WorldFAIR

Project Title	Global cooperation on FAIR data policy and practice
Project Acronym	WorldFAIR
Grant Agreement No	101058393
Instrument	HORIZON-WIDERA-2021-ERA-01
Topic, type of action	HORIZON-WIDERA-2021-ERA-01-41 HORIZON Coordination and Support Actions
Start Date of Project	2022-06-01
Duration of Project	24 months
Project Website	http://worldfair-project.eu

D5.3 Guidelines for implementing Geochemistry FIPs

Work Package	WP05 - Geochemistry
Lead Author (Org)	Alexander Prent (AuScope, Australia)
Contributing Author(s) (Org)	Rebecca Farrington (AuScope, Australia), Lesley Wyborn (AuScope, Australia), Angus Nixon (University of Adelaide, Australia), Kirsten Elger (GFZ Potsdam, Germany), Marthe Klöcking (University of

	Münster, Germany), Dominik Hezel (Goethe Universität Frankfurt, Germany), Kerstin Lehnert (Columbia University, USA)
Due Date	29.02.2024
Date	26.02.2024
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.10712808

Dissemination Level

- [X] PU: Public
- [ ] PP: Restricted to other programme participants (including the Commission)
- [ ] RE: Restricted to a group specified by the consortium (including the Commission)
- [ ] CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Authors	Notes
0.1	22.01.2024	Alexander Prent, Lesley Wyborn, Rebecca Farrington	Draft for OneGeochemistry and WP05 internal review
0.2	31.01.2024	Alexander Prent, Rebecca Farrington, Lesley Wyborn, Angus Nixon, Kirsten Elger, Marthe Klöcking, Dominik Hezel, Kerstin Lehnert.	Content ready for WorldFAIR internal review
0.3	14.02.2024	Jeff Ollerton, Iseult Lynch	Review of the draft
0.4	20.02.2024	Marthe Klöcking, Angus Nixon, Dominik Hezel	Suggestions and revisions
1.0	22.02.2024	Alexander Prent, Rebecca Farrington	Inclusion of feedback

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Disclaimer

WorldFAIR has received funding from the European Commission's WIDERA coordination and support programme under the Grant Agreement no. 101058393. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

Abbreviations and Acronyms

DOI	Digital Object Identifier; https://www.doi.org/
FAIR	Findable, Accessible, Interoperable, Reusable
FER	FAIR Enabling Resource
FIP	FAIR Implementation Profile
Geochemistry	The study of the chemical composition of the Earth and solid bodies in the solar system, their rocks and minerals, including the distribution, circulation and abundance of elements, their ions and isotopes, molecules and fluids
netCDF	network Common Data Form; a file format for storing multidimensional scientific data (variables)
PID	Persistent Identifier (long-lasting reference to a document, file, web page, or other digital object) for example, a DOI
QA/QC	Quality Assurance / Quality Control
WP05	WorldFAIR Geochemistry Work Package

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Executive summary

As a long-tail scientific discipline with highly specific and heterogeneous analytical methods, the geochemistry community faces challenges in achieving FAIR data compliance. While many repositories satisfy the Findable and Accessible principles of FAIR, increased modernisation of existing standards and development of additional data standards are required to achieve Interoperability and Reusability of data.

This third deliverable of the WorldFAIR Geochemistry Work Package (WP05) aims to guide the geochemistry data infrastructure community towards convergence by identifying FAIR Enabling Resources (FERs) that are currently being used by the geosciences community. Promulgation of used resources and their uptake by other infrastructure providers is part of the push towards convergence. The WorldFAIR Geochemistry Work Package proposes creating a reference FIP or catalogue of FERs to promote interoperability and prevent duplication of efforts. The geochemistry reference FIP is designed as a living document, allowing continual updates by the community. It serves as a tool for laboratories, repositories, and infrastructure providers to enhance data FAIRness. Through the provision of a reference FAIR Implementation Profile (FIP) or catalogue of FERs as part of this report, data providers and producers are provided with a tool to help select FERs when building or updating their infrastructure to become more FAIR.

Together with the second deliverable (D5.2)[1] of the WorldFAIR Geochemistry Work Package which outlined the usefulness and importance of FIPs, this report and the associated reference FIP can be used by the geochemistry community - particularly by data creators and providers - to improve their FAIRness. We recommend that new and emerging geochemistry data producers and providers consult the geochemistry reference FIP and ideally choose to implement existing FERs, although the selection and implementation of FERs should align with the principles and community needs that the specific data system serves.

