Geosyst.* **4**, 8004 (2003).
3. Goldstein, S. L., Lehnert, K. & Hofmann, A. W. Managing open and FAIR data in geochemistry: Where are we a decade after the Editors Roundtable? In *EGU General Assembly 2021* <https://doi.org/10.5194/egusphere-egu21-16420> (EGU, 2021)
4. Goldstein, S. L., Hofmann, A. W. & Lehnert, K. A. *Requirements for the Publication of Geochemical Data Version 1.0* <https://doi.org/10.1594/IEDA/100426> (IEDA, 2014).
5. Stall, S. et al. Make scientific data FAIR. *Nature* **570**, 27–29 (2019).
6. Lehnert, K. & Wyborn, L. A. OneGeochemistry: Toward a global network of geochemical data. In *AGU Fall Meeting 2019* (AGU, 2019).
7. Wyborn, L. & Lehnert, K. OneGeochemistry: Creating a global FAIR-way to access and share geochemical data. In *Goldschmidt Conference 2020* <https://doi.org/10.46427/gold2020.2910> (Goldschmidt, 2020).

Acknowledgements

K.J.C. and D.J.M. acknowledge support from NERC (NE/S007520/1). K.A.L. acknowledges support by NSF (EAR-1948806) and NASA (80NSSC19K1102). I.M.M. acknowledges support by JSPS KAKENHI grant no. JP20K14586. G.W. and the DIGIS project are supported by the German Science Foundation (Wo362-53-1). K.J.C. and D.J.M. thank the participants of the Volcano-Tectonic Links workshop where the idea for this Comment was seeded.

Competing interests

The authors declare no competing interests.

RELATED LINKS

AstroMat database: <https://www.astromat.org/>
 AuScope Geochemistry Network (AGN): <https://www.auscope.org.au/agn>
 DIGIS initiative: <https://www.uni-goettingen.de/de/643369.html>
 EarthChem Library: <https://earthchem.org/ecl/>
 GEOROC database: <http://georoc.mpch-mainz.gwdg.de/georoc/>
 Multi-Scale Laboratories of the European Plate Observing System (EPOS): <https://www.epos-eu.org/epos-services-list>
 North American Volcanic and Intrusive Rock Database (NAVDAT): <https://www.navdat.org/>
 OneGeochemistry: <https://www.earthchem.org/communities/onegeochemistry/>
 OSIRIS-REx: <https://www.asteroidmission.org>
 PetDB database: <https://search.earthchem.org>