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

D5.2 Geochemistry Methodology and Outreach

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 German Research Centre for Geosciences, Germany), Marthe Klöcking (University of Münster, Germany), Dominik Hezel (Goethe Universität Frankfurt, Germany), Kerstin Lehnert (Columbia University, USA)
Reviewing Author(s) (Org)	Beth Knazook, Leah McEwen
Due Date	29.02.2024
Date	02.02.2024

WorldFAIR

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

Version	2.0
DOI	10.5281/zenodo.10406332

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	09.05.2023	Alexander Prent	Collection of papers and ideas towards completion of D5.2
0.2	13.06.2023	Alexander Prent	Strategy and action items to complete for D5.2
0.3	24.07.2023	Alexander Prent	Structuring of document and writing part of Introduction from MS6 report
0.4	21.10.2023	Alexander Prent, Lesley Wyborn, Rebecca Farrington	Introduction, FIPs, FERs & explaining the need for a reference FIP
0.5	06.11.2023	Alexander Prent, Rebecca Farrington, Lesley Wyborn	First draft of the complete deliverable, for review of project proponents
1.0	19.12.2023	Alexander Prent, Rebecca Farrington, Lesley Wyborn, Angus Nixon, Kirsten Elger, Marthe Klöcking, Dominik Hezel, Kerstin Lehnert.	Project proponents' feedback and revisions incorporated. Draft ready for internal WorldFAIR project review.

'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.

1.1	14.01.2024	Beth Knazook, Leah McEwen, Laura Molloy	WorldFAIR internal review, suggestions, rephrasing, and grammatical edits
2.0	30.01.2024	Alexander Prent, Rebecca Farrington, Kirsten Elger, Marthe Klöcking, Lesley Wyborn, Dominik Hezel, Angus Nixon, Kerstin Lehnert	Incorporated WorldFAIR internal feedback ready for submission to EC.

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.

'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.

Terms and Abbreviations

15 FAIR Guiding Principles	Subdivision of the FAIR acronym in 15 principles that are integral to the FAIR concept and making data consistently and persistently Findable, Accessible, Interoperable and Reusable (Jacobsen et al., 2020; Wilkinson et al., 2016)
21 Questions to create a FIP	21 questions that are phrased in such a way that the answers lead to a list of FAIR Enabling Resources a community has chosen to use in addressing all the 15 FAIR Guiding Principles, which combined makes up a FIP. 21 because some of the questions apply to metadata and data pertaining the 15 principles (Schultes et al., 2020).
CODATA	Committee on Data of the International Science Council, https://codata.org/
DDI Alliance	Data Documentation Initiative Alliance. https://ddialliance.org/
DDI–CDI	DDI Cross Domain Integration
ESIP	Earth System Information Partners, https://www.esipfed.org/
FAIR	Findable, Accessible, Interoperable, Reusable
FER	FAIR Enabling Resource
FIP	FAIR Implementation Profile
GenSC	Genomics Standards Consortium, https://www.gensc.org/
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
IUPAC	International Union of Pure and Applied Chemistry, https://iupac.org/
IUSSP	International Union for the Scientific Study of Population (USSP) CODATA Working Group, https://codata.org/initiatives/decadal-programme2/fair-vocabularies/iussp-codata-working-group-on-fair-vocabularies/
OneGeochemistry	Global data standardisation initiative of geochemistry data repository managers, www.onegeochemistry.org
OneGeology	The OneGeology Initiative, https://onegeology.org/about/governance.html
QA/QC	Quality Assurance / Quality Control
RDA	Research Data Alliance, https://www.rd-alliance.org/
RDA Vocabulary Services IG	RDA Vocabulary Services Interest Group https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html
Resource maturity tiers	Resources, such as vocabularies, can be more or less mature when they are used by (1) local, (2) community or (3) international groups

'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

Together with the earlier WorldFAIR Milestone 6[1], this D5.2 report focuses on advocating the utility and significance of FAIR Implementation Profiles (FIPs) for the geochemistry community, culminating in presenting a set of policy and organisational recommendations. The primary goal of this report is to foster alignment across the complex and heterogeneous geochemistry community, in producing and integrating FAIR data for the huge diversity of sample types and target analytes of this community, each often having numerous analytical methods. This document presents various ways in which the community can increase FAIRness through the publication of FERs for different levels of data granularity and FAIR community size and complexity (Figure 2). Additionally, interoperability of data between methodologies is suggested to be overcome through data abstraction (Box 1).

Following the FIP methodology, this D5.2 report makes reference to the fifteen FAIR Principles, divided into scientific and technical components. Scientific component implementations, and related community engagement, are to be based on best practice publications that outline data reporting and methodology descriptions from within specific geochemistry sub-disciplines. Parts of these publications, including tables and images in PDF or document formats, could be converted into machine actionable FAIR-enabling resources (FERs), and be part of a generic FIP for geochemistry. Technical components need to be generated, reviewed and assessed by geochemistry data infrastructure and repository technical staff, along with the development of additionally needed FERs in consultation with other FAIR data management expert groups (e.g., CODATA-DDI Alliance activity, the DDI-CDI group, the RDA Vocabulary Services Interest Group, IUPAC, etc.) and the “Ten Simple Rules for making a vocabulary FAIR” (Cox et al. 2021).

This report is the result of interactions with the geochemistry community through the OneGeochemistry Initiative, its board members, research infrastructure experts, analytical facilities and international leaders in geochemistry data management systems (EarthChem[2],

[1] WorldFAIR MS6: https://zenodo.org/doi/10.5281/zenodo.7963469
[2] EarthChem. https://www.earthchem.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.

DIGIS[3]-GEOROC[4], AGN[5]–AusGeochem[6], GFZ Data Services[7], NFDI4Earth[8], and EPOS[9] MSL Laboratories[10]).

[3] DIGIS: Digital Geochemical Infrastructure. https://www.uni-goettingen.de/de/643369.html
[4] GEOROC: Geochemistry of Rocks of the Oceans and Continents. https://georoc.eu/
[5] AGN: Australian Geochemistry Network. https://www.auscope.org.au/agn
[6] AusGeochem. https://ausgeochem.auscope.org.au
[7] GFZ Data Services. https://bib.telegrafenberg.de/dataservices/
[8] NFDI4Earth: German National Research Data Infrastructure for Earth. https://www.nfdi4earth.de/
[9] EPOS: European Plate Observing System. https://www.epos-eu.org/
[10] EPOS Multi-Scale Laboratories. https://www.epos-eu.org/tcs/multi-scale-laboratories

‘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

The Horizon Europe-funded WorldFAIR project focuses on making research data more interoperable and reusable within its eleven domain-specific work packages and is identifying ways to enable interoperability across multiple scientific domains. The WorldFAIR Geochemistry Work Package (WP05) is developing ways to make geochemical research data more Findable, Accessible, Interoperable and Reusable, or 'FAIR' (Wilkinson et al., 2016), and will produce three interconnected deliverables aligned with this goal. Klöcking et al. (2023) defines the required components to achieve this goal, as determined in a community workshop held in mid-2022; many of these components have been incorporated into WP05 activities.