The goal is to facilitate the implementation of commonly-used FERs, and so improving data FAIRness, with a resource that fosters interoperability, accelerates convergence on data standards, and ultimately enhances the accessibility and reusability of geochemical data. This report and the reference FIP aim to encourage the reuse of available resources, prevent duplication, and enhance convergence on data standards within the geochemistry community. Community collaboration, the continuous evolution of the living reference FIP document to support FAIR compliance and convergence towards standardisation are needed to continue improving FAIRness in the geochemistry data community.

[1] https://doi.org/10.5281/zenodo.10406332

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Table of contents

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

1. Introduction

Geochemistry, with its highly diverse data producing community, has difficulty in complying with the FAIR principles (Wilkinson et al., 2016). Many repositories and databases are internally consistent and can be considered to satisfy the F and A in FAIR. However, to satisfy the I and the R, data from these sources should be interoperable and reusable not just within the immediate community they serve but also outside of it, being interoperable with other countries' databases on geochemistry, as well as with domains beyond geochemistry.

In order to expand the FAIRness of databases, data collections and datasets for both humans and machines, the resources that are used to structure and describe them need to be made available and publicised. Sufficient metadata about the technical solutions and resources - such as vocabularies that were implemented in the database or used by the data collection or the datasets - need to be available in order to improve FAIRness of these resources. There are a reasonable number of data-handling best practice publications and community recommendations for data curation published by the geochemistry community, and an overview of these has been collated by the OneGeochemistry initiative[2]. However, most are available only as PDF files, usually as appendices within research papers or reports: very few are FAIR-compliant for both humans and machines and identifiable with a globally unique persistent identifier (PID). Thus, there is a need for the geochemistry community to modernise many of these existing 'best practices' and reporting templates, which in turn, will require re-curation of many existing datasets to bring them up to the standards that are now required for FAIR compliance.

In WorldFAIR Milestone 6 (MS6; Prent et al., 2023)[3], the WorldFAIR Geochemistry Work Package (WP05) provided advice on publishing vocabularies and other FAIR Enabling Resources (FERs) (Shultes et al., 2020) at any community tier at which they are generated, be that local, regional, international or global (Box 1). This makes it clear how data from that community tier were formatted, and clarifies which terminologies are being used. For example, FERs can contain the common variables used by the community that publicised them. The second deliverable of the WorldFAIR Geochemistry Work Package (D5.2; Prent et al., 2024)[4] spelled out the importance of FAIR Implementation Profiles (FIPs; Shultes et al., 2020), and the role FIPs could take in the standardisation of reporting about which FERs are being used by communities.

[2] http://www.onegeochemistry.org
[3] https://doi.org/10.5281/zenodo.7977116
[4] https://doi.org/10.5281/zenodo.10406332

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Having a standardised manner to report which FERs are used will help accelerate convergence even if FERs are reported but not published. There are almost as many formats and term definitions of variables available as there are laboratories and other data-producing capacities. Few commonly-used variables are being made FAIR-compliant and published by communities using these laboratories and data-producing capacities. The term definitions and other FERs that exist and are used by local, regional, international communities should ideally be promulgated for reuse to accelerate convergence on usage of similar terms and data structures in a larger community (used in a higher tier, Box 1).

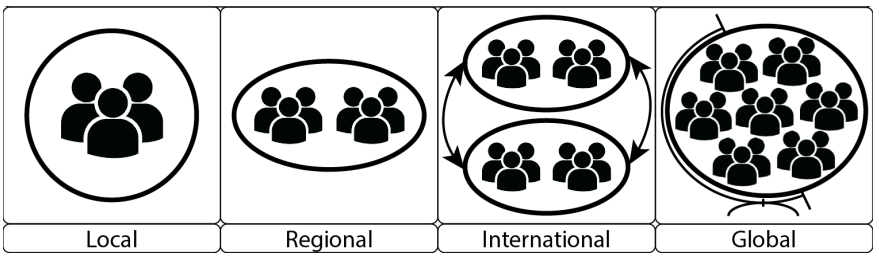
To address this, in this deliverable we aim to provide a reference source, or a catalogue of available FERs that are being used by WP05-associated geochemistry data repositories. This reference catalogue can be consulted and used by laboratories and repositories, as a means to encourage reuse of available resources and prevent 'reinventing of the wheel'. We view this manner of publicising which FERs are being used in any tier of the community, in the FIP format, as the best way, aggregating resources under the same FAIR principles.

Box 1: Community levels by size and/or complexity

In order to make a distinction between the size and/or complexity of a community that works together (FAIRly), four community levels were defined in the second WorldFAIR Geochemistry Work Package deliverable (D5.2; Prent et al., 2024).

The four community levels are:

1. A local community consists of a single laboratory or university.
2. A regional community is one within a city or between cities and universities, multiple laboratories or the community of a discipline within a country.
3. An international community acts as a collaboration between countries.
4. A global community acts as one with all of the community involved.



