frontiers in Earth Science

PERSPECTIVE
published: 31 January 2020
doi: 10.3389/feart.2020.00003

Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures

Daniele Bailo[1], Rossana Paciello[1], Manuela Sbarra[2], Riccardo Rabissoni[2], Valerio Vinciarelli[2] and Massimo Cocco[1]*

[1] European Plate Observing System, ERIC-Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy, [2] Istituto Nazionale di Geofisica e Vulcanologia, Seismology and Tectonophysics, Rome, Italy

FAIR principles have become reference criteria for promoting and evaluating openness of scientific data and for improving datasets Findability, Accessibility, Interoperability, and Reusability. This also applies to Research Infrastructures (RIs) in the solid Earth domain committed to provide access to seismological data, ground deformations inferred from terrestrial, and satellite observations, geological maps, and laboratory experiments. Such RIs have been indeed committed for a long time, well before the appearance of FAIR principles, to engage scientific communities involved in data collection, standardization, and quality control as well as in implementing metadata and services for qualification, storage and accessibility. By addressing open science and managing scientific data, they are working to adopt FAIR principles, thus having the onerous task of turning these principles into practices. In this work we argue that although FAIR principles have the merit of creating a common background of knowledge to engage communities in providing data in a standard way thus easing interoperability and data sharing, in order to make the adoption of FAIR principles less onerous there is an urgent need of clear models, reference architectures and technical guidelines which can support RI implementers in the realization of FAIR data provision systems. We therefore discuss the state of the art of FAIR principles ecosystem and open new perspectives by discussing a four-stages roadmap that reorganizes FAIR principles in a way that better fits to the approach of RI implementers, and a FAIR adoption process that relates FAIR principles to technologies for their implementation.

Keywords: research infrastructure, FAIR principles, solid Earth science, multidisciplinary data, system interoperability

OPEN ACCESS

Edited by:
Susanne Buiter,
Geological Survey of Norway, Norway

Reviewed by:
Juliane Dannberg,
University of Florida, United States
Matthias Rosenau,
German Research Centre for Geosciences, Helmholtz Centre Potsdam, Germany

***Correspondence:**
Daniele Bailo
daniele.bailo@ingv.it;
bailo.daniele@gmail.com

Specialty section:
This article was submitted to Solid Earth Geophysics, a section of the journal Frontiers in Earth Science

Received: *10 October 2019*
Accepted: *10 January 2020*
Published: *31 January 2020*

Citation:
Bailo D, Paciello R, Sbarra M, Rabissoni R, Vinciarelli V and Cocco M (2020) Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures. Front. Earth Sci. 8:3. doi: 10.3389/feart.2020.00003

INTRODUCTION

FAIR (Findable, Accessible, Interoperable, Reusable) data principles (Wilkinson et al., 2016) are gaining consensus within scientific communities, fostered by the participation in designing pan-European initiatives such as the European Open Science Cloud (EOSC) (EOSC, 2018), where they are used as driving concepts to support data interoperability among standardized repositories compliant to a shared set of requirements (Mons et al., 2017).

FAIR principles were discussed and launched at a FORCE11[1] workshop in 2014, enabling the publication of a first paper with their detailed definition (Wilkinson et al., 2016).

EOSC initiative subsequently emphasized how these principles just provide guidelines, without including any technical requirement nor explicitly suggesting technologies for their adoption (Mons et al., 2017), which may lead to ambiguities regarding their implementation; this led proponents already involved in their definition to establish a FAIR Metrics group for measuring FAIRness of data (Wilkinson et al., 2018). Such an approach, however, does not consider the impact of FAIRness on the full research data lifecycle and on the related development activities (Boeckhout et al., 2018).

On the same page, initiatives have been undertaken to provide support for FAIR principles adoption: FAIRsharing[2] provides a platform for enhancing the discoverability of resources even outside of a community (Sansone et al., 2019); GO-FAIR[3] offers supporting materials and tools (Schultes et al., 2018); Enabling FAIR data project[4] promotes FAIR principles implementation with the document "Commitment Statement to Enabling FAIR Data in the Earth, Space, and Environmental Sciences" (E. F. D. Community, 2018). In parallel, further work is carried out by the Research Data Alliance (RDA[5]) "FAIR Data Maturity Model WG Case Statement" Interest Group for establishing core criteria to assess the implementation level of FAIR data principles (FAIR Data Maturity Model WG, 2019).

In this framework, Environmental Research Infrastructures (ENVRI) have set up a joint 4 years program, ENVRI-FAIR (Petzold and Glaves, 2018; Petzold et al., 2020), with the overarching goal of promoting findability, accessibility, interoperability, and reusability of digital assets through the development of FAIR compliant common policies and technical solutions following expectations of the authorities promoting FAIRness (Cocco et al., 2019).

Similarly, several scientific communities are attempting to fill existing gaps between FAIR principles and viable practices for FAIRness by exploring pilot implementations as in life science case (Wilkinson et al., 2017) that propose standards-based reference technologies. Furthermore, an action plan for FAIR adoption has been presented by the European Commission Expert Group with the "Turning FAIR data into reality" report (Collins et al., 2018).

Relying on the experience matured in the framework of the European Plate Observing System (EPOS) (www.epos-eu.org) RI (Bailo et al., 2016), we argue that the aforementioned search for technological solutions need to be further elaborated. Already existing RIs, in particular those involved in the EOSC initiative (EOSC, 2018), would incredibly benefit from a clear and technically consistent roadmap to make technologies already in place better adhere to FAIR principles.

[1]https://www.force11.org/
[2]https://fairsharing.org/
[3]https://www.go-fair.org/
[4]http://www.copdess.org/enabling-fair-data-project/
[5]https://www.rd-alliance.org/

PERSPECTIVES ON ADOPTION OF FAIR PRINCIPLES

Perspective view and expertise related to technical implementation of the FAIR principles matured in the context of EPOS RI, which engages 10 different scientific communities in the solid Earth domain and integrates more than 250 research infrastructures and data providers.

On the basis of experience and know-how in the EPOS community of practitioners, experts, and engineers, a common approach was observed, which is described by the re-organization of FAIR principles into a four-stages roadmap. Such roadmap is part of a more general FAIR adoption process that may enable potentially any RI in the solid Earth domain and beyond to build FAIR Research Infrastructures. List of abbreviations to ease readability of the text is reported in the **Supplementary Material**.

Four-Stages Roadmap Approach

The driving concept is that FAIR principles are not easily understood by every domain scientist, data practitioner and IT specialist (here summarized as *RI implementers* as they all contribute to the technical construction of an RI), because such professionals have a specific mindset when it comes to building a system for data stewardship in the scientific domain.

The roadmap hence reflects the pragmatic approach undertaken by RI implementers and considers four different stages of development (**Figure 1A**), each of them referencing detailed FAIR principles [as described in (Wilkinson et al., 2016)]: (a) *data* stage, (b) *metadata* stage, (c) *access* stage, and (d) *use* stage.