The current report, D5.2, expands on the initial geochemistry FAIR Implementation Profile (FIP; Schultes et al., 2020), created as part of the output of WorldFAIR Work Package 02 Cross-Domain Interoperability Framework (Gregory and Hodson, 2022) and provides a detailed explanation of what FIPs are, why they are useful and how they can be generated, updated and used within the geochemistry community. The goal of this report is to highlight the importance of using FIPs as a tool for FAIRification of data and facilitate convergence between datasets, data collections, and data-serving infrastructures. The report also outlines how FIPs can be applied to interrogate the quality, interoperability and reusability of data publications as part of journal article publications or separate to them.

It is expected that this document will be especially useful for data repositories, database creators, data producers and data infrastructure managers, by showing them how to optimise existing solutions from the wider FAIR data community and apply these specifically to geochemical data. Building on these FAIR resources, and engaging current geochemistry infrastructure, organisations and authoritative bodies will help develop, endorse and maintain standards for the community (Hodson and Gregory, 2023).

The recommendations from Klöcking et al. (2023) include: i) developing an overview, thereby gaining a better understanding, of existing standards, best practices, schemas and vocabularies in geochemistry; ii) gaining endorsement from the diverse geochemistry sub-communities for the collective OneGeochemistry initiative; and iii) engaging with and educating the community through promulgating knowledge of best practices, standards and

'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.

vocabularies to make geochemical data FAIR (including organising workshops and science sessions at relevant conferences).

WP05 has acted on the recommendations of Klöcking et al. (2023) by aggregating known 'best practice' publications for geochemical methodologies on the OneGeochemistry website[11], and by completing a lightweight information-gathering exercise on the usage of FAIR Enabling Resources (FERs) in the geochemistry community. An initial geochemistry FIP[12] was produced as part of the WorldFAIR WP02 ('Engagement, Synthesis, Recommendations and FAIR Assessment') task of creating an overview of the existing technical and community resources that were available at the commencement of the WorldFAIR project (Gregory and Hodson, 2022).

WP05 views using FIPs as an excellent tool for facilitating FAIR convergence within the heterogeneous geochemistry community. However, it is important for data infrastructure managers firstly to know how to generate, update and communicate FIPs relevant to the geochemical data assets in their care; and secondly, to understand what can be achieved with broader community knowledge of, and access to, FIPs. A community-driven systematic approach proposed in WorldFAIR Milestone 6, WP05 (Prent et al., 2023) will enable the comparison of FIPs across different tiers of community complexity - defined in Milestone 6 as 'local', 'community' and 'international'. As not all these terms are an indication of group size or complexity of the group, we propose to change the terms to 'local', 'regional', 'international' and 'global'. Further, an understanding of the FERs needed to produce FAIR data is necessary before potentially tailoring these resources to the heterogeneous geochemistry community, particularly for enabling machine-to-machine interoperability of globally-distributed geochemical data. Once a FIP has been generated for a particular data granularity, community tier and sub-discipline, a formal governance mechanism is required to guarantee sustainability (preservation) and relevance (evaluation) to ensure the persistence of the particular FIP.

[11] OneGeochemistry website: www.onegeochemistry.org
[12] WorldFAIR WP05 initial geochemistry FIP, October 2022: https://docs.google.com/spreadsheets/d/e/2PACX-1vTFCeI4OaA4xsgboHPEfRrCJh3A253Cz41khi1rcxKAV3zGglPlHydTHUoedioS3nMAxuapEq3nVdn6/pubhtml

'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.

2. Making data FAIR

With data performing critical roles in scientific progress (as they form the basis to test hypotheses, develop theories, and expand knowledge), their sustainable curation, storage and organisation to support reuse are highly valuable activities. Therefore, investing in processes to ensure data are FAIR has the potential to transform science and should be a necessity for all research disciplines. Doing so stimulates collaboration, accelerates discovery, prevents duplication of efforts, supports robust and thorough analysis of complex and interdependent processes and, above all, enables researchers to build upon existing knowledge.

With the wide uptake of the FAIR principles in many scientific communities, the ability to assess the FAIR alignment of a community, a data collection or a dataset, and to articulate how they can be made FAIRer is highly beneficial. For example: for effective reuse, a dataset must be accompanied by additional information (metadata), and enable all of the 15 FAIR Guiding Principles (Jacobsen et al., 2020; Wilkinson et al., 2016; Figure 1). For reference and consistent usage, the 15 FAIR Guiding Principles, as interpreted by Jacobsen et al. (2020), have been summarised in Appendix 1 following the interpretation of the GO FAIR Foundation[13]. In recent years, several tools have been developed to test maturity and alignment with the FAIR principles (e.g., Devaraju et al., 2021; Krans et al., 2022; Peters-von Gehlen et al., 2022). However, to have utility within a heterogeneous scientific community such as geochemistry, a large degree of freedom is necessary to enable new science, and the developed tools are insufficiently flexible to quantitatively define FAIR alignment.

Currently, the requirements for interoperability and reuse are set by the community that works with and produces a particular set or type of data. How the 15 principles are addressed is influenced by the availability of technical solutions and agreed community standards and resources (e.g., metadata schemas, vocabularies) used to find, access, interoperate and reuse their data. Domain repositories set up by and for the community provide data storage and organisation solutions, and together with their community they can build on existing resources. In order to align the community further, an 'Editors' Roundtable' was held during the International Goldschmidt 2023 conference in Lyon[14], where journal data submission

[13] https://www.gofair.foundation/interpretation

[14] Presentation slides for the ESIP COPDESS Session, July 2023: https://docs.google.com/presentation/d/1wqDQAO7QcAdqHA_pxwuFYCLWbjnhmdBa6SMAuMOJQuU/edit#slide=id.g25ad9e9fd99_0_1164 (slides 35-41)

'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.

practices were discussed, advocating for submission to domain repositories. Journal editors play an important role in setting (publication) standards and so can support the push for FAIR data publication practices. Repositories and editors will continue to meet at the domain level at Goldschmidt and other leading international conferences to increase the understanding of the geochemical data ecosystem and to fulfil journal expectations.

2.1. FAIR Guiding Principles; technical and community aspects

The FAIR Guiding Principles can be divided into principles concerning Information Technology and infrastructural implementations (‘IT’), and principles that address the data content and community practices (‘Com’) (Schultes, 2023). The two types of principles are highlighted in Table 1 using *IT* and *Com*. Principles focused on IT are generally implemented by, or are the responsibility of, the infrastructure that provides or creates data. Principles focused on data content and community involve domain-specific implementations and choices made by experts, and concern data formats, standards, schemas, vocabularies, ontologies, etc. Although alignment with these principles is primarily facilitated by domain experts, these considerations need ultimately to be converted to machine-readable forms that support FAIR machine-readable data assets. Some IT and infrastructure principles overlap with content and community considerations, and should therefore be developed/implemented in consultation with the community that the infrastructure serves. From a governance perspective, for initiatives wanting to FAIRify the geochemistry ecosystem as a whole (e.g. OneGeochemistry, universities, local/regional/global governments), this distinction may help to manage who to consult in realising the implementation of the fifteen FAIR Guiding Principles.

2.2. FAIR Enabling Resources

The individual resources supporting FAIRer data, either IT implementations or community developed, can be considered FAIR Enabling Resources (FERs) (Schultes et al., 2020). Examples of these include specific persistent identifiers and their minting services, schemas, registered vocabularies, and best practices. To satisfy this set of fifteen FAIR Guiding Principles and enable interoperability and reuse of common datasets, a community must make choices as to which set of FERs to use for addressing each particular principle. If specific resources needed to satisfy any of the principles do not yet exist or are not used by a particular community, additional resources may need to be created or adopted from other community examples.