'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

As it is unlikely that there will be a single common set of variables and structures of data reporting that are internationally agreed to, some flexibility of reporting standards and FERs is required. In their second deliverable report[5], the WorldFAIR Ocean Science and Development Work Package (WorldFAIR WP11) came to a similar conclusion (Buttigieg, 2023). A certain level of flexibility and variability is required to be able to create innovative and original science whilst, at the same time, allowing for local to regional variations in data and metadata standards and vocabularies that are not universally applicable.

The geochemistry reference FIP or FERs catalogue reported on here is intended to become a living document that groups and individual researchers can add to. The reference FIP will be available as an open, version-controlled Google spreadsheet; the community is invited to add resources as they see fit. These additions will be reviewed by the OneGeochemistry interim board on a quarterly basis and new versions can be published via the OneGeochemistry Zenodo community page[6].

In the future an improved method of publication and revision can be developed as per the IUPAC Cookbook example described in the second WorldFAIR Chemistry Work Package deliverable[7] (D3.2; Chalk et al., 2024). Further in this deliverable report we will explain the manner in which to use the reference FIP and how geochemistry data producers and providers should keep their FIPs updated.

This document should be used by those working in laboratories that provide a service to the community by making available commonly-used formatting to their data outputs, as well as by repositories and database infrastructure providers who are looking to find sources that help make their data holdings FAIR, and more specifically, interoperable with other existing (geochemistry) data infrastructures.

2. A geochemistry reference FAIR Implementation Profile

Within the data community, there are many published and online FERs that can be used or implemented to make data FAIR. The incorporation of commonly-used FERs in geochemistry data-providing infrastructures will support interoperability within and between communities, growing the size of the FAIR community. An overview of FERs that are in use by the geochemistry community can drive the implementation of them by new or developing geochemistry

[5] https://doi.org/10.5281/zenodo.10219933
[6] https://zenodo.org/communities/onegeochemistry
[7] https://doi.org/10.5281/zenodo.10711950

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

infrastructure providers. Knowing which FERs are in use can also drive the development of mappings and crosswalks between actively-used FERs, furthering interoperability between geochemistry communities. The incorporation of more generically-used FERs from interdisciplinary areas can lead to cross-domain interoperability which will be increasingly sought after in the near future.

Although the concepts of FIPs and FERs are at present relatively unknown in the geochemistry data infrastructure community, infrastructure providers use resources that make their data FAIR. These resources can be considered FERs although they are often not published or made available, or presented as such, and thus there is limited knowledge of which resources are actually being used. Internal formatting of data according to a template (e.g., for ingestion of a dataset into a repository) uses specific FERs embedded within, implicitly advocating the usage of specific FERs. Community 'best practice' advice on data reporting practices (e.g., usage of metadata profiles, data formats, terminologies/vocabularies, etc.) can also be considered FERs and often form the basis for data ingestion templates. Data producers and providers thus use FERs informally and rarely publish or promulgate them despite the fact that they are foundational building blocks for interoperability, reuse and machine readable data. Having access to these FERs and creating an overview of the ones that are being used within the geochemistry community could support data infrastructure providers in achieving greater compatibility and interoperability, as well as preventing them from unnecessarily 'reinventing the wheel' due to lack of awareness of relevant existing solutions. Compiling a list of technical solutions and resources commonly used in geochemistry data systems, or in other domains that could be implemented also for geochemistry, collated within a reference FIP or catalogue of FERs as a tool to the community could stimulate data interoperability and convergence by facilitating usage of the same FERs. It will enable each infrastructure provider to choose whether to implement the same resources, map their used resources to existing ones, or build from the ground up a system that conforms with the resources generally in use.

The geochemistry reference FIP provided here[8,9] lists all the used FERs by domain repositories amongst the WorldFAIR Geochemistry Work Package proponents' infrastructures (Table 1). This reference FIP for geochemistry is a catalogue providing options for resources in making geochemistry data FAIR. This catalogue is organised by each of the 15 FAIR Guiding Principles, stating the set of FERs that go with the respective principle and question used to create a FIP. This overview includes the resource name and reference (nanopublication), its usage (related to a specific method or general usage), the user(s) of the FER (repository, laboratory or other

[8] https://docs.google.com/spreadsheets/d/1MXJ2AgLd-pOny5TdtSa8DHJRFFnhqAtnMtHs59Iggh0/edit?usp=sharing

[9] WorldFAIR FIPs will also be added by project end to the WorldFAIR FIP Collection at https://zenodo.org/communities/worldfair-project-fips/records

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

infrastructure provider) and the level of data granularity it applies to (dataset, data collection, disciplinary/synthesis database, repository; see D5.2; Prent et al., 2024). The notes added provide information for infrastructure providers that intend to implement FERs from this list to help choose the ones that are most applicable for their use case, facilitating convergence.