In the following paragraphs, references to detailed FAIR principles will be done by using letters of the acronym F1, F2, etc. for Findability detailed principles, A1, A1.1, etc. for Accessibility detailed principles and so on, as defined in Box 2 of Wilkinson et al. (2016).

The first concern of RI implementers is indeed the data, which constitutes the main business in the scientific domain and whose collection process is often hard, effort-consuming and resulting from ingenious and complex experimental techniques.

The first stage (S1) therefore considers datasets, data products, research objects and any other resource to which we refer to as *data* in this context. Wherever data come from—laboratory or field experiment, selection or collection, monitoring sensors, whether organized in networks, or standalone, etc.—they usually need to be harmonized when provided in heterogeneous formats in order to ensure interoperability. As a consequence, data providers in a certain domain should agree on community standards (R1.3), broadly applicable language and formats for representation of knowledge (I1, I2, I3); data should then be uniquely and globally identified using persistent identifier (PID) (F1), and indexed in a searchable resource (F4), taking also into account licensing aspects to guarantee intellectual property rights and appropriate credit (R1.1).

Once data is properly collected and managed, RI implementers are concerned with attaching information to facilitate findability encompassing searchability and

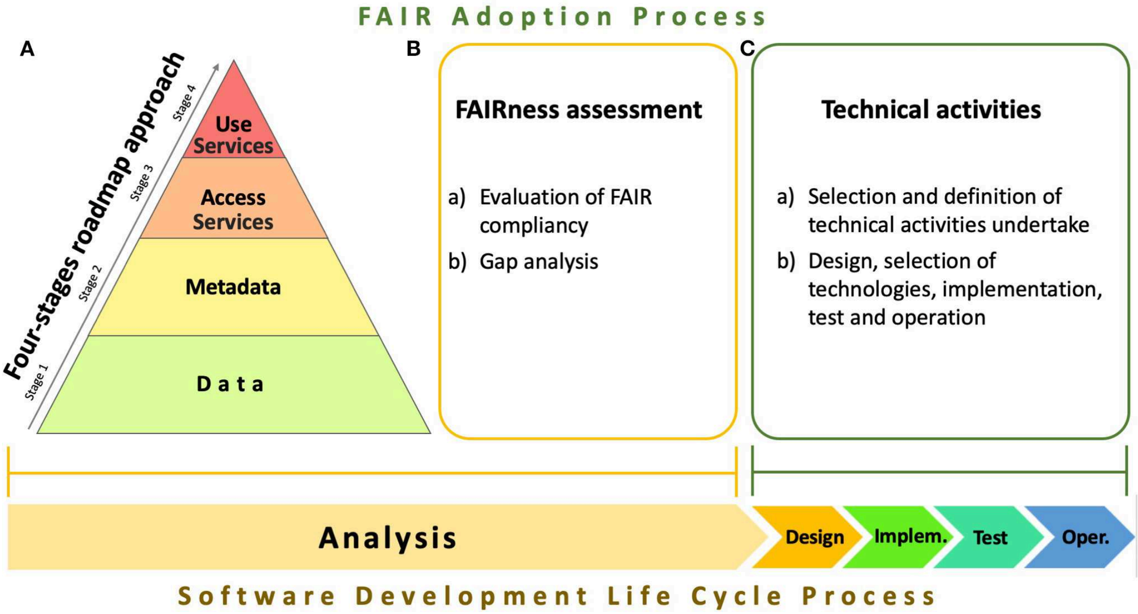

FIGURE 1 | FAIR adoption and Software Development Life Cycle (SLDC) processes are shown in parallel and correspondences between the two are highlighted. FAIR adoption is a three-step process: the first step **(A)** is represented by the four-stages roadmap (Bailo, 2019), which reorganizes detailed FAIR principles (Wilkinson et al., 2016) according to the mindset of RI implementers; the second step **(B)** is devoted to the evaluation of re-organized FAIR principles at each stage, and to a gap analysis to make emerge the priority by which FAIR principles at each stage need to be addressed; the third step **(C)** is devoted, on the basis of the gap analysis, to the selection of technical activities to undertake, and then to actual implementation. Software Development Life Cycle, described in section FAIR adoption process, is here compared to the FAIR adoption process. What emerges is that step one and two of the FAIR adoption process are carried out in the Analysis Phase in the SDLC; indeed, FAIR principles are elicited as requirement and gap analysis in order to understand which ones will be implemented at each "round" of implementation is performed. Phases from design to operation are executed in the "technical activities" step of the FAIR adoption process.

contextualization. *The second stage (S2)* thus concerns metadata implementation, which allows proper description of data—if rich enough (F2, I3)—and supports provenance recording (R1.2) and license information (R1.1). Usage of standards (R1.3), based on existing vocabularies (I2) and formal languages (I1) also facilitates the process. Similarly to S1, also metadata needs to be referenced by a unique persistent identifier (F1, F3) and to be indexed in searchable resources (F4). Metadata are supposed to be persistent and accessible even if data no longer exist (A2).

Well-described and contextualized data need to be accessible. Hence, *the third stage (S3)* deals with the provision of services that allow both human and machine access to data and metadata. They should include search functionalities (F4) and enable access (e.g., download) also through persistent identifiers (F1). Such services need to use standardized communication protocols (A1.1) and must also support authentication and authorization mechanisms if applicable (A1.2).

Access to data is not, however, the end of the scientific process that aims at producing meaningful and interpretable data. This is implicitly assumed by the FAIR principles that include reusability (and lately reproducibility), but also apparent in several Data Lifecycle models where other steps related to the usage of data are envisaged (Ball, 2012). Such steps are concerned with all functionalities that go beyond data access, for instance data analysis and processing.

FAIR RIs and data stewardship systems should then address a *fourth stage (S4)* concerned with services that *make use* of data (S1) and metadata (S2) FAIRly accessed (S3) and produce new data products as output. Such a stage guarantees that analysis performed on FAIR data still produce FAIR data as output. A common use case is indeed represented by services that use machine-readable and machine-actionable data to perform data analysis, visualization and processing, and produce what is often referred to as "data products" also in other environmental RIs.

FAIR Adoption Process

RI implementers setting up or upgrading an existing RI usually follow a system development life-cycle (SDLC) process (Blanchard, 2004). In the current work, an SDLC inspired by the waterfall development model was used, which encompasses the following steps: (a) *analysis*, including use cases and requirements collection; (b) *design*, including architecture design, and identification of architectural components matching requirements, (c) *implementation,* through software developments and adoption of suitable technologies, (d) *test*, (e) *operation and maintenance.*

Although FAIR principles claim to be technically non-prescriptive (Mons et al., 2017), they need to intersect and be part of SDLC since its early phases, leading to three main questions:

i where FAIR principles have to be considered in the SDLC

ii what FAIR detailed principles should be implemented first, and what would be a correct sequence and time-line
iii how FAIR principles should be technically addressed

FAIR principles define *what* the system should provide and *how* it should be provided, they can therefore be considered as *requirements* in the SDLC and should be taken into account during the analysis phase.