‘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.

The FAIR Guiding Principles			
To be Findable:			
F1.	IT	Com	(meta)data are assigned a globally unique and persistent identifier
F2.	-	Com	data are described with rich metadata (defined by R1 below)
F3.	IT	-	metadata clearly and explicitly include the identifier of the data it describes
F4.	IT	-	(meta)data are registered or indexed in a searchable resource
To be Accessible:			
A1.	IT	-	(meta)data are retrievable by their identifier using a standardised communications protocol
A1.1	IT	-	the protocol is open, free, and universally implementable
A1.2	IT	Com	the protocol allows for an authentication and authorization procedure, where necessary
A2.	-	Com	metadata are accessible, even when the data are no longer available
To be Interoperable:			
I1.	IT	-	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2.	-	Com	(meta)data use vocabularies that follow FAIR principles
I3.	-	Com	(meta)data include qualified references to other (meta)data
To be Reusable:			
R1.	-	Com	meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1	-	Com	(meta)data are released with a clear and accessible data usage licence
R1.2	-	Com	(meta)data are associated with detailed provenance
R1.3	-	Com	(meta)data meet domain-relevant community standards

Table 1: The 15 FAIR Guiding Principles of Wilkinson et al. (2016) indexed as in Schultes (2023). "IT" highlights information technology and infrastructure implementations while "Com" highlights the principles related to data content and community practices.

Ideally, FERs are available as published, reusable, machine-actionable resources such as a controlled registered vocabulary (e.g. Richard et al., 2023), methodological description or published schema for data reporting for a specific method. In some cases, however, infrastructure providers may develop in-house solutions, lists and schemas without openly publishing, which limits their ability to interoperate and provide FAIR data outside of their immediate communities. FERs therefore need to be shared, published, promulgated, promoted and, where possible, endorsed by scientific unions, societies, associations, etc., to

'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.

ensure utility and enhance broader intercommunity FAIRness. Their maintenance and governance also needs to allow for further development by the community (Prent et al., 2023, p8).

Currently, community guidelines in geochemistry are mostly described and published in PDF documents that are not machine actionable. They include figures and tables denoting workflows and schemas for a particular data type (e.g. Deines et al., 2003; Flowers et al., 2022; Horstwood et al., 2016; Peng et al., 2022; Schaen et al., 2020; Walker et al., 2008). Tables often exist in these PDFs which note the desired format of data reporting, order of terms and units to be used. These 'best practice' publications can be used as the basis of templates for data reporting defined by a user community, a data repository or data-accepting database. These are not ideal for machine-to-machine interoperability but as they explicitly state which resources have been used, they are worth citing until they are themselves made FAIR-compliant. Such schema and tables are available in best practice publications (e.g. Figure 1), highlighted on the OneGeochemistry website[16] and, with time, will be developed further into machine-actionable FERs and templates for the (broader) community, that capture the essential (meta)data for reproducibility and reuse.

2023-11_Flowers-et-al-2022_(U-Th)/He-template

Checklist 1 | Checklist 2 | Checklist 3 | (U-Th)/He Reporting template | Explanation and Example

TABLE 1. CHECKLIST OF NEEDED (OR REFERENCED) SAMPLE INFORMATION AND METHODS TO REPORT FOR (U-Th)/He ANALYSES	
Sample Information	
☐	Geographic coordinates in a specified geodetic datum: Latitude, longitude, elevation
☐	Lithology
☐	Type of sample collected: In situ sample, loose rock, river sand
☐	Geologic context
☐	Any additional geochronologic or thermochronologic data available for the sample
Methods (Either describe or cite an appropriate reference)	
☐	Mineral separation procedures: type of rock disaggregation (e.g., standard crushing and grinding, SelFrag), water table, heavy liquids, drying temperature, and magnetic separation, as well as any pre-concentration techniques for detrital samples or chemical cleaning procedures.
☐	Mineral selection procedures: microscopy, magnification used, and any additional tools.
☐	Criteria used for grain selection: size, morphology, clarity, color, lack of mineral and fluid inclusions, etc.
☐	Methods for He measurements: instrumentation used, laser or furnace heating, type of He analysis (e.g., isotope dilution or external calibration).
☐	Methods for mineral dissolution and U, Th, Sm measurements: dissolution protocol, instrumentation used, type of analysis (e.g., isotope dilution).
☐	Secondary mineral standards used for quality control (e.g., Durango apatite, Fish Canyon Tuff zircon).

Figure 1. Recommendations for sample information and methodology metadata from (U–Th)/He best practice paper of Flowers et al. (2022).

'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.

2.3. FAIR Implementation Profiles

With the strong uptake of the FAIR data principles, communities are implementing a wide variety of sometimes conflicting FERs to address them. To record these choices and the combinations in which they are commonly used, and to better foster FAIR data practices, the GO FAIR community has developed a methodology comprising twenty-one questions addressing all fifteen principles (Schultes et al., 2020, summarised in Table 2 below). The collation of responses to these questions forms a FAIR Implementation Profile or 'FIP' (e.g. initial geochemistry FIP[17]). A FIP documents the implementation choices adopted by a community to enhance data FAIRness within datasets, collections, repositories, or communities. The FIP helps to identify how a dataset was made FAIR, listing which technical and community FERs were used to do so. A FIP can also be used to identify where a community has not developed or implemented resources needed to make a data instance FAIR. The community or infrastructure provider can then decide to develop or find additional resources to advance that FAIRness. As such, the ease with which a community or infrastructure provider can advance FAIRness between datasets, data collections, other communities and infrastructures depends in part on knowledge and online availability of FERs.

FIPs are key to understanding which resources or technologies are used to make a defined community's data compliant with the FAIR principles. To facilitate machine-to-machine readability of data in geochemistry, FIPs are to be generated and associated with various data levels: 1) dataset, 2) data collection and 3) data repository (discussed in Milestone 6, Prent et al., 2023, further defined in the next section). Associating a FIP to those three levels of (aggregated) data makes it clear to data users and contributors what resources were chosen to make that data FAIR. Further interoperation at the appropriate level can then be achieved by the adoption or mapping of FERs.

Principle	Question	FER types
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type
F2	Which metadata schemas do you use for findability?	Metadata schema
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism
F4	In which search engines are your metadata records indexed?	Search engines
F4	In which search engines are your datasets indexed?	Search engines
A1.1	Which standardised communication protocol do you use for metadata records?	Communication protocol

'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.

A1.1	Which standardised communication protocol do you use for datasets?	Communication protocol
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique
A2	Which metadata longevity plan do you use?	Metadata longevity
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies

Principle	Question	FER types
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema
I3	Which models, schema(s) do you use for your datasets?	Data schema
R1.1	Which usage licence do you use for your metadata records?	Data usage licence
R1.1	Which usage licence do you use for your datasets?	Data usage licence
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model

Table 2. FAIR Implementation Profile questionnaire excerpt from GO FAIR; FAIR principle definition, FER type definition and answer columns have been omitted here.[15]

2.4. FIPs and data granularity levels, community tiers and sub-disciplines

The WorldFAIR Milestone 6 report by WP05 recommended association of FIPs to each of three levels of data granularity. Here we would like to introduce a fourth tier: the disciplinary / synthesis database. From fine to coarse, the four data granularity levels defined here are **dataset, data collection, disciplinary/synthesis database, repository** (Figure 2):

[15] https://www.go-fair.org/how-to-go-fair/fair-implementationprofile/

'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. A dataset[16] is an organised collection of data or objects in a computational format, that are generated or collected by researchers in the course of their investigations, regardless of their form or method, that form the object on which researchers test a hypothesis. This includes the full range of data: raw unprocessed datasets, proprietary generated and processed data and secondary data obtained from third parties. The presentation of the data in the application is enabled through metadata.