The geochemistry reference FIP is created to be used by new and existing data producers and providers to implement the same FERs or decide on how to build crosswalks between commonly-used FERs in order to improve data FAIRness. Crosswalks provide a means to interoperate between resources and could take the form of a tabular mapping of existing terms of both or more resources enabling translation between standards used (e.g., Crystal-Ornelas et al., 2022) or else through ontology mappings (e.g., Matentzoglu et al., 2022)

The geochemistry reference FIP is by no means exhaustive, but it does provide an overview of the resources being used by geochemistry repositories, databases, platforms and collaborating infrastructures currently active in the WorldFAIR project. The geochemistry reference FIP is considered a living document which could form the basis of a geochemistry specific Best Practice System website, yet to be designed and funded, akin to the Oceans Best Practices System website[10] and search portal[11]. This website enables publication, from any organisation involved in data standards and best practices related to ocean research, of FAIR-related documents including standards, netCDF conventions, ontologies, specifications for data collected by a specific instrument type, QA/QC validation procedures, QA/QC validation documents, etc. (Przeslawski et al., 2023).

Table 1. WorldFAIR geochemistry specific repositories on which the reference FIP v1.0 is based.

Geochemical repository name	**Description (short)**	**Reference / weblink**
DIGIS Geochemical Data Repository	The DIGIS Geochemical Data Repository, hosted by GFZ Data Services as a dedicated data centre, hosts research data within the scope of the GEOROC database: geochemical compositions of rocks, glasses, minerals and inclusions from all geological settings on Earth. The repository is curated by the Digital Geochemical Data Infrastructure (DIGIS) project at Göttingen University.	https://uni-goettingen.de/en/georoc+data+repository/651621.html
EarthChem Library	The EarthChem Library (ECL) is an open-access repository	https://www.earthchem.org/ecl/

[10] https://www.oceanbestpractices.org/
[11] https://search.oceanbestpractices.org/

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

	operated as part of the IEDA2 suite of data services. The ECL publishes and preserves laboratory analytical data and field observations including geochemical, mineralogical, petrological, geochronological, and experimental data from a wide range of disciplines (e.g., volcanology, geodynamics, paleoclimate, environmental sciences). The EarthChem Library curation services offer long-term archiving, DOI registration with DataCite, peer review of datasets, metadata templates, and machine-readable metadata that include persistent links to samples via IGSN and related publications via DOI.	
GFZ Data Services	GFZ Data Service is a repository for research data and scientific software across the Earth System Sciences, hosted at GFZ. The curated data are archived, persistently accessible and published with digital object identifier (DOI). They range from large dynamic datasets from global monitoring networks with real-time acquisition, to international services in geodesy and geophysics, to the full suite of small and highly heterogeneous datasets collected by individual researchers or small teams ("long-tail data"). In addition to the DOI registration and data archiving itself, GFZ Data Services team offers comprehensive consultation by domain scientists and IT specialists. Among others, GFZ Data Services is data publisher for several international services and networks in the geodesy and geophysics, as well as the official repository for DIGIS/ GEOROC since 2023. GFZ Data Services is also hosting the service and catalogue for International Generic Sample Numbers (IGSN) at GFZ.	https://dataservices.gfz-potsdam.de
AusGeochem	AusGeochem is a community oriented platform for easily visualising, analysing, extracting and interpreting georeferenced data produced by geochemistry laboratories around the world. This service allows users to directly upload data during the research process while offering privacy options prior to publication, and provides services for minting dataset DOIs and sample IGSNs for compliance with journal discoverability requirements. The project is an operation of the AuScope Geochemistry Network (AGN), an NCRIS enabled consortium of Australian Earth science institutes and researchers.	https://www.auscope.org.au/ausgeochem

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

3. Recommendations

This report has generated an overview of FERs in use by WorldFAIR Geochemistry Work Package associated data repositories aggregated into a reference FIP or catalogue of FERs which can be used by new and emerging data producers and providers to select from and implement FERs that are already in use, leading to convergence in the technologies used to make data FAIR in the geochemistry community.

This report makes the following recommendations:

3.1. Geochemistry community policy recommendations

Stakeholders: geochemistry associations, unions and societies, geochemistry journals, geochemistry data standardisation initiatives.

Recommendations:

1. A geochemistry website should be built that follows on the example of the Oceans initiative where FERs from the reference FIP and other resources can be published, promulgated and updated by the relevant local, regional, international and global communities.
2. A call should go out to the geochemistry community to list their existing FERs used in respective repositories and data producing facilities into the living document so the reference FIP can grow and stimulate convergence.
3. Data infrastructures are recommended to document their own FIPs utilising the reference FIP and other FERs as per their current or planned future practice.
4. Similar FERs should be identified and discussions about building crosswalks between them should be initiated.