However, how to manage contexts where a Research Infrastructure already exists and needs to be upgraded to be compliant to FAIR principles is still an open question. This is indeed a common status for many RIs, for instance all those in the ENVRI-FAIR initiative.

The reorganization of the FAIR detailed principles into a four-stages roadmap together with potential technical activities to implement them at each of the stages of the roadmap (**Table 1A**) provide a perspective to answer to this latter question and also to questions (ii) and (iii).

Although the four-stages roadmap makes it easier for RI implementers to understand FAIR principles within a framework they are familiar with, a clear process is required for going from the roadmap and principles to actual technical implementations fulfilling FAIR principles.

A three-step FAIR adoption process can be used, and each step is detailed as follows:

1. *Per-Stage Implementation Approach.* For Research Infrastructures that are built from scratch, the pyramidal roadmap is likely to be followed in a sequential way starting from the bottom data layer (S1) to the top services for usage layer (S4), with a potential exception for processing services producing FAIR data products in output using FAIR dataset as input, as discussed in the next section. Indeed it wouldn't make sense to consider, for example, services for accessing data (S3) when data is not defined, stored, and not described by a (rich) metadata standard. For existing RIs that need to be enhanced to reach an higher level of FAIRness, the four layers can be considered also in a non-sequential way, as the work can be carried out only where needed. So, for instance, if a minimal level of FAIRness exist at data stage (e.g., dataset are at least stored in an accessible repository) and at metadata stage (e.g., simple metadata available for stored datasets), it is then possible to only improve services for accessing such assets (S3).
2. *FAIRness Assessment.* In order to enhance a specific stage, its compliance to the re-organized detailed FAIR principles needs to be evaluated and a gap-analysis performed to define technical activities addressing criticalities emerged during the evaluation. Currently, several initiatives are devoted to evaluating compliancy to FAIR principles[6, 7, 8] (Wilkinson et al., 2019); as consensus about one or more evaluation methods is reached within the scientific community, they might be adopted as canonical evaluations.
3. *Technical Activities Definition.* On the basis of the FAIRness evaluation results and the emerged gaps, the actual technical activities needed to fulfill a specific FAIR principle can finally be defined and executed in the domain specific context by following steps from design (b) to operation (e) of the SLDC.

In this three-step FAIR adoption process the SDLC is not confined to the technical implementation step (3), as its analysis phase encompasses also the *Definition of the stages to implement (1)* and *FAIRness evaluation (2)* steps, which is where FAIR detailed principles are defined and elicited as requirements (**Figure 1**).

In the following paragraph we describe the technical activities required for implementing FAIR requirements at each stage.

Technical Implementation

FAIR principles, in their definition, implicitly refer to the adoption of certain technological solutions, for instance web services or PIDs. Existing technical-related resources (Collins et al., 2018) usually mention these elements as best practices, avoiding any mandatory guideline. However, a clear process for the implementation of FAIR principles, like the three-step FAIR adoption process described in the previous section, requires an explicit definition of technical activities.

By *technical activities* in this context we refer to: (a) *tasks* with a clear technical outcome (e.g., community agreements about data standards, R1.3); (b) activities for construction of *architectural building blocks* of a system, including software or hardware, needed to provide features defined by FAIR requirements (e.g., metadata catalog building block for rich metadata storage - F2); (c) activities aiming at defining *conceptual models* and paradigms (e.g., adoption or extension of existing ontologies for knowledge representation—I1). We do not refer nor propose any specific software or tool, since they would change as technology evolves.

Table 1A provides a compact overview at each stage (column 1) of technical activities (column 3) RI implementers need to undertake in order to fulfill specific FAIR principles (column 2). Such table, by establish a correspondence between specific FAIR principles and technical activities at each stage, represents a pragmatic approach implementing the three-step FAIR adoption process.

FAIR ADOPTION EXAMPLE

In this section we describe the FAIR adoption process—including the four-stages roadmap—described in the previous section, in the context of the EPOS pan-European Research Infrastructure.

European Plate Observing System

The European Plate Observing System (EPOS) is building a pan-European research infrastructure for solid-Earth sciences. Its Preparatory Phase (PP), funded under the European Commission's FP7 Work Programme, ran from 2010 to 2014 and reached the ambitious goal of creating the conditions for the integration of existing and future national and international research infrastructures (RIs) in Europe with the final goal of

[6]http://blog.ukdataservice.ac.uk/fair-data-assessment-tool/
[7]https://www.go-fair.org/2017/12/11/metrics-evaluation-fairness/ [http://aims.fao.org/activity/blog/put-fair-principles-practice-and-enjoy-your-data]
[8]https://www.biorxiv.org/content/biorxiv/early/2018/09/25/418376.full.pdf

TABLE 1 | At each stage of the roadmap starting from the bottom (**Data Sheet 1**) to the top stage (Services For Use, S4), specific FAIR principles (Wilkinson et al., 2016) related to each layer, together with corresponding technical activities, are listed.

Table 1A | "Technical Activities" are described (bold text) and shortly discussed (plain text) in order to outline the main challenges to be addressed and issues that may be encountered in technical implementation.

