REPORT
doi:10.26443/seismica.v4i1.1537

SEISMICA

DOI, licence and citation uptake for seismological waveform data after 10 years of implementation effort

Helle A. Pedersen * [1,2], **Jonathan Schaeffer** [2], **Florian Haslinger** [3], **Rob Casey** [4], **Javier Quinteros** [5], **Lesley Wyborn** [6,7], **Elisabetta D'Anastasio** [8], **Jonathan B. Hanson** [8], **Jerry Carter** [4]

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France, [2]Univ. Grenoble Alpes, CNRS, INRAE, IRD, METEO-FRANCE, OSUG, 38000 Grenoble, France, [3]Swiss Seismological Service at ETH Zürich, Sonneggstr. 5, 8092 Zurich, Switzerland, [4]EarthScope Consortium, 1200 New York Avenue NW, Suite 454, Washington, DC 20005, USA, [5]GFZ German Research Centre for Geosciences, 2.4 Seismology, Potsdam, Germany, [6]National Computational Infrastructure Australia, Canberra, Australia, 2601, [7]AuScope Ltd, Melbourne, Australia, 3053, [8]GNS Science Te Pū Ao, 1 Fairway Dr., Lower Hutt, WEL, 5011, New Zealand

Author contributions: *Conceptualization*: Helle A. Pedersen, Florian Haslinger, Rob Casey, Javier Quinteros, Lesley Wyborn, Elisabetta d'Anastasia, Jonathan B. Hanson, Jerry Carter. *Methodology*: Helle A. Pedersen, Jonathan Schaeffer, Florian Haslinger, Javier Quinteros. *Software*: Jonathan Schaeffer. *Validation*: Helle A. Pedersen, Jonathan Schaeffer, Florian Haslinger, Rob Casey, Elisabetta d'Anastasia, Jonathan B. Hanson. *Formal Analysis*: Helle A. Pedersen. *Investigation*: Helle A. Pedersen, Florian Haslinger. *Writing - Original draft*: Helle A. Pedersen. *Writing - Review & Editing*: Florian Haslinger, Rob Casey, Javier Quinteros, Lesley Wyborn, Elisabetta d'Anastasia, Jonathan B. Hanson, Jerry Carter. *Visualization*: Helle A. Pedersen, Rob Casey.

Abstract The International Federation of Digital Seismic Networks (FDSN) has championed online open access to seismological waveform data for almost four decades. In 2014, FDSN recommended using DataCite Digital Object Identifiers (DOIs) for seismic networks to enhance data attribution, citation, and impact metrics. This study evaluates the level of adoption of DOIs and licences across FDSN-registered networks, analyzing their influence on data citation and compliance with FAIR (Findability, Accessibility, Interoperability, and Reusability) principles. 73% of seismic networks that have an assigned FDSN network code have adopted DOIs, with more than 80% DOI coverage for networks created after 2014. Licence adoption, not covered by present FDSN recommendations, remains low (8%), with significant regional variations. The main challenges are presently barriers to systematic data citation, whether on scientist or publisher side. Citations have increased substantially, but improvements are needed to support and implement correct data citation across all levels, including networks, data centers, scientists and journals. Of specific concern is the limitation on references set by some journals, which renders proper attribution impossible for studies using data from many seismic networks. This work highlights best practices and provides a set of recommendations for improving attribution, citation, and FAIRness of seismological waveform data, the latter including that FDSN should recommend licence on waveform data and a limited set of recommended licences. It also explores emerging ethical considerations, like the CARE principles, for Indigenous Data Governance. These insights aim to guide future FDSN strategies and foster enhanced alignment with FAIR and CARE principles. An added value of the assessment was that many minor errors and inconsistencies were identified and fixed at FDSN and in the seismological metadata.

Production Editor:
Andrea Llenos
Handling Editor:
Sam Teplitzy
Copy & Layout Editor:
Hannah F. Mark

Signed reviewer(s):
Tim Ahern
Barbara Romanowicz

Received:
December 16, 2024
Accepted:
March 11, 2025
Published:
April 1, 2025

1 Introduction

Seismologists have, through the International Federation of Digital Seismic Networks (FDSN, https://www.fdsn.org), promoted open seismological waveform data for almost four decades (Romanowicz and Dziewonski, 1986). From the outset, FDSN dedicated significant efforts to data and metadata exchange formats (Romanowicz and Dziewonski, 1986; Dziewonski, 1994), setting the seismological community on a path to open data sharing. But, despite the success of the FDSN open data policy in terms of increased data use, this increase was not reflected in the impact metrics of data sets and data generators (e.g. citation count, h-index). The typical reference to the data set(s), if any, was a sentence in the acknowledgment section of scientific papers. Therefore, FDSN decided in 2014 to recommend the use of DataCite Digital Object Identifiers (DOIs) (Evans et al., 2015) to promote proper citation of seismic network data.