2. A data collection[17] is a logical grouping of (research) datasets that share a common aspect or concept. Highest entity in the hierarchy of data groupings (data collection, dataset, row or record in a dataset). Comprises a grouping of datasets that have a strong connection and it is organised coherently around a single element or concept such as a model or instrument.
3. A disciplinary or synthesis database represents a source for a large collection of data from the same/or similar types with the main purpose to make the data (values) discoverable across many publications. Examples for synthesis databases for geochemical data are PetDB and GEOROC. They represent sources for geochemical measurements derived from hundreds of research articles that were retrieved from the articles or, more recently, ingested directly through data templates from DOI-referenced data publications. Queries in these synthesis databases can be made, e.g., for the value of a specific property of the data (e.g. the Fe_2O_3 content of mid-ocean ridge basalts or any other major or minor elements) across the entire database. These synthesis databases are essential for regional or global data analyses and to gain a more complete picture of the Earth's geochemistry.
4. A (research) data repository; Repositories preserve, manage, and provide access to many types of digital materials in a variety of formats. Materials in online repositories are curated to enable search, discovery, and reuse. There must be sufficient control for the digital material to be authentic, reliable, accessible and usable on a continuing basis.
    a. They, especially domain repositories, collect rich metadata (including PIDs like ORCID, ROR, Grants, Funders) that support data discovery on the dataset or data collection level. A query on the data value level, as described above in the synthesis database, is often not intended. However, data repositories can play

[16] CODATA RDM Terminology (2023 version): https://doi.org/10.5281/zenodo.10626170
[17] CODATA RDM Terminology (2023 version): https://doi.org/10.5281/zenodo.10626170

'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.

an important role in the promotion of standards when they request, e.g., to use existing data templates, like the EarthChem Templates (EarthChem Team, 2022) for geochemical data publications. They also play an important role in the whole publication process; as a consequence of the work of the Coalition of Publishing Data in the Earth and Space Sciences (COPDESS), more and more research journals require the publication (with DOI) of data underlying the scientific outcomes described in research articles as a condition to publish the articles and allow data citations in the reference sections (e.g. Hanson et al. 2015, Stall et al, 2018). The publishers also require that a dataset be lodged in a trusted repository and linked to the paper through the DOI.

Similarly, a tiered approach of publication was suggested in WorldFAIR Milestone 6 (Prent et al., 2023) to achieve convergence on interoperability for vocabularies. The tiered approach is based on the extent of the community that defined the resource. Introducing and defining the four community tiers; **local, regional, international and global** (Figure 2) as:

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.

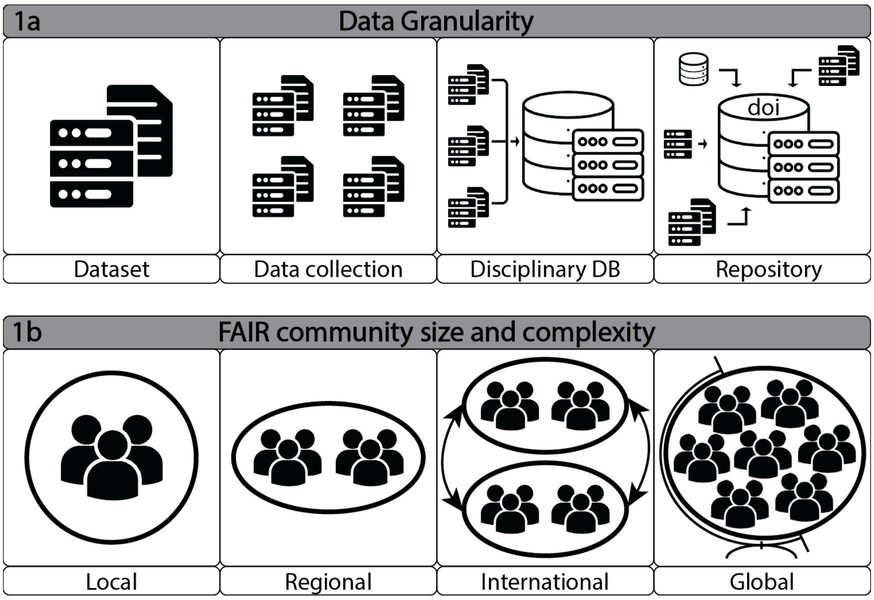

Figure 2. FAIR evolution pathways.

For data to be FAIR within a community, members must be able to find, access, interoperate and reuse data within that community. However, intra-community FAIR compliance does not necessarily mean that this data is FAIR for those outside that community. To enable broader intercommunity access and use of FAIR data - for example, from local to regional - a local community needs to specify and publicise the FERs used to make their data FAIR.

This will enable a full understanding of the data and metadata and clarify the annotation and vocabularies used. In publicising FERs, a local community can facilitate FAIR data reuse for a broader community at the regional, international or even global level. Multiple local groups can converge local standards into ones adopted more regionally to globally, or build crosswalks and/or mappings for them to enable broader interoperability. In this way, a community can provide FAIR data and facilitate machine-to-machine interoperability by articulating their chosen FERs in a completed FIP irrespective of how broad their community coverage is.

Building on these ideas, data from sub-disciplines can be tailored to interoperate with other sub-disciplines. The many sub-disciplines in geochemistry collect diverse suites of data and different data types, structures and terminology; they are thus not naturally interoperable. Yet

'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.

differing sub-disciplines are commonly combined to form coherent interpretations (see Box 1), with datasets connected by the sample or geolocation level.

2.5. FIPs as tools towards interoperability and convergence

Fundamentally, a FIP can be used to describe precisely what resources were used to enable each FAIR principle. However, the number of FERs required increases with the complexity of a data collection or data repository, and FIPs that apply to a coarser level of data granularity (e.g. repository) will note more FERs per FAIR Principle unless all the data of finer granularity in that coarse level is completely standardised following a single set of FERs. It is therefore beneficial for users of data publishing infrastructure to publish a FIP that applies to a particular data granularity level when making data available, in order to enable interoperability across different granularity levels of data between different communities. The availability of a FIP with the appropriate data granularity level then facilitates interoperability and convergence. It is important that each FIP has a DOI and is version-controlled and linked to the relevant dataset/data collection as a Related Identifier.

The availability of FIPs and their generation serves a dual purpose: it informs data consumers on the implementation choices and resources used to make data FAIR, and it guides data-producing communities on identifying and creating missing resources to enhance data FAIRness.

'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.

Box 1: Sub-disciplinary, local tier geochemistry interoperability

An example where interoperability between sub-disciplines has been achieved is with Lu/Hf and U–Pb geochronology. The mantle extraction age calculation from Lu/Hf data depends on the interpreted crystallisation age obtained from U–Pb analyses. Interoperability between these sub-domains is achieved across a different level of data interpretation, i.e. the interpreted age from U–Pb analysis (level 2 data; **Table 3**) is used as input to calculate mantle extraction ages based on Lu/Hf analyses (going from a Lu/Hf level 1 to level 2 data; see Boone et al., 2022 for an explanation on data interpretation levels).