3.2. Data infrastructure recommendations

Stakeholders: data producers, data publishers, data/informatics specialists, digital architects, software developers and team leads, data- or software-focused researchers, (research) data managers.

Recommendations:

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

1. New and emerging geochemistry data producers and data providers should consult the geochemistry reference FAIR Implementation Profile in order to use and implement existing FERs in their data infrastructures.
2. In using the reference FIP, look for a FAIR principle for which your system currently has no FAIR Enabling Resources implementation as yet. Then choose a FER that has been published (in a machine actionable fashion) and has community uptake.
3. Select and implement a FER that best suits the community behind creating, updating or advancing of the data system, determined on a case by case basis.

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

3. References / bibliography

Buttigieg, P. L. (2023). WorldFAIR (D11.2) New interoperability specifications and policy recommendations. https://doi.org/10.5281/ZENODO.10219933

Crystal-Ornelas, R., Varadharajan, C., O'Ryan, D., Beilsmith, K., Bond-Lamberty, B., Boye, K., Burrus, M., Cholia, S., Christianson, D.S., Crow, M., Damerow, J., Ely, K.S., Goldman, A.E., Heinz, S.L., Hendrix, V.C., Kakalia, Z., Mathes, K., O'Brien, F., Pennington, S.C., Robles, E., Rogers, A., Simmonds, M., Velliquette, T., Weisenhorn, P., Welch, J.N, Whitenack, K., Agarwal, D.A. (2022). Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. Sci Data 9, 700. https://doi.org/10.1038/s41597-022-01606-w

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., … Schultes, E. (2020). FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence, 2(1–2), 10–29. https://doi.org/10.1162/dint_r_00024

Matentzoglu, N., Balhoff, J. P., Bello, S. M., Bizon, C., Brush, M., Callahan, T. J., Chute, C. G., Duncan, W. D., Evelo, C. T., Gabriel, D., Graybeal, J., Gray, A., Gyori, B. M., Haendel, M., Harmse, H., Harris, N. L., Harrow, I., Hegde, H. B., Hoyt, A. L., … Mungall, C. J. (2022). A Simple Standard for Sharing Ontological Mappings (SSSOM). Database, 2022, baac035. https://doi.org/10.1093/database/baac035

Prent, A., Wyborn, L., Farrington, R., Lehnert, K., Klöcking, M., Elger, K., Hezel, D., ter Maat, G., Profeta, L., & Rawling, T. (2023). WorldFAIR Project (MS6) Geochemistry Scientific Content Component. https://doi.org/10.5281/ZENODO.7977116

Prent, A., Farrington, R., Wyborn, L., Nixon, A., Elger, K., Klöcking, M., Hezel, D., Lehnert, K. (2024). WorldFAIR Project (D5.2) Geochemistry Methodology & Outreach. https://doi.org./10.5281/zenodo.10406332

Przeslawski, R., Barrett, N., Carroll, A., Foster, S., Gibbons, B., Jordan, A., Monk, J., Langlois, T., Lara-Lopez, A., Pearlman, J., Picard, K., Pini-Fitzsimmons, J., Van Ruth, P., & Williams, J. (2023). Developing an ocean best practice: A case study of marine sampling practices from Australia. Frontiers in Marine Science, 10, 1173075. https://doi.org/10.3389/fmars.2023.1173075

Schultes, E., Magagna, B., Hettne, K. M., Pergl, R., Suchánek, M., & Kuhn, T. (2020). Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In G. Grossmann & S. Ram (Eds.), Advances in Conceptual Modeling (Vol. 12584, pp. 138–147). Springer International Publishing. https://doi.org/10.1007/978-3-030-65847-2_13

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N.,

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Appendix: Reference FAIR Implementation Profile for Geochemistry