	FAIR Principle	Technical activity
SERVICE FOR USE (S4)	FAIR principles in this stage do not address computation or visualization services	***Implementation of computational services, analysis or visualization services.*** Processing tools providing services that use FAIR data as input and perform computations should take into account: (i) standard protocols for transferring data (e.g., gridFTP[a]), because common HTTP based web-services may encounter issues when transferring huge amounts of data; or (ii) pointers to data (e.g., file URLs) encoded in a machine-readable format; (iii) technologies for authentication and authorization, usually needed when computational tasks are required as the system needs to account users' resource usage; (iv) mechanisms to track provenance information, attach appropriate metadata, as well as dealing with assignment of PID to the data product, in order to ensure that data products obtained still comply with FAIR principles.
SERVICE FOR ACCESS (S3)	**F4**. (meta)data are registered or indexed in a searchable resource. **A2.** metadata are accessible, even when the data are no longer available. **A1**. (meta)data are retrievable by their identifier using a standardized communications protocol. **A1.1** the protocol is open, free, and universally implementable.	***Implementation of (web) services for making (meta)data catalog resources accessible and searchable, by using open standard communication protocols.*** Services for access should include search and access functionalities and consider that standard communication protocols together with standard technologies should be selected. A wide range of options is available, spanning from RESTful (Richardson and Ruby, 2007) web services to other SOAP approach implementing search functionalities, (meta)data object referencing by identifier which should also persist when actual (meta)data is no longer available.
	A1.2 the protocol allows for an authentication and authorization procedure, where necessary.	***Implementation of authentication and authorization services.*** Authentication and Authorization (AA) should be also supported in order to achieve the "as open as possible, as closed as necessary" principle mentioned in Guidelines on FAIR Data Management in Horizon 2020 (European Commission, 2016). Setting up AA system includes: (a) selection of interoperable authentication protocols (e.g., Oauth 2.0), (b) management of users within a catalog, (c) assignment and handling of appropriate Authorization schemas. Complexity is increased by legalistic and policy aspects related to users' authorizations, and by potential interoperation with existing federated AA service providers, e.g., EDUgain network[b]
METADATA (S2)	**F2**. data are described with rich metadata (defined by R1). **F3**. metadata clearly and explicitly include the identifier of the data it describes. **R1.** metadata are richly described with a plurality of accurate and relevant attributes. **I1**. metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation. **I2**. metadata use vocabularies that follow FAIR principles. **I3**. metadata include qualified references to other (meta)data. **R1.3**. metadata meet domain-relevant community standards	***Selection of a metadata model which allows description of concepts of interest in a formal language, using a common vocabulary and serialized in a machine-readable format.*** Metadata model choice should fall upon an existing standard or its extension taking into account (a) the need for a rich metadata model which includes information about context, quality, condition, or characteristics of data. This guarantees that all concepts in the context of interest are captured, also these going beyond the immediate description of dataset and related to concepts like projects and funding used to produce data; (b) the need to manage ontologies and vocabularies: the driving criteria here should to use or extend existing semantic models for representation and concepts linking, e.g., DCAT-AP[c]; (c) metadata models and ontologies should be easily serializable in formats that relies on machine readable standards and schemas (e.g., RDF/turtle, RDF/XML, JSON-LD).
	F1. metadata are assigned a globally unique and persistent identifier.	***Selection of a PID system which guarantees technical reliability, authority, and ensures a long-term viability.*** PIDs require dedicated machinery and software to be issued, resolved and managed. A common solution is to rely on organizations—e.g., Datacite (Jan, 2009), ePIC[d]—but other options might be considered (Sicilia et al., 2019).
	R1.1. metadata are released with a clear and accessible data usage license.	***Discussion of a metadata policy and consequent adoption of a license ensuring metadata collected are made available under clear usage conditions.*** Licenses require rich enough information to be machine-readable and properly cited. A relevant example is creative commons which provide an easy way to choose a license and to share it in a machine-readable way[e] Besides technical aspects, the establishment of policies for metadata may need community-wide and, depending on the size of the community, they may require implementation of appropriate communication, community building, and organizational strategies.

(Continued)

TABLE 1A | Continued

	FAIR Principle	Technical activity
	F4. metadata are registered or indexed in a searchable resource. **A2.** metadata are accessible, even when the data are no longer available.	***Selection of metadata catalog, structured according to international standards or schemes.*** Such activity points out the need for a specific technical building block, i.e., the metadata catalog, which supports at least search functions and CRUD (Create, Read, Update, Delete) operations. Such catalog should take into account a rich logic schema (for representing rich metadata) and fit for purpose technologies for metadata storage (e.g., RDBMS, triplestore, noSQL etc.).
	R1.2. metadata are associated with detailed provenance.	***Selection of metadata provenance model for representing Information concerning the creation, attribution, or version history of managed data.*** Depending on the complexity and accuracy of provenance information, different technologies can be used, spanning from PIDs for simple data producer citation, to more complex workflow management and tracking systems (Filgueira et al., 2015) for tracking information about full history of the processing chain.
DATA (S1)	**I1**. data use a formal, accessible, shared, and broadly applicable language for knowledge representation. **R1.3**. data meet domain-relevant community standards **I2**. data use vocabularies that follow FAIR principles. **I3**. data include qualified references to other (meta)data.	***Selection or creation of a data model which provides a standard format for information resource description, supporting an ontology and providing standard vocabulary of terms.*** A main step for a FAIR data provision is its harmonization by adopting a standard format within a domain-specific community (e.g., shapefiles ESRI, 1998); need for harmonization frequently occurs typical "long-tail" contexts where data are produced by several different instruments on a per-sample basis, for instance in laboratories. Some formats embed in their serialization also an ontology (e.g. stationXML[f]); in such cases, appropriate FAIR, community-agreed ontologies need to be selected and adopted, possibly reusing or inheriting from existing ones (e.g., EnvO Buttigieg et al., 2016).
	F4. data are registered or indexed in a searchable resource.	***Selection of a Data catalog which supports discovery of datasets, and adoption of appropriate storage and preservation strategies.*** A system for data storage should be selected on the basis of data type, e.g., georeferenced layered data might need a database with GIS support, while other types of data might require filesystem with hierarchical folders or noSQL databases. Furthermore, Certification of Data repository (Dillo and de Leeuw, 2018) might be also considered.
	F1. data are assigned a globally unique and persistent identifier.	***Selection of a PID system which guarantees technical reliability, authority, and ensures a long-term viability.*** PIDs require dedicated machinery and software to be issued, resolved and managed. A common solution is to rely on organizations—e.g., Datacite (Jan, 2009), ePIC[g]—implementing the above, but other options might be considered (Sicilia et al., 2019).
	R1.1. data are released with a clear and accessible data usage license.	***Development and adoption a Data policy which ensures that data collected or created by the communities, once quality controlled, are made available under clear usage conditions licenses.*** From a technical point of view, licenses require rich enough information to be machine-readable and properly cited. A relevant example is creative commons which provide an easy way to choose a license and to share it in a machine-readable way[h]. Besides technical aspects, for establishing data policies community-wide agreements are needed and, according to the size of the community, they may require implementation of appropriate communication, community building, and organizational strategies.
	R1.2. data are associated with detailed provenance.	***Selection of data provenance model for representing Information concerning the creation, attribution, or version history of managed data.*** Information about provenance of datasets are usually tracked by means of appropriate metadata models. This activity hence focuses on the creation, extension or update of metadata models for tracking provenance, that should be implemented then at metadata stage.

[a] *https://www.globus.org/sites/default/files/GFD-R.0201.pdf*
[b] *https://edugain.org/*
[c] *https://joinup.ec.europa.eu/release/dcat-ap/11*
[d] *https://www.pidconsortium.eu*
[e] *https://creativecommons.org/choose/*
[f] *https://www.fdsn.org/xml/station/*
[g] *https://www.pidconsortium.eu*
[h] *https://creativecommons.org/choose/*

TABLE 1B | Technical implementation in the EPOS use case is described either at Thematic Core Services (TCS) level and at the Integrated Core Service (ICS) level.