Data type	Data level	Description
Sample	Sample	Sample information (e.g., IGSN, Sample ID, location info, lithology, etc.)
Raw	Level 0	Unprocessed instrumental data and metadata (digital imagery, raw mass spectrometry counts, etc.)
Hard	Level 1	Reduced or analysed raw data (e.g., calculated isotopic ratios, chemical concentrations, digital image analyses, etc.)
	Level 2	Calculated parameters and ages using Level 1 data
Soft	Level 3	Derivative models determined using Level 1 and/or 2 data (e.g., thermal history models)
	Level 4	Derivative models combining/interpolating Level 1, 2 and/or 3 data over an area (e.g., spatial interpolations, regional cooling/heating maps)

Table 3. Data level type terminologies after Boone et al. (2022) adapted from NASA's Data Processing Level hierarchy (NASA 2021).

3. Making geochemical data FAIR

Geochemistry is an important research discipline that helps progress our understanding of the evolution of Earth and planetary systems. Geochemistry research supports societal needs as described in the the UN Sustainable Development Goals (SDGs)[18], particularly addressing clean water and sanitation (SDG 6), affordable and clean energy (SDG 7), decent work and economic growth (SDG 8), industry, innovation and infrastructure (SDG 9), climate action (SDG 13), and life on land (SDG 15). Drawn from complex methodologies and processes, geochemical data are a primary resource, often painstakingly generated, and as such highly valuable. As interpretations can change based on additional data availability or the development of new

[18] The UN Sustainable Development Goals: https://sdgs.un.org/goals

'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.

algorithms to process the same data, the generation of reusable data or facilitating reusability of data is paramount.

3.1. Data from a long-tail community and the complexity of making them FAIR

In the pursuit to better understand the geology of Earth and planetary systems, new methods and techniques to analyse element and isotope compositions are constantly being developed and refined. Data are generated using various techniques at different scales, from continental or even planetary scales, to sub-micron scale. The resulting geochemical data are as diverse in their structure and notation as they are in content: major elements, trace elements, isotopes of elements, U–Pb, Fission Track – helium (FT/He), etc. This growing diversity and heterogeneity means that it has been difficult for the geochemistry community to standardise and make their data FAIR, particularly regarding machine-to-machine interoperability.

With this large variety of data types comes a similar number of communities and sub-disciplines. Individual groups and laboratories often develop in-house data processing workflows that are maintained locally, causing regional or international groups to be divided over the best approach to achieve reproducibility and standardisation. Combined with the "long-tail" nature of geochemical data, many separate data silos exist at local levels. Introducing and operationalising the FAIR data principles therefore proves to be an enormous challenge, and the capacity to reproduce, replicate or even reuse datasets is increasingly a community issue.

Making geochemistry data FAIR requires each community associated with a data type/method/technique to determine the key minimum variables that make each data type reproducible and interoperable. Such minimum variables and practices do exist, but most are published as or part of research articles and/or PDFs that contain tables and prescribe what needs to be done to the data to become FAIR. These resources need to be further developed to become machine-actionable FERs. These FERs, whether existing or newly developed, will need to be implemented using FAIR processes and used with corresponding datasets, data collections and repositories.

'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.

3.2. Using FIPs to facilitate steps towards FAIR geochemistry data

To achieve FAIR data across multiple geochemical datasets, communities and sub-disciplines, it is essential for data to be understood at a core level. A FIP is an excellent tool that firstly provides a standardised method for geochemists to provide machine-readable information on how a specific dataset is described and compiled. Secondly, when FIPs become available in the community they can serve as a guide to understanding the scientific resources and technologies utilised by specific geochemical communities. FIPs can offer insights into the FAIR evolution of a dataset or a community. FIPs can also serve as a tool for data producers and data providers (infrastructures), who can refer to a specific dataset's FIP to provide the information required to understand and use the dataset. Ideally, a FIP should be published with a persistent identifier (e.g., DOI) and be linked to the dataset and vice versa (e.g., using the "related identifier" property of the DataCite Metadata Schema). Declaring resources used to FAIRify data in FIPs will support the standardisation of how groups can become aware of these resources. There is a need to generate FIPs for data of each level of data granularity (as in Figure 2).

Comparing FIPs associated with the same or different levels of data granularity can be beneficial. This comparison can help to identify overlapping, redundant or different usages of FERs in use for data FAIRification. This can highlight where there are differences between FERs (e.g., metadata profiles, vocabularies, ontologies, formats), and define where crosswalks need to be developed and published, or other techniques utilised to enable interoperability (e.g., SSSOM – Simple Standard for Sharing Ontological Mappings; Matentzoglu et al., 2022).

Open publication and wide promulgation of available resources that enhance interoperability (such as agreed minimum variables, formats, vocabularies and ontologies) would support the geochemical communities in adopting the same FER, resulting in larger communities capable of interoperating, and ultimately enabling international networks of standardised machine-readable data. However, it can take years, if not decades, of dedicated collaboration to develop agreed, respected and endorsed standards. Currently, due to the increasing necessity for making geochemical data FAIR from major funders and other drivers (including the U. S. National Science Foundation, the Australian Research Council, the European Commission, the European Open Science Cloud, the German Research Foundation (DFG), industries, and industry groups such as the Open Forum Group, among others), multiple standards and vocabularies are being published online. While most are produced by local

'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.

groups, others are developed and published through larger research infrastructure projects (e.g., AuScope, EPOS, EarthChem). However, only a few are developed by collaborations involving various national research groups and projects and there is no agreed single international authority to endorse any as a globally agreed standard.

Given that new technologies (e.g., the surge in uptake of AI and ML) are creating an urgent need for FAIR machine-actionable data, Milestone 6 suggested that vocabularies (a type example of a FER), alongside other types of FERs required for enabling interoperability of data, should always be published even as a representation of a local level FER, and be time-stamped and version controlled. From such local publications, encouragement to collaborate regionally and internationally on the convergence of FERs will further the interoperability of datasets, data collections and repositories. For example, a close collaboration is underway to make interoperable the DIGIS[19], EarthChem[20] and Astromaterials[21] data systems. The collaboration includes the comparison of infrastructure FIPs and assesses each of the system's FERs and resource solutions for further convergence (Klöcking et al., 2023, Richard et al., 2023). Ultimately, when an agreed Global Geochemical Standards community is identified, the various local and community (regional – international) resources can be harmonised and endorsed at the global level, for actionability by both humans and machines (Prent et al., 2023).

Publicising and promulgating vocabularies as well as other FERs developed at each and any of the community tiers (local, regional, international, global) should be encouraged. Vocabularies define what terms and definitions go with each data set making it more FAIR. Where possible, scientific unions and associations, and where relevant government agencies, should encourage harmonisation of these local vocabularies into international and ultimately global vocabularies that are authoritatively endorsed to stimulate sharing and reduce time spent preparing data for (re)analysis and/or reuse. Besides generation and publication of these resources, the promulgation of their existence will help enable the interoperability of geochemical data. With such resources openly available, sub-communities can extract profiles consisting of multiple sources that satisfy their needs without recreating entire vocabularies and removing the need to publish, curate and maintain their own resources.

[19] DIGIS: https://www.uni-goettingen.de/en/643369.html
[20] EarthChem: https://www.earthchem.org/
[21] Astromaterials: https://www.astromat.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.