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

WorldFAIR

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

TABLE NAME	**Geochemistry FER list per FIP question**				
TABLE Description	**An overview of resources available and relevant to the geochemistry community (sometimes already in use)**				
TABLE Resources	**FIPs Questions and FAIR Enabling Resource Types**				
First a table follows that lists all the questions that make up a FAIR Implementation Profile (including their explanations and FER type definition) -- it is presently collapsed					
FAIR Principle name	Referring to MetaData/Data	Definition of FAIR (sub)Principle	FIP question	Type of FAIR Enabling Resource	FER Type definition
F1 (6 questions total)	MD	(meta)data are assigned a globally unique and persistent identifier	What globally unique, persistent, resolvable identifier service do you use for metadata records?	Identifier service	A **service** that provides for metadata (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistence and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F1	D	data are assigned a globally unique and persistent identifier	What globally unique, persistent, resolvable identifier service do you use for datasets?	Identifier service	A **service** that provides for data (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistence and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	MD	data are described with rich metadata	What metadata schemas do you use for findability?	Metadata schema	A **specification** that specifies metadata fields describing attributes of data or other digital objects in terms of semantics, syntax and optionality.
F3	D	metadata clearly and explicitly include the identifier of the data it describes	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	MD	(meta)data are registered or indexed in a searchable resource	In what lookup services are your metadata records indexed?	Search engines	A **service** that indexes metadata and provides search over that index.
F4	D	(meta)data are registered or indexed in a searchable resource	In what lookup services are your datasets indexed?	Search engines	A **service** that indexes data and provides search over that index.
A1.1 (5 questions total)	MD	(meta)data are retrievable by their identifier using a standardized communications protocol: the protocol is open, free, and universally implementable	What standardized communication protocol do you use for metadata records?	Communication protocol	A **specification** of how messages are structured and exchanged.
A1.1	D	(meta)data are retrievable by their identifier using a standardized communications protocol: the protocol is open, free, and universally implementable	What standardized communication protocol do you use for datasets?	Communication protocol	A **specification** of how messages are structured and exchanged.
A1.2	MD	(meta)data are retrievable by their identifier using a standardized communications protocol: the protocol allows for an authentication and authorization procedure, where necessary	What authentication & authorisation technique do you use for metadata records?	Authentication and Authorization	A **service** that mediates access to digital objects according to specifed conditions
A1.2	D	(meta)data are retrievable by their identifier using a standardized communications protocol: the protocol allows for an authentication and authorization procedure, where necessary	Which authentication & authorisation technique do you use for datasets?	Authentication and Authorization	A **service** that mediates access to digital objects according to specifed conditions
A2	MD	metadata are accessible, even when the data are no longer available	What metadata longevity policy do you use?	Metadata longevity policy	A **data policy** that describes the conditions under which metadata should be provided in the future.
I1 (6 questions total)	MD	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	What knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation langu	A language **specification** that enables knowledge to be processed by machines.
I1	D	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	What knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation langu	A language **specification** that enables knowledge to be processed by machines.
I2	MD	(meta)data use vocabularies that follow FAIR principles	What structured vocabularies do you use to annotate your metadata records?	Controlled vocabulary	A **specification** for a controlled list of uniquely identified and unambigous concepts with their definitions represented using web standards.
I2	D	(meta)data use vocabularies that follow FAIR principles	What structured vocabularies do you use to encode your datasets	Controlled vocabulary	A **specification** for a controlled list of uniquely identified and unambigous concepts with their definitions represented using web standards.
I3	MD	(meta)data include qualified references to other (meta)data	What models / schema(s) do you use for your metadata records?	Metadata schema	A **specification** that defines qualified relations between entities describing data or other digital objects according to the Linked Data principles. This can include semantic data models and ontologies.
I3	D	(meta)data include qualified references to other (meta)data	What models / schema(s) do you use for your datasets?	Data schema	A **specification** that defines qualified relations between entities describing data or other digital objects according to the Linked Data principles. This can include semantic data models and ontologies.
R1.1 (4 questions total)	MD	(meta)data are richly described with a plurality of accurate and relevant attributes: (meta)data are released with a clear and accessible data usage license	Metadata: Which usage license do you use for your metadata records?	Data usage license	A **data policy** that specifies and guides FAIR data practices for a community.
R1.1	D	(meta)data are richly described with a plurality of accurate and relevant attributes: (meta)data are released with a clear and accessible data usage license	Which usage license do you use for your datasets?	Data usage license	A **data policy** that specifies and guides FAIR data practices for a community.
R1.2	MD	(meta)data are richly described with a plurality of accurate and relevant attributes: (meta)data are associated with detailed provenance	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	A **specification** that specifies metadata fields describing the origin and lineage of data or other digital objects.
R1.2	D	(meta)data are richly described with a plurality of accurate and relevant attributes: (meta)data are associated with detailed provenance	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	A **specification** that specifies metadata fields describing the origin and lineage of data or other digital objects.
F1 Metadata: What globally unique, persistent, resolvable identifier service do you use for metadata records?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
ORCID | Open Researcher and Contributor ID	http://purl.org/np/RA1MzU1MPio-mtLzm1P7zfTBSMnTNc2I8HLNARhPjpif8#ORCID	AusGeochem	Current		
IGSN (DataCite)		AusGeochem	Current		