	FAIR Principle	Technical implementation	
		ICS	**TCS**
		Implementation of computational services, analysis or visualization services.	
SERVICE FOR USE (S4)	FAIR principles in this stage do not address computation or visualization services	Computational and visualization services are implemented through the so called ICS-D, i.e., distributed external services, whose interoperation with ICS-C (central hub) is still ongoing[a], as in the case of Enlighten web tool, a Jupiter notebook based system where tools for processing and visualization of seismic data are made available. Communication between ICS-C and Enlighten is designed to be APIs based, and pointers to data can be provided in CWL (Common Workflow Language[b]) format. Another ICS-D case, based on the work from DARE project[c], also worked on Provenance Solutions based on S-ProvFlow[d] Main challenge of ensuring that data products obtained by FAIR dataset processing still comply with FAIR principles is still ongoing.	Some TCS hold computational facilities that are candidate for ICS-D status, for instance Geological community claim to provide services for calculating Borehole Geometry and Gravity field, while Anthropogenic Hazard community can make available generic computation services (based on computing resources of Polish National Grid). All these are however still at prototype level.
SERVICE FOR ACCESS (S3)	**F4**. (meta)data are registered or indexed in a searchable resource. **A2.** metadata are accessible, even when the data are no longer available. **A1**. (meta)data are retrievable by their identifier using a standardized communications protocol. **A1.1** the protocol is open, free, and universally implementable.	***Implementation of (web) services for making (meta)data catalog resources accessible and searchable, by using open standard communication protocols.***	
		ICS-C node provides machine-readable and machine-actionable HTTP RESTful Web APIs (**A1.1**) to interact with the metadata catalog, to trigger system functionalities (e.g., convert data) and to search for data and metadata (**F4**). The latter can be obtained also by referencing (meta)data identifiers (**A1**). Discussions are undergoing whether to make metadata available when data is no longer available (**A2**).	Thematic communities repositories currently provide access to data and metadata by means of HTTP(s) web services. Some communities use existing standards (e.g., OGC WMS and WFS in Geology, FDSN services in seismology) which all allow for (meta)data search (**F4**), implemented with open protocols (**A1.1**). Communities starting from scratch right after the issuing of FAIR principles were guided to adopt known and open standards (usually EU recommendations) complying with **F4**, **A2**, **A1.1**.
	A1.2 the protocol allows for an authentication and authorization procedure, where necessary.	***Implementation of authentication and authorization services.***	
		Authentication and Authorization (AA) are implemented as an external service. Users requests have to go through a proxy, developed with Nginx, which takes advantage of auth_request and proxypass directives to route the requests to the AA service. The service issues a token to redirected users that need to be authenticated or checks the token validity and authorization to be used for communication with TCS secured repositories over https (**A1.2**). Such service leverages on existing AA architecture models (e.g., AARC blueprint architecture[e]) and software (e.g., Unity IDM[f]) With this solution ICS don't impose Authentication over all the domain services it integrates, but can propagates the AA policies adopted at community level.	Authentication and Authorization (AA) might be of limited applicability for some data in the solid Earth domain, i.e., data that do not envisage commercial usage, nor include personal information; yet there is a portion of data that requires AA in order to manage embargoed, commercial, and security-sensible datasets, or data products requiring processing (e.g. satellite "on demand" images). In addition to that, AA systems are fundamental to account the usage of the data in a detailed way, showing what type of users have requested access to data. TCS to which AA can be applied are encouraged to adopt standard AA technologies like OAuth2[g] or join some federated Identity Provider like EDUgain, as in the case of Seismological Data.
METADATA (S2)	**F2**. data are described with rich metadata (defined by R1). **F3**. metadata clearly and explicitly include the identifier of the data it describes. **R1.** metadata are richly described with a plurality of accurate and relevant attributes. **I1**. metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation. **I2**. metadata use vocabularies that follow FAIR principles. **I3**. metadata include qualified references to other (meta)data. **R1.3**. metadata meet domain-relevant community standards	***Selection of a metadata model which allows description of concepts of interest in a formal language, using a common vocabulary and serialized in a machine-readable format***	
		CERIF (Bailo and Jeffery, 2014) formal conceptual model for research domain was selected for metadata storage. It encompasses many metadata information (i.e. rich standard) (**F2, R1**) (e.g., Project, Person, Organization, Services, Datasets, Facility etc.), includes a semantic layer, geographic binding, and implements time stamps and roles. Each of the entities include an attribute for identifiers (**F3**). CERIF can be expressed by means of an Entity Relationship model and can be mapped to more known standards (I1, R1.3) (e.g., DCAT-AP). It has its own vocabulary but through the semantic layer any vocabulary can be mapped and used (I2). CERIF is a superset of many metadata models used by communities (e.g., Dublin Core, stationXML, DCAT-AP etc.) meaning that they can be mapped to CERIF (R1.3) and, in the other direction, can be produced from CERIF. In order to collect scientific metadata and data-related metadata from Thematic Communities, and ingest it into the CERIF based metadata catalog, the EPOS-DCAT-AP metadata model, an extension of DCAT-AP, with RDF/turtle serialization was used (Trani et al., 2018). The usage of such human understandable (but also machine readable), wide known, popular standard simplifies the process of TCS metadata provision. EPOS-DCAT-AP extension is rich enough (**F2**) to represent scientific metadata, it includes identifiers of the data (**F3**), and uses W3C DCAT vocabulary (**I1**). The extension was created also to enrich the original Application Profile, thus complying with R1 principle as well. Metadata encompassing elements that go beyond scientific data related information (e.g., projects, processing facilities) are not captured by EPOS-DCAT-AP and directly ingested into CERIF.	Metadata model at TCS level present a heterogenous landscape: many communities already used standard models (e.g., Geology which strongly relies on OGC standards and INSPIRE directive[h] which usually comply to I1. In this case, communities were encouraged to enhance their standard, if necessary, to be rich enough (**F2**) and to comply with related FAIR principles (**I2, F3**). Other communities started from scratch, as in the case of Laboratory community that set up a metadata and vocabulary task force[i] and Geomagnetic community which leveraged on ISO 19139[j] In such cases, a continuous interaction and guidance action was put in place in order to make them comply to FAIR principles. In most cases, quite a work is required for complying to the rich metadata FAIR requirement **R1**.