Creation and publication of FERs for geochemical data should thus be encouraged at all of the tiers: local, regional, international and global.

4. Recommendations

The work leading up to this report has explored geochemical best practices and pathways to make the geochemistry data community more compliant with the FAIR principles. We have found the generation and use of FIPs an effective tool to pinpoint commonalities between existing systems and highlight where the community can improve.

This report and the WorldFAIR Milestone 6 report have the following recommendations:

4.1. Policy 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:

1. **Reference the size of the community that is creating and using FAIR data infrastructure.** Use the terms 'local', 'regional', 'international' and 'global' to provide information as to the extent and inclusivity of the community. As FAIRness can be achieved within a community, the size of that community should be clearly communicated, as follows:
    a. A local community consists of a single laboratory or university.
    b. A regional community is within a city or between cities and universities, multiple laboratories or the community of a discipline within a country.
    c. An international community acts as a collaboration between countries.
    d. A global community acts as one with all of the community involved.
2. **Publish online and make FAIR any resource used to enable compliance with any of the FAIR Principles.** A community that uses any resources that describe the data they produce and that support making these data FAIR, should publish and make these resources available as persistent online resources, each with its own PID. It is essential to publish any resource regardless as to whether this is at the 'local', 'regional',

'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.

'international' or 'global' level in order to drive convergence towards FAIR. Publication of resources as FAIR Enabling Resources (FERs) is a key step towards this goal.

3. **Generate FIPs as a means of declaring which FERs are used to implement each of the FAIR Guiding Principles.** Data-providing and data-generating infrastructures should generate FIPs as a way to clarify the FERs they use to make datasets, data collections and the data infrastructure (repository / database) FAIR. This is a systematic way to provide specific information about the data in a standardised manner with links to the exact resource that is used.

4. **Publish individual FIPs at any data granularity tier as version-controlled resources, each with a PID.** Data infrastructures should publish FIPs as version-controlled resources with PIDs that can be linked to the dataset, data collection(s) or repository to which they apply. These FIPs can serve publishers and funders in order to assess the FAIR maturity level of any dataset that publishers or funders are reviewing or funding. If included as a linked resource in the metadata of data published in generic repositories, the FIP can provide domain-specific information about that data asset, which generic repositories are currently unable to provide.

5. **Use FIPs as a means of assessing FAIR maturity of and between data assets.** Evaluating and comparing FIPs of data assets, particularly for whether each principle has been addressed and for what FER implementation was used, can be used to assess how FAIR such assets are. Community uptake, technical advancement and maturity of the implemented FER plays an important role in the FAIR maturity evaluation too.

6. **Use FIPs as a means to develop and publish machine-actionable crosswalks between 'conflicting' resources**. The systematic approach of creating FIPs can be used to standardise comparison of FERs in use between data repositories and inform the decision-making process of how to become FAIRer across these data assets. The comparison of one FIP to another can help decide if a crosswalk between FERs needs to be developed or if mapping between resources can suffice.

7. **Compare and evaluate FAIR data enabling resources of larger communities, complexer data aggregations and where used to connect (sub)disciplines.** At a technical, abstract level the geochemical community and the data community in general should start to think about FAIR applying to different sizes or complexity of a 'community' (local, regional, international, global), complexity or size of 'storage' (dataset – data collection – database), and 'discipline' (single, dual, multi). For example interoperability between (sub)disciplines is likely only possible when data of higher abstraction levels are integrated.

'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.

4.2. Organisational recommendations

Stakeholders: OneGeochemistry WorldFAIR Work Package, OneGeochemistry Initiative.

Recommendations:

1. **Produce a reference/catalogue FIP for the WorldFAIR Geochemistry Work Package.** A reference/catalogue FIP can serve dual purposes: a) convergence through providing a catalogue from which new 'builders' of repositories can choose FERs; and b) an overview that can compare used resources and choose where mappings/crosswalks should be developed with highest impact.
2. **Socialise and promulgate policy recommendations to geochemistry stakeholders.** The policy recommendations and this report should be socialised and promulgated within the geochemistry community with special attention to sharing them with data producers, data publishers, data/informatics specialists, digital architects, software developers and team leads, data- or software-focused researchers, (research) data managers.

'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.

References

Boone, S. C., Dalton, H., Prent, A., Kohlmann, F., Theile, M., Gréau, Y., … Rawling, T. (2022). AusGeochem: An open platform for geochemical data preservation, dissemination and synthesis. Geostandards and Geoanalytical Research, 46(2), 245-259. https://doi.org/10.1111/ggr.12419

CODATA RDM Terminology Working Group (2024). CODATA RDM Terminology (2023 version): overview (Version 2023). Zenodo. https://doi.org/10.5281/zenodo.10626170

Cox, S. J. D., Gonzalez-Beltran, A. N., Magagna, B., Marinescu, M. C. (2021). Ten simple rules for making a vocabulary FAIR. *PLOS Computational Biology*, *17*(6). https://doi.org/10.1371/journal.pcbi.1009041

Deines, P., Goldstein, S.L., Oelkers, E.H., Rudnick, R.L., and Walter, L.M. (2003). Standards for publication of isotope ratio and chemical data in chemical geology. Chem. Geol.202, 1–4. https://doi.org/10.1016/j.chemgeo.2003.08.003.

Devaraju, A., Mokrane, M., Cepinskas, L., Huber, R., Herterich, P., de Vries, J., Akerman, V., L'Hours, H., Davidson, J., Diepenbroek, M. (2021). From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. Data Science Journal, 20: 4, pp. 1–14. https://doi.org/10.5334/dsj-2021-004

EarthChem Team (2022). Bulk Analysis Data Template v. 3.1 (Version 1.0) [Data set]. Interdisciplinary Earth Data Alliance (IEDA). https://doi.org/10.26022/IEDA/112263

Flowers, R., Zeitler, P., Danišík, M., Reiners, P., Gautheron, C., Ketcham, R., Metcalf, J., Stockli, D., Enkelmann, E., Brown, R., 2022. (U-Th)/He chronology: Part 1. Data, uncertainty, and reporting. Geol. Soc. Am. Bull. 135, 104–136. https://doi.org/10.1130/b36266.1.

Gregory, A., Hodson, S. (2022). WorldFAIR Project (D2.1) 'FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?'. https://doi.org/10.5281/ZENODO.7378109

Hanson, B., Lehnert, K. L., Cutcher-Gershenfeld, J. (2015). Committing to Publishing Data in the Earth and Space Sciences. In Eos (Vol. 96). https://doi.org/10.1029/2015eo022207

Hodson, S., Gregory, A. (2023). WorldFAIR Project (D1.3) First policy brief (Version 1). Zenodo. https://doi.org/10.5281/zenodo.7853170

Horstwood, M.S.A., Košler, J., Gehrels, G., Jackson, S.E., McLean, N.M., Paton, C., Pearson, N.J., Sircombe, K., Sylvester, P., Vermeesch, P., Bowring, J.F., Condon, D.J., Schoene, B. (2016). Community-derived standards for LA-ICP-MS U-(Th-)Pb geochronology – uncertainty propagation, age interpretation and data reporting. Geostand. Geoanal. Res. 40, 311–332. https://doi.org/10.1111/j.1751-908x.2016.00379.x.

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

'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.