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

AusGeochem itself provides a service to publish metadata. All metadata is indexed and provides a search over that index.		AusGeochem	Current		
EarthChem Library; DataCite; GeoCODES; DataONE		EarthChem	Current		
GFZ Data Services portal/catalogue, IGSN Portal (own development, SolR search engine with underlying MySQL metadatabases		GFZ Data Services, DIGIS	current	GFZ Data Services: https://dataservices.gfz-potsdam.de/portal IGSN Catalogue: https://dataservices.gfz-potsdam.de/igsn-new/	
F4 Datasets: Which service do you use to publish your datasets?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
Datasets are published using the DOI minting feature for datapackages.					
re3data | Registry of Research Data Repositories	http://doi.org/10.17616/R31NJNAQ	AusG	Current		
EarthChem Synthesis if data are appropriate; EarthChem Portal;		EarthChem	Current		
DataCite DOIs / IGSNs		GFZ Data Services, DIGIS	current		
A1.1 Metadata: Which standardized communication protocol do you use for metadata records?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
REST-API | Representational state transfer	http://purl.org/np/RAszH6lU-Zc3UO7MHPKj1Lb0dmMmaTJrRvQ0jqpXMyFY4#REST	AusG	Current		
REST		EarthChem	Current		
OAI-PMH | Open Archives Initiative Protocol for Metadata Harvesting		GFZ Data Services, DIGIS	current		
A1.1 Datasets: Which standardized communication protocol do you use for datasets?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
REST-API | Representational state transfer	http://purl.org/np/RAszH6lU-Zc3UO7MHPKj1Lb0dmMmaTJrRvQ0jqpXMyFY4#REST	AusG	Current		
OGC-API		AusG			
REST		EarthChem	Current		
HTTPS | Hypertext Transfer Protocol Secure		GFZ Data Services, DIGIS	current		
A1.2 Metadata: Which authentication & authorisation service do you use for metadata records?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
Authentication is enabled by password protected JWT (Jason Web Token). Authorization is enabled by an data package based Access Control Layer.		AusG	Current		
OAuth		EarthChem	Current		
Open Data		GFZ Data Services, DIGIS	current		
A1.2 Datasets: Which authentication & authorisation service do you use for datasets?					
FER (name)	Reference (link)	Used by	Implementation status	Note	

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

Authentication is enabled by password protected JWT (Jason Web Token). Authorization is enabled by an data package based Access Control Layer.		AusG	Current		
OAuth		EarthChem	Current		
Open Data		GFZ Data Services, DIGIS	current		
A2: What metadata preservation policy do you use?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
RDA Core Trust Seal Certification	http://purl.org/np/RAmSQlUhDvNjDr1ECxPD-lKSTPj5hLAutEZ1zjc6Pnmy8#RDA_CTS_Cert	AusG	Future		
CoreTrustSeal?		EarthChem	Future		
RDA Core Trust Seal Certification (in prep)		GFZ Data Services, DIGIS	future		
I1 Metadata: What knowledge representation language (allowing machine interoperation) do you use for metadata records?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
A knowledge representation language is the analogue in AI of the language of thought in cognitive science. AusGeochem does not implement a knowledge representation language.		AusG	Current		
XML		EarthChem	Current		
XMLS|eXtensible Markup Language Schema; JSON-LD (schema.org)		GFZ Data Services, DIGIS	current		
I1 Datasets: What knowledge representation language (allowing machine interoperation) do you use for datasets?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
A knowledge representation language is the analogue in AI of the language of thought in cognitive science. AusGeochem does not implement a knowledge representation language.		AusG	Current		
ECL spreadsheet templates		EarthChem	current; to be replaced		
various ASCII and binary formats, the use of community standards is encouraged		GFZ Data Services, DIGIS	current	GFZ Data Services is serving the entire geosciences community. It has a focus on long-tail data, but additionally serves larger communities in geodesy and geophysics - each with their specific, community-agreed standards	
I2 Metadata: What structured vocabulary do you use to annotate your metadata records?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
Mixture of internal and IGSN proposed / ODM3 / Others... In general AusGeochem uses the object oriented (OO) pattern of extension.	none	AusG	Current	Needs to be replaced by a structured and published vocabulary	
Specific domain vocabulary.					
Domain vocabulary aggregator					
Cross-domain vocabulary aggregator (e.g. Research Vocabularies Australia)					
Research Vocabs Australia		EarthChem	Future		