(Continued)

TABLE 1B | Continued

	FAIR Principle	Technical implementation ICS	TCS
	F1. metadata are assigned a globally unique and persistent identifier.	***Selection of a PID system which guarantees technical reliability, authority, and ensures a long-term viability***	
		Intense debate is currently ongoing, pointing out that sustainability aspects need to be addressed first in order to be able to assign PIDs released by international organizations (e.g., DOIs or ePIC). Potential usage of UUID or GUID is also discussed.	As for the data stage, most of the TCS adopted Datacite DOIs in different technical implementations flavors. However, issues are being encountered when it to link data to the landing-page information in a machine-readable manner.
	R1.1. metadata are released with a clear and accessible data usage license.	***Discussion of a metadata policy and consequent adoption of a license ensuring metadata collected are made available under clear usage conditions.***	
METADATA (S2)		ICS are compliant with EPOS data policy[k], which states that "in order to ensures the widest dissemination and publicity for EPOS managed services, assumes that metadata are easily and freely accessible at any time, with as few restrictions as possible. Suppliers are thus encouraged to affix open licenses, preferably Creative Commons 4.0 CC:BY, to their metadata. The machine-readable version of this license will allow User(s) to identify the relevant datasets through search engines licenses filters." As a consequence, metadata is provided with CC:BY.	Likewise ICS, all TCS have agreed to the EPOS Data Policy which encourage to provide metadata with CC:BY licensing schema. Convergence and consensus building of all communities toward this policy has required more than 3 years.
	F4. metadata are registered or indexed in a searchable resource. **A2.** metadata are accessible, even when the data are no longer available.	***Selection of metadata catalog, structured according to international standards or schemes.***	
		CERIF (Bailo and Jeffery, 2014) formal conceptual model for research domain was selected, as it supports the management research-related entities such as people, projects, organizations, publication, products, etc. and the relationships between them. The catalog is implemented in PostgreSQL RDBMS which enable metadata search (**F4**). Metadata records can still exist also when pointers to dataset or services are no longer available (**A2**). It supports formal syntax and declared semantics, and guarantees referential and functional integrity. TCS metadata and TCS service description are ingested into the main central CERIF catalog at ICS-C level.	Similarly to the Data Stage, communities adopted heterogeneous solutions fitting to the existing RI landscape. Solutions are often Institution dependant as in the case of Seismology, where Seismic networks are federated internationally through web services, so that specific implementation choices are left to the national Data Centers. Most straightforward solutions were usage of relational databases (e.g., SQL based DB for FDSN services at INGV), but also MongoDB[l] based catalogs were implemented (Near Fault Observatories—INGV).
	R1.2. metadata are associated with detailed provenance.	***Selection of metadata provenance model for representing Information concerning the creation, attribution, or version history of managed data.***	
		Metadata is stored with the CERIF format, which supports provenance because of the time-stamped linking entities. Nevertheless, according to (Bailo et al., 2016) there is still necessity to further develop in CERIF some provenance aspects such as the integration of causal-effect relationships among the entities and activities involved and re-used across processing tasks.	**R1.2** has not been tackled in a methodical way with the exception of a few communities that claim to be ready for producing and storing provenance information (e.g., seismology).
DATA (S1)	**I1**. data use a formal, accessible, shared, and broadly applicable language for knowledge representation. **R1.3**. data meet domain-relevant community standards **I2**. data use vocabularies that follow FAIR principles. **I3**. data include qualified references to other (meta)data.	***Selection or creation of a data model which provides a standard format for information resource description, supporting an ontology, and providing standard vocabulary of terms.***	
		Being ICS-C the node integrating datasets provided through TCS, no activities at DATA stage were carried out. TCS indeed have the ownership of data and are responsible for storage, preservation and data quality control.	A harmonization process was carried out transversally among TCS communities with similar data represented in different formats, as in the case of Volcanological and Seismological communities: Volcano observatories included seismic stations providing non-standard seismic waveforms format, while seismological communities had a long lasting tradition in using the FDSN web services standard[m] for providing access to waveforms in mini-seed format. As a result, Volcano Observatories and Seismological community converged toward the usage of FDSN and miniseed. A similar process was carried within National Research Infrastructures (NRI) at TCS level, as in the case of GNNS community which converged toward the provision of raw satellite navigation system data in RINEX format (Gurtner and Estey, 2007) and other standards for other types of data (full list in WP10 Service Validation Report[n] This did not happen in other communities as Geology which converged toward the usage of OGC compliant file formats (e.g., jpg) years ago. Interestingly, the harmonization process required limited technical skills, while leveraged on governance, communication, and consensus building capabilities.
	F4. data are registered or indexed in a searchable resource.	***Selection of a Data catalog which supports discovery of datasets, and adoption of appropriate storage and preservation strategies.***	
			Communities adopted heterogeneous solutions fitting to the existing RI landscape: filesystem with relational databases (e.g., Seismic waveforms at INGV), Cloud Object Storage provided by Amazon S3[o], Copernicus DIAS[p] or OVH Cloud[q] (Satellite Data TCS) and other solutions. Certification was not mentioned by any community, but all catalogs provided search capabilities (**F4**).

(Continued)

TABLE 1B | Continued

	FAIR Principle	Technical implementation	
		ICS	TCS
DATA (S1)	**F1.** data are assigned a globally unique and persistent identifier.	***Selection of a PID system which guarantees technical reliability, authority, and ensures a long-term viability.***	
			A few TCS communities implemented DOIs. Some of them adopted Datacite DOIs in a heterogeneity of situations, for instance DOIs to identify seismic networks, or DOIs to identify publications or experiments related to Laboratory data, or DOIs for Induced Seismicity episodes; some communities are testing ePIC Identifiers for daily seismic waveforms (Seismology in the framework of EUDAT). Although its widespread usage, DataCite DOI may present issues when it comes to referencing URLs pointing to data object in a machine-readable manner.
	R1.1. data are released with a clear and accessible data usage license.	***Development and adoption a Data policy which ensures that data collected or created by the communities, once quality controlled, are made available under clear usage conditions licenses.***	
			For addressing data policy related activity (**R1.1**), Thematic communities and EPOS Management office co-developed and agreed upon the EPOS data-policy[r]Tuesday, January 21, 2020 4:31 pm, where default licenses have been identified (CC:BY and CC:BY:NC from the Creative Commons 4.0 licenses set). These licenses place a legal obligation on the end user to acknowledge the owner of the TCS services, and also define the conditions upon which the data can be used.
	R1.2. data are associated with detailed provenance.	***Selection of data provenance model for representing Information concerning the creation, attribution, or version history of managed data.***	
			R1.2 has not been tackled in a methodical way in the context of EPOS, with the exception of a few communities that claim to be ready for producing and storing provenance information (e.g., seismology).

[a]*https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5be0cc293&appId=PPGMS;* [b]*https://www.commonwl.org/;* [c]*http://project-dare.eu/epos/;* [d]*https://www.knmi.nl/kennis-en-datacentrum/project/s-provflow.;* [e]*https://aarc-project.eu/wp-content/uploads/2017/04/AARC-BPA-2017.pdf;* [f]*https://www.unity-idm.eu/.;* [g]*https://oauth.net/2/;* [h]*https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32007L0002&rid=1;* [i]*https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5be0c154a&appId=PPGMS;* [j]*https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b1725572&appId=PPGMS..;* [k]*https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b190e31a&appId=PPGMS;* [l]*https://www.mongodb.com/;* [m]*https://www.fdsn.org/;* [n]*https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5be0b3e9f&appId=PPGMS.;* [o]*https://aws.amazon.com/it/s3/;* [p]*https://eo4society.esa.int/2018/04/20/copernicus-dias-data-and-information-access-services/;* [q]*https://www.ovh.it/;* [r]*https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b190e31a&appId=PPGMS.*

improving access to data, products, and services. The technical architecture (Jeffery et al., 2018) was built following a co-design approach, with a continuous interaction among communities stakeholders, data practitioners, scientists, and engineers who shared their skills and experiences.