Klöcking, M., Wyborn, L., Lehnert, K. A., Ware, B., Prent, A. M., Profeta, L., Kohlmann, F., Noble, W., Bruno, I., Lambart, S., Ananuer, H., Barber, N. D., Becker, H., Brodbeck, M., Deng, H., Deng, K., Elger, K., De Souza Franco, G., Gao, Y., … Zhou, T. (2023). Community recommendations for geochemical data, services and analytical capabilities in the 21st century. *Geochimica et Cosmochimica Acta*, *351*, 192–205. https://doi.org/10.1016/j.gca.2023.04.024

Krans, N. A., Ammar, A., Nymark, P., Willighagen, E. L., Bakker, M. I., Quik, J. T. K. (2022). FAIR assessment tools: Evaluating use and performance. NanoImpact, 27, 100402. https://doi.org/10.1016/j.impact.2022.100402

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

NASA (2021) Data processing levels. Accessed 31 January 2024. https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels

Peng, G., Lacagnina, C., Downs, R.R., Ganske, A., Ramapriyan, H.K., Ivánová, I., Wyborn, L., Jones, D., Bastin, L., lin Shie, C., Moroni, D.F., 2022. Global community guidelines for documenting, sharing, and reusing quality information of individual digital datasets. Data Sci. J. 21, 8. https://doi.org/10.5334/dsj-2022-008.

Peters-von Gehlen, K. P., Höck, H., Fast, A., Heydebreck, D., Lammert, A., Thiemann, H. (2022). Recommendations for Discipline-Specific FAIRness Evaluation Derived from Applying an Ensemble of Evaluation Tools. Data Science Journal, 21, 7. https://doi.org/10.5334/dsj-2022-007

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

Richard S.M., Lehnert, K. A., Klöcking, M., Luzi-Helbing, M., Johansson, A. Profeta, L. (2023) Analytical methods for geochemistry and cosmochemistry, (2023) Astromaterials Data System. https://vocabs.ardc.edu.au/viewById/650

Schaen, A.J., Jicha, B.R., Hodges, K.V., Vermeesch, P., Stelten, M.E., Mercer, C.M., Phillips, D., Rivera, T.A., Jourdan, F., Matchan, E.L., Hemming, S.R., Morgan, L.E., Kelley, S.P., Cassata, W.S., Heizler, M.T., Vasconcelos, P.M., Benowitz, J.A., Koppers, A.A., Mark, D.F., Niespolo, E.M., Sprain, C.J., Hames, W.E., Kuiper, K.F., Turrin, B.D., Renne, P.R., Ross, J., Nomade, S., Guillou, H., Webb, L.E., Cohen, B.A., Calvert, A.T., Joyce, N., Ganerød, M., Wijbrans, J., Ishizuka, O., He, H., Ramirez, A., Pfänder, J.A., Lopez-Martínez, M., Qiu, H., Singer, B.S., 2020. Interpreting and reporting 40Ar/39Ar geochronologic data. Geol. Soc. Am. Bull. 133, 461–487. https://doi.org/10.1130/b35560.1.

'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.

Schultes, E. (2023). The FAIR hourglass: A framework for FAIR implementation. FAIR Connect, 1(1), 13–17. https://doi.org/10.3233/FC-221514

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

Stall, S., Yarmey, L., Boehm, R., Cousijn, H., Cruse, P., Cutcher-Gershenfeld, J., Dasler, R., de Waard, A., Duerr, R., Elger, K., Fenner, M., Glaves, H., Hanson, B., Hausman, J., Heber, J., Hills, D., Hoebelheinrich, N., Hou, S., Kinkade, D., … Wyborn, L. (2018). Advancing FAIR Data in Earth, Space, and Environmental Science. In Eos (Vol. 99). https://doi.org/10.1029/2018eo109301

Walker, D.J., Condon, D., Thompson, W., Renne, P., Koppers, A., Hodges, K., Reiners, P., Stockli, D., Schmitz, M., Bowring, S., Gehrels, G., 2008. Geochron Workshop Reports Sponsored by EarthChem and EARTHTIME. https://doi.org/10.5281/ZENODO.4313859.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., 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 1: The FAIR Guiding Principles interpreted following Jacobsen et al (2020).

1. Findable

Findable 1

Principle F1 underscores the importance of assigning digital resources, such as data and metadata, a globally unique and persistent identifier. This identifier must serve as a permanent machine-readable reference. "Globally unique" means the identifier must unambiguously refer to the intended resources on a universal scale, not just within a local context like a single database. "Persistence" requires that this identifier is never reused, even if the resource it identifies no longer exists or moves to a different digital environment. The GO FAIR Foundation also supports "resolvability," ensuring that the identifier can be linked to the object it represents in a large-scale automated environment. Furthermore, predictable identifier resolution behaviour is expected, allowing consistent resolution across multiple requests. In essence, the GO FAIR Foundation promotes the use of Globally Unique, Persistent, and Resolvable Identifiers (GUPRIs) for FAIR data implementations.

Findable 2

Principle F2 complements F1 by emphasising the discoverability of digital resources. To achieve this, these resources must be described with rich metadata, including detailed content descriptors linked to their unique identifiers. While defining the minimum metadata richness can be challenging, more comprehensive metadata, suitable for both humans and computers, greatly enhances the resource's findability in refined searches. Descriptive metadata plays a crucial role in cross-domain search and interdisciplinary use cases. To enable discoverability, the essential metadata requirements for citations are creator, title, publication date, publisher, and identifier, as specified in various documents. Additional core metadata requirements for data discovery may also be considered. This principle is primarily concerned with making well-described digital resources easily discoverable, even when the user is unfamiliar with the resource. It encourages data providers and domain experts to support various search facets and provide both generic and domain-specific descriptors for global and local search engines to locate the resource effectively.

Findable 3

Principle F3 highlights the importance of linking metadata clearly and explicitly to the digital resources it describes via a globally unique and persistent identifier (GUPRI), including it in the

'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.

metadata record. Consistent linkage of metadata to resources is key, since there are various methods used in the community to declare this relationship it is difficult for humans and machines to identify the described resource from the metadata record and consistency can be improved.

Findable 4

Principle F4 states that digital resources must be registered or indexed in a searchable resource (e.g., a search engine). The searchable resource provides the infrastructure by which a metadata record (made accessible with a GUPRI, F1) can be discovered, using either the attributes in that metadata (F2) or via the identifier of the resource itself (F3).

2. Accessible

Accessible 1

Principle A1 underscores the importance of enabling efficient access to digital resources. It emphasises that the access protocol used to retrieve a resource should not present any additional barriers when permitted access is granted. This means that whether a machine or human agent is retrieving the record, the identifier should follow a globally-accepted schema tied to a standardised communication protocol. While FAIR encourages mechanised access, it doesn't rule out non-mechanized access, as long as the process is well-described and a machine can identify the next steps, even if it involves human intervention. In such cases, the identifier must still unambiguously indicate the requested record. The emphasis is on the standardised communication protocol, which provides a consistent way for agents to access a resource, whether access is automated or facilitated by human actions, such as sending access requests via email or phone.

Accessible 1.1

Principle A1.1 emphasises that the protocol used to access a digital resource, like querying it, should not create bottlenecks or obstacles. This concept primarily concerns the process of accessing the resource and isn't directly related to restrictions on using the resource itself. The ideal protocols, like HTTP for the World-Wide Web, should be open, free, and universally implementable. These characteristics reduce the cost of accessing digital resources by being well-defined and open, allowing anyone to create their own standards-compliant

'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.

implementations. The accessibility of protocol specifications for free ensures that even those with limited financial resources can access and implement them without incurring monetary obligations. Additionally, the universality of these protocols ensures that the technology is available to all, without restrictions based on location or sub-communities, embodying both the 'gratis' (free of cost) and 'libre' (freedom to use) aspects of 'free.'