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

(1) GCMD|Global Change Master Directory; (2) GeoSciML: simple lithology, geochronology, etc (3) GEMET|General Multilingual Environmental Thesaurus (4) Analytical Methods (via RVA)		GFZ Data Services, DIGIS	(1) current (2) current (3) current - but not active at the moment (4) future	RDF vocabularies	
I2 Datasets: What structured vocabulary do you use to encode your datasets?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
Custom vocabulaires to be published through Research Vocabularies Australia		AusG	Future		
Sample Identifiers and Metadata to Support Data Management and Reuse in Multidisciplinary Ecosystem Sciences	http://doi.org/10.5334/dsj-2021-011	AusG	Future		
PaCTS 1.0: a crowdsourced reporting standard for paleoclimate data	https://doi.org/10.1029/2019PA003632	AusG	Future		
Interpreting and Reporting 40Ar/39Ar Geochronological Data	https://doi.org/10.1130/B35560.1	AusG	Future		
Guidelines for reporting and archiving 210Pb sediment chronologies to improve fidelity and extend data lifecycle	https://doi.org/10.1016/j.quageo.2019.04.003	AusG	Future		
Geochron Workshop reports sponsored by EarthChem and EARTHTIME	https://doi.org/10.5281/zenodo.4313859	AusG	Future		
Data reporting standards for publication of U-series data for geochronology and timescale assessment in the earth sciences.	https://doi.org/10.1016/j.quageo.2017.03.001	AusG	Future		
Community-Derived Standards for LA-ICP-MS U-(Th-)Pb Geochronology – Uncertainty Propagation, Age Interpretation and Data Reporting	https://doi.org/10.1111/j.1751-908X.2016.00379.x	AusG	Future		
Community Established Best Practice Recommendations for Tephra Studies-from Collection through Analysis	https://doi.org/10.5281/zenodo.3866266	AusG	Future		
A Vision for the Future Low-Temperature Geochemical Data-scape	https://doi.org/10.31223/X5ZP5W	AusG	Future		
(U-Th)/He chronology: Part 1. Data, uncertainty, and reporting	https://doi.org/10.1130/B36266.1	AusG	Future		
AusGeochem: An Open Platform for Geochemical Data Preservation, Dissemination and Synthesis	https://doi.org/10.1111/ggr.12419	AusG	Future		

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

(1) GCMD|Global Change Master Directory; (2) GeoSciML: simple lithology, geochronology, etc (3) GEMET|General Multilingual Environmental Thesaurus (4) Analytical Methods (via RVA)		GFZ Data Services, DIGIS	(1) current (2) current (3) current - but not active at the moment (4) future	RDF vocabularies	
I3 Metadata: What semantic model do you use for your metadata records?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
AGN designed relational model		AusGeochem			
DataCite: relatedIdentifier properties		GFZ Data Services, DIGIS	current	we cite all related work (papers, data, samples, software) using the assigned PIDs and include these in the DataCite Metadata Schema	
I3 Datasets: What semantic model do you use for your datasets?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
AGN designed relational model		AusGeochem			
none		GFZ Data Services, DIGIS	current	we have too many too different datasets	
R1.1 Metadata: Which usage license do you use for your metadata records?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
CC BY 4.0 | Attribution 4.0 International	http://purl.org/np/RAQ__sGdY_Qc7l1O_zmn4nr-pMBOxKU04Ur9s998rS6Fc#CC-BY-4.0	AusGeochem	Current		
Open metadata?				used for EO data	
CC0 or CC-BY		EarthChem	Current		
no licence assigned, metadata are in the public domain.		GFZ Data Services, DIGIS	current		
R1.1 Datasets: Which usage license do you use for your datasets?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
CC BY 4.0 | Attribution 4.0 International	http://purl.org/np/RAQ__sGdY_Qc7l1O_zmn4nr-pMBOxKU04Ur9s998rS6Fc#CC-BY-4.1	AusG	Current		
CC0 or CC-BY		EarthChem	Current		
CC BY 4.0|Attribution 4.0 International; CC BY-NC 4.0|Attribution-NonCommercial 4.0 International; GNU General Public License v3.0 or later and other open source software licences (MIT, Apache, LGPL...)		GFZ Data Services, DIGIS	current		
R1.2 Metadata: What metadata schema do you use for describing the provenance of your metadata records?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
The metadata schema to define orgin and lineage consists of two fields: dataSource and sourceId.					

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.

none, unless included in the DataCite metadata		GFZ Data Services, DIGIS	current		
R1.2 Datasets: What metadata schema do you use for describing the provenance of your datasets?					
FER (name)	Reference (link)	Used by	Implementation status	Note	
The metadata schema to define orgin and lineage consists of two fields: dataSource and sourceId.					
ODM2; Observations & Measurements		EarthChem			
none, when data are updated, we use a new version and make the connection between both versions in the DataCite Metadata		GFZ Data Services, DIGIS	current		
R1.3: Your community uses this FAIR Implementation Profile to link to domain-relevant community standards. Please acknowledge this statement by clicking on 'Read and understood'.					

'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.