EPOS also undertook an Implementation Phase from 2014 to September 2019, where the implementation of a pre-operational Integrated Core Services (ICS) system and of interoperable Thematic Core Services (TCS) were carried out, and the technical architecture confirmed.

Such an architecture is composed by three fundamental elements:

- *Thematic Core Services (TCS)*, representing datasets and services provided by the domain specific communities. At such community level, EPOS has promoted and stimulated the harmonization of data management, access methods and policies, as well as services (e.g., processing, visualization) and resource provisioning by: (1) Fostering the creation of new European-wide thematic hubs; and (2) Supporting existing organizations (e.g., ORFEUS29 for seismology).
- *Integrated Core Services-Centralized (ICS-C)*, representing the novel system which integrates resources provided by the TCS. Interoperability between ICS and TCS was implemented by activities also envisaged by the four-stages roadmap in the FAIR adoption process, i.e., metadata related activities.
- *Integrated Core Services-Distributed (ICS-D),* which constitute the distributed part of the ICS, devoted to computational or visualization tasks, designed as services offered by e-Infrastructure providers and resource providers that—under clear procurement policies or SLAs—make resources available (e.g., HPC, HTC, data storage and data transport).

FAIR Implementation in EPOS

As many RI in the environmental domain, EPOS relies on already existing data providers, and as such contains peculiarities that need to be remarked, before addressing actual implementation activities (**Table 1A**).

The first remark is that technical implementation of FAIR principles in EPOS required consistent non-technical efforts, i.e., keeping the full community together by adopting community building actions, communicating results, manage legalistic, and governance aspects. The actual magnitude of these efforts needs to be taken into account as it influences the technical work, and make the entire FAIR adoption process difficult to fit to the timelines envisaged by some policy makers, for instance at EOSC level[9]. The FAIR adoption process in EPOS started years ago, is cyclically being refined and the experience suggests that this will be an iterative activity still going on in the medium term in the future.

Secondly, in EPOS the process of creating the conditions for interoperability among existing RIs has actually started in 2010, long before the FAIR principles were made explicit by FORCE11 (Wilkinson et al., 2016). This entails two considerations: (a) EPOS architecture was FAIR-compliant since its conception because some of its driving aspects are key concepts for FAIR principles as well, i.e., metadata and ontologies, identifiers, technological interoperability; (b) the methodology was applied to an existing research infrastructure, and in the wider Environmental RIs landscape this is often the case in the experience of the authors.

The third remark is that the evaluation step in the FAIR adoption process was carried out on the basis of IT expertise and skills of EPOS RI implementers, being rigorous methods to evaluate and making the gap analysis not yet available before 2016 and currently still under discussion (Wilkinson et al., 2019). Importantly, such FAIR evaluation methods developed at different levels and in different continents, are taking advantage of domain specific RIs like EPOS and RIs clusters like ENVRI-FAIR, when it comes to have an harmonized application, test, and feedback of the proposed evaluation methods and questions, as in the case of the work being carried out in ENVRI-FAIR WP5 and WP7.

Details about technological activities carried out in EPOS are reported in **Table 1B**.

DISCUSSION

Perspective on FAIR Principles

Nowadays, FAIR principles are largely accepted by a wide stakeholders range, all over Europe (and beyond); the acronym well-reflects the concept of "establishing equally a common technical background" for all those involved in data provision; also, it has the merit of making technical concepts like "interoperability" understandable by non-technical audience; FAIR concepts discussions are cross-disciplinary and applicable also to non-technical domains [see "FAIR policies" discussions (Boeckhout et al., 2018)]; finally, being it a driving concept in European initiatives like EOSC, it has rapidly become a reference for all those RIs whose sustainability relies on European funding.

However, the road from principles to real FAIR RIs is still long: at now many initiatives focus primarily on the establishment of criteria for assessing "FAIRness" of data stewardship systems and then on complementary guidelines (Collins et al., 2018) or support actions (GO-FAIR) working on best practices and not yet attempting to formalize technical references architectures.

The current work aims at making a step forward on the "principles-to-reality" track, by emphasizing the need for technical guidelines and by proposing a four-stages roadmap related to technological solutions that ease real implementation.

The roadmap in **Figure 1**, rather than prescribing a sequential temporal approach, aims at capturing the mindset and approach of RI implementers. The roadmap also demonstrates that an approach that follows sequentially the letters of the FAIR acronym (i.e., implementing first technologies for Findability, secondly technologies for Accessibility, and so on), is not technically viable and does not fit with development practices within RIs.

The roadmap is part of a FAIR adoption process which also include an SDLC and reflects the real workflow RI implementers use in the actual technical work. The SDLC discussed here is simple but efficient and commonly used; others may be used to meet more advanced implementational techniques, and the FAIR adoption process would still work as an implementation guideline method.

Importantly, the FAIR adoption process can be applied also to existing RIs, which in the current FAIR landscape (for instance in the EOSC) represent the majority.

With the conceptual solutions proposed (**Table 1A**) and their actual technical implementation in EPOS (**Table 1B**), the authors want to foster efforts in providing clear technical guidelines for building or upgrading existing RIs data systems. In perspective, an agreed common view of the architectural technical building blocks may also facilitate the establishment of metrics for FAIR assessment of systems, not only of the data they steward.

FUTURE WORK

The fourth stage, which deals with services taking FAIR data as input to produce FAIR data products in output, paves the way for future work related to the extension of FAIR principles also to processing services, which seem not to have a habitat yet in the FAIR environment (definition, guidelines, etc.) although they are a key element for re-usability based on interoperability. Additional consideration on provenance aspects and on what building blocks are needed to produce FAIR provenance records should be done.

In the EPOS case study we observed that the common approach to RI system development influences RI architectures that turn out to be very similar across the EPOS Thematic Communities, in terms of system components. Hence, additional effort in the definition of FAIR reference architectures would make any RI, especially those providing access to similar resources, benefit of a pool of reference architectural building blocks and technical solutions that would ease implementation of FAIR compliant systems.

AUTHOR CONTRIBUTIONS

DB developed the main idea and wrote the manuscript. RP provided support in manuscript writing and contributed to the definition of technological solutions. MS, RR, and VV participated to EPOS team technical work and discussions that

[9]https://op.europa.eu/en/publication-detail/-/publication/78ae5276-ae8e-11e9-9d01-01aa75ed71a1

paved the way for roadmap elaboration, produced figures, and tables and reviewed the text. MC overviewed the manuscript and as EPOS coordinator, managed strategic FAIR RI activities in the EPOS framework.

FUNDING

This research has been supported by EPOS-IP EU project (No. 676564). VRE4EIC Grant agreement ID: 676247. ENVRI-FAIR Grant agreement ID 824068.