Accessible 1.2

Principle A1.2 highlights that adhering to the FAIR guiding principles does not mean that all data must be 'open' and accessible to everyone. Some digital resources, due to ethical, legal, or contractual constraints, require additional steps to access. These steps typically involve verifying the identity of the access requester (authentication), ensuring their profile and credentials align with the resource's access conditions (authorization), and confirming that the intended use aligns with permitted use cases. Additionally, when technical protocols do not already define authentication and authorization, they should be specified. Requesters can be either humans or machine agents, with machines often acting as proxies for human users. The principle mandates that a FAIR resource must offer such a protocol, but the specifics of the protocol are not further specified. In practice, an Internet of FAIR Data and Services cannot operate effectively without implementing Authentication and Authorization Infrastructure, including Authentication and Authorization for machines, which should be Ontology-based and machine actionable.

Accessible 2

Principle A2 emphasises the importance of ensuring continued access to relevant digital resources in the future. There are situations where data may become inaccessible either intentionally (due to limited resources or legal requirements) or accidentally. However, because these data may have been used and referenced by others, it's crucial that consumers, including machines, can access high-quality, machine-readable metadata that describes these resources in a way that allows a basic understanding of their nature and source, even when the data itself is no longer available. This principle heavily relies on the "second purpose" of principle F3, which states that metadata records should contain the identifier of the data. In cases where the data is no longer accessible, there must be a clear and precise way to find its historical metadata record. This idea of ensuring accessibility even when data is no longer available is further explained in the Joint Declaration of Data Citation Principles.

'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. Interoperable

Interoperable 1

Principle I1 addresses the challenge of making digital resources understandable and usable by both humans and machines. Often, the lack of clear and unambiguous content descriptors, particularly in non-machine-interpretable data formats like tables or generic XML, leads to difficulties in understanding and combining digital resources. While community-defined data exchange formats work well within specific contexts, they often fail to enable efficient interoperation and integration beyond those contexts, making it a costly and sometimes impossible task. This not only hampers human efforts but also prevents machines from effectively utilising digital resources, which goes against the core goal of the FAIR principles.

To address this, Principle I1 emphasises the need for a globally understood "language" for machines to achieve a common understanding of digital resources. This language, in the form of a machine-readable "grammar," defines entities and their relationships within the data's structure. By using such a language, producers of digital resources enable mechanised interpretation, allowing machines to grasp the information's content to some extent. This is a crucial step toward creating a common understanding of digital resources for machines, which is essential for building a functional Internet of FAIR Data and Services. Various technologies can be employed to implement Principle I1, enhancing the overall usability and interoperability of digital resources.

Interoperable 2

Principle I2 addresses the importance of using structured vocabularies in the FAIR data principles to represent concepts within a domain clearly. While shared and structured vocabularies are crucial, they should also be FAIR in their own right. This means that the vocabularies used for metadata or data need to be findable, accessible, interoperable, and reusable so that users, including machines, can fully understand the meaning of the terms used in the metadata.

The principle acknowledges that merely using a label, such as "temperature," is insufficient for a machine to understand its context and intent (e.g., body temperature or melting temperature) or how it should be linked to other similarly labelled data. To address this, I2 requires that vocabulary terms used in knowledge representation languages (as per Principle

'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.

I1) must be machine-distinguishable, allowing machines to resolve their intended meanings. This ensures that machines can detect and prevent both "false agreements" and "false disagreements" regarding the exact meaning of the identifier, enhancing data understanding and interoperability.

Interoperable 3

Principle I3 emphasises the importance of connecting data and metadata to create a meaningful network of interlinked resources. A "qualified reference" is a reference to another resource that includes a clear specification of the relationship between them. For example, when dealing with metadata files, it's valuable to link to prior or next versions using specific relationships like "prior version" or "next version," following community standards that align with FAIR principles.

In the context of data, consider a dataset that contains information about global city populations. To be FAIR, it should include links to related resources such as city data from Wikidata, geographical and geospatial data, or other domain-specific resources generated by those cities. These links should be qualified references with meaningful and interpretable relationships.

It's also important to note that various metadata files or containers, as individual FAIR digital resources, can point to the same target object, such as a dataset or workflow. For instance, a FAIR Digital Object, like a nanopublication, can have intrinsic metadata, provenance metadata, and "secondary" metadata created by reusers, all describing the same digital resource from different perspectives. This principle also highlights the need to distinguish clearly between metadata files or containers and the resources they describe.

4. Reusable

Reusable 1

Principle R1 may seem similar to F2 at first, but there's a key distinction. F2 is about enabling effective attribute-based search, while R1 focuses on helping both machines and humans assess if a discovered resource is suitable for a specific task. For instance, not all gene expression data related to a particular gene are relevant to a study on heat stress effects. R1 addresses the ability to evaluate and filter discovered data based on its suitability for a

'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.

particular purpose. This highlights the importance of data stewards considering not only high-level metadata for generic searches but also detailed metadata that offers operational instructions for reusability.

In this context, a broad range of factors may be necessary to determine a resource's suitability for inclusion in an analysis and how to process it effectively. The term "plurality" suggests that metadata authors should provide as much metadata as possible to support a wide range of use-cases and agent needs, without making narrow assumptions about secondary consumers. The sub-principles R1.1, R1.2, and R1.3 further specify critical attribute types that contribute to R1.

Reusable 1.1

Principle R1.1 states that all digital resources and their metadata must include a licence that defines the conditions for their use, even if it's an "unconditional" permission. This is essential to provide legal clarity for resource use. It's crucial that the licence is discoverable by agents; otherwise, it's as good as not having a licence at all. Additionally, the licence for data resources may differ from that of their metadata, impacting how metadata is indexed and found. This underscores the importance of separating and providing permanent links for data and metadata. It's essentially a clear declaration of public domain status or equivalent terms of use for digital operations, like smart contracts. The absence of a licence doesn't imply "open" but rather creates legal uncertainty that can discourage or legally prevent reuse. Combinations of resources with varying licence conditions can have adverse effects and may hinder their use for specific purposes. To promote reuse, the chosen licence should be as open as possible.

Reusable 1.2

Principle R1.2 emphasises the importance of detailed provenance information, which includes aspects like how a resource was created, why it was created, who created it, the conditions under which it was created, the starting data or source resource, the funding or resources used, data ownership, attribution, and any post-generation cleaning processes. Provenance details assist both humans and machines in evaluating if a resource aligns with their criteria for reuse and what data manipulation procedures may be required for appropriate reuse.

'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.

Reusable 1.3

Principle R1.3 highlights the importance of adhering to community standards and best practices for data archiving and sharing. Many research communities have established Minimal Information Standards that outline the essential metadata needed to evaluate data quality, support reproducibility, and enable reuse. While these standards are a good beginning, it's important to recognize that achieving true interdisciplinary reusability typically demands more extensive metadata. The level of detail required in provenance metadata depends on the norms established in closely related research communities. These standards provide a foundation for effective data sharing and reuse. For a list of such standards, FAIRsharing is a valuable resource to consult.

'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.