ACKNOWLEDGMENTS

We would like to thank the EPOSWP6–WP7 Technical Team.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2020.00003/full#supplementary-material

REFERENCES

Bailo, D. (2019). *Four-Stages FAIR Roadmap - FAIR "Pyramid,"*. Zenodo. Available online at: http://doi.org/10.5281/zenodo.3299353

Bailo, D., and Jeffery, K. G. (2014). EPOS: a novel use of CERIF for data-intensive science. *Proc. Comput. Sci.* 33, 3–10. doi: 10.1016/j.procs.2014.06.002

Bailo, D., Jeffery, K. G., Spinuso, A., and Fiameni, G. (2016). "Interoperability oriented architecture: the approach of EPOS for solid Earth e-infrastructures," in *Proceedings of the 2015 IEEE 11th International Conference on e-Science*, (Munich) 529–534. doi: 10.1109/eScience.2015.22

Ball, A. (2012). *Review of Data Management Lifecycle Models,* 14. Available online at: http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf. (accessed January 23, 2020).

Blanchard, B. S. (2004). *Systems Engineering Management.* Hoboken, NJ: John Wiley & Sons, Ltd,.

Boeckhout, M., Zielhuis, G. A., and Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: fair enough? *Eur. J. Hum. Genet*. 26, 931–936. doi: 10.1038/s41431-018-0160-0

Buttigieg, P. L., Pafilis, E., Lewis, S. E., Schildhauer, M. P., Walls, R. L., and Mungall, C. J. (2016). The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *J. Biomed. Semant.* 7, 1–12. doi: 10.1186/s13326-016-0097-6

Cocco, M., Bailo, D., Paciello, R., Freda, C., and Jeffery, K. (2019). The FAIRness horizon: from principles to practice, the EPOS experience in solid Earth science. *Geophys. Res. Abstr.* 21:13137. Available online at: https://ui.adsabs.harvard.edu/abs/2019EGUGA..2113137C. (accessed January 23, 2020).

Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., et al. (2018). *Turning FAIR Data Into Reality: Interim Report From the European Commission Expert Group on FAIR Data*. Interim Rep. from Eur. Comm. Expert Gr. FAIR Data.

Dillo, I., and de Leeuw, L. (2018). Coretrustseal. *VOEB-Mitteilungen*. 71, 162–170. doi: 10.31263/voebm.v71i1.1981

E. F. D. Community (2018). *Commitment Statement to Enabling FAIR Data in the Earth, Space, and Environmental Sciences.*

EOSC (2018). *Prompting an EOSC in Practice*. High Level Expert Group.

ESRI (1998). Environmental systems research institute, ESRI shapefile technical description. *Comput. Stat*. 16, 370–371.

European Commission (2016). *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*. Available online at: https://www.unive.it/pag/fileadmin/user_upload/SBA/documenti/BDA/H2020_Guidelines_on_FAIR_Data_Management_Version3.0_26_luglio_2016.pdf (accessed February 7, 2018).

FAIR Data Maturity Model WG (2019). Available online at: https://www.rd-alliance.org/groups/fair-data-maturity-model-wg (accessed June 29, 2019).

Filgueira, R., Krause, A., Atkinson, M., Klampanos, I., Spinuso, A., and Sanchez-Exposito, S. (2015). "Dispel4py: an agile framework for data-intensive eScience," in *Proc. - 11th IEEE Int. Conf. EScience, EScience*, (Munich) 454–464. doi: 10.1109/eScience.2015.40

Gurtner, W., and Estey, L. (2007). *Rinex-the Receiver Independent Exchange Format-Version 3.00.* Bolulder, CO: Astronomical Institute, University of Bern and UNAVCO.

Jan, B. (2009). "DataCite - a global registration agency for research data," in *4th Int. Conf. Coop. Promot. Inf. Resour. Sci. Technol. COINFO.* (Beijing) 257–261.

Jeffery, K., Harrison, M., Atakan, K., Bailo, D., Card, C., Michalek, J., et al. (2018). *D7. 2 Second (final) Report on EPOS-ICS Architecture*, 1–30.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., Da Silva Santos, L. O. B., and Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR data guiding principles for the European Open Science Cloud. *Inf. Serv. Use* 37, 49–56. doi: 10.3233/ISU-170824

Petzold, A., Glaves, H. (2018). ENVRI-FAIR - the next step towards FAIRer environmental research. *Am. Geophys. Union Fall Meet*. 2019:15684.

Petzold, G., Asmi, H. M., Bailo, A., Bundke, D., Vermeulen, U., Schaap, A., et al. (2020). *ENVRI - FAIR - Interoperable Environmental FAIR Data and Services for Society, Innovation and Research*. Zenodo, doi: 10.1109/eScience.2019.00038

Richardson, L., and Ruby, S. (2007). *RESTful Webservices.* O'Reilly Media, Inc.

Sansone, S. A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A. L., et al. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nat. Biotechnol*. 37, 358–367. doi: 10.1038/s41587-019-0080-8

Schultes, E., Strawn, G., and Mons, B. (2018). Ready, set, go fair: accelerating convergence to an internet of fair data and services. *CEUR Workshop Proc.* 2277, 19–23. Available online at: http://ceur-ws.org/Vol-2277/paper07.pdf. (accessed January 23, 2020).

Sicilia, M. A., García-Barriocanal, E., Sánchez-Alonso, S., and Cuadrado, J. J. (2019). "Decentralized persistent identifiers: a basic model for immutable handlers," in *Procedia Comput. Science*, eds M.-A. Sicilia, E. Simons, A. Clements, P. de Castro, and J. Bergström (Umeå University), 123–130. doi: 10.1016/j.procs.2019.01.087

Trani, L., Atkinson, M., Bailo, D., Paciello, R., Filgueira, R. (2018). Establishing core concepts for information-powered collaborations. *Futur. Gener. Comput. Syst*. 89, 421–436. doi: 10.1016/j.future.2018.07.005

Wilkinson, M. D., Dumontier, M., Ij.Aalbersberg, J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*. 3:160018. doi: 10.1038/sdata.2016.18

Wilkinson, M. D., Dumontier, M., Sansone, S., da Silva Santos, L. O. B., Prieto, M., Batista, D., et al. (2019). Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Sci. Data* 6:174. doi: 10.1038/s41597-019-0184-5

Wilkinson, M. D., Sansone, S. A., Schultes, E., Doorn, P., Da Silva Santos, L. O. B., and Dumontier, M. (2018). Comment: a design framework and exemplar metrics for FAIRness. *Sci. Data* 5:180118. doi: 10.1038/sdata.2018.118

Wilkinson, M. D., Verborgh, R., Bonino da Silva Santos, L. O., Clark, T., Swertz, M. A., et al. (2017). Interoperability and FAIRness through a novel combination of Web technologies. *PeerJ Comput. Sci.* 3:e110. doi: 10.7717/peerj-cs.110

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.