This work assesses the improvement in data attribution and citation after 10 years of using DOIs to identify seismic networks in an international effort led by FDSN. Other improvements, more recent and with differing levels of maturity, are licences, AAI (Authentication and Authorization Infrastructure), as well as interoperability, and other related issues. We believe this assessment, carried out partly as part of the Geo-INQUIRE project (https://www.geo-inquire.eu), can feed into future strategies of FDSN, ORFEUS (Observatories and Research FAcilities for European Seismology, https://orfeus-eu.org/), EPOS (European Plate Observing System, https://www.epos-eu.org/), the EarthScope Consortium[1] (https://www.earthscope.org/), individual seismo-

*Corresponding author: helle.pedersen@univ-grenoble-alpes.fr

[1]The seismological services that were often referred to as 'IRISDMC'

logical data centers, seismic network operators and, hopefully, other communities.

The present document assesses the uptake of the FDSN recommendations of DataCite DOIs for seismic networks, to identify bottlenecks and problems and pave the road for solutions, improvements, and wider adoption. We will present our assessment broadly within the framework of the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles (Wilkinson et al., 2016), even though seismological waveform data distribution precedes FAIR by decades. Because data licences are a central issue for Reusability of data, we also analyze the uptake of licences across FDSN networks even though FDSN does not yet have any formal recommendations for licences.

The document is separated into six parts. The first one describes the organization of the technical elements relevant to the assessment and the workflow. The second and third parts present the outcome of the assessment and the inconsistencies that we detected. The fourth part presents some first findings on citation uptake, and the difficulties encountered to assess citations. The fifth part further analyses the results and places the results in the broader framework of FAIR and CARE. Finally, the sixth part presents lessons learned and suggestions for future improvements by scientists, network operators, data centers, and FDSN.

2 Building blocks for FDSN network data and metadata distribution, and workflow of the analysis

We here present the building blocks for seismological waveform data and metadata distribution, some of them going back several decades, and touch upon their relation to the FAIR principles (see also Section 6.2)

FDSN publishes seismology standards both for data and metadata formats (ensuring Interoperability), for data and metadata delivery services (ensuring Accessibility, International Federation of Digital Seismograph Networks, 2013), and maintains a registry of data centers that use these standards (ensuring Findability). Such standardization is key, because the evolving data sets are stored in and distributed by regional, national, and international repositories, and users access them through multiple well-defined access methods. The main services for requesting metadata and data are the FDSN web services fdsnws-station and fdsnws-dataselect (https://www.fdsn.org/webservices/), adopted in 2013. These services ship metadata and data based on user-defined requests. Users can access data via web interfaces such as interactive data center or federation webpages; however, most users access data via smart clients and programmatic interfaces (e.g. Python, Matlab). Open-source seismic data libraries and tools like ObsPy (Beyreuther et al., 2010) play a major role in supporting seismic data usage and increasing interoperability and accessibility.

FAIR for seismological waveform data is in this way partly achieved largely through FDSN standardization, since 2014 combined with network DOIs, which improve findability and meet reuse conditions through complete metadata and licence information. As discussed below, using DOIs offers many advantages, and can increase FAIRness considerably. FDSN has historically given a high priority to findability and accessibility; a short analysis of all aspects of FAIR with respect to FDSN standardized data, metadata, and services is presented in Section 6.2.

DataCite metadata and FDSN standardized metadata have complementary functions. DataCite metadata enables globally standardized ways of identifying a single digital object across multiple domains. FDSN metadata contain seismology-related attributes, making it possible for the user to explore the data sets and reuse data.

2.1 Data identification and seismic network description at FDSN

FDSN has defined a unique data stream identifier, known as the Source Identifier (https://docs.fdsn.org/projects/source-identifiers/en/v1.0/). It is composed of four hierarchical elements: network, station, location, and channel, the latter subdivided into band, source, and subsource. A data center can manage, and a user can request, data based on this identifier and a time window through a variety of services, and in particular the standardized FDSN web services. Depending on service implementation, geographical coordinates (from metadata) can also be used to constrain searches for data, either from a station or a seismic event (earthquake) perspective, or a combination of both.

The highest hierarchical element of the Source Identifier is the "network" code, which is assigned by FDSN upon request, usually by a network operator. FDSN maintains an openly accessible directory of networks at https://fdsn.org/networks/, where each network page provides high-level information on the network, including a description, citation and data access information, and a map of stations belonging to that network. That information is provided and can be updated by the network operator. A *network* can have any number of *stations*, each station can have several instrument *locations*, and many different data types (some non-seismic), that can be described in technical components of *channel* (band, source, sub source). Figure 1 shows the seismic stations available through the FDSN standard metadata webservice (see Section 2.3) using the FDSN datacenter registry (see Section 2.4).

Figure 2 shows the evolution of the number of FDSN networks over time. We observe that there were more permanent networks than temporary networks until 2002. Since then, new networks are mostly temporary experiments. For example, out of 98 new network codes created by FDSN in 2023, only 7 (7%) were permanent networks. For comparison, FDSN created 82 network codes in 2014, out of which 12 (14%) were permanent. The time evolution of DOI and licence uptake presented

were developed with funding from the NSF to the Incorporated Research Institutions for Seismology (IRIS). After 2018 the NSF funded IRIS to operate the Seismological Facility for the Advancement of Geoscience (https://www.earthscope.org/about/gage-sage-facilities/). IRIS operated this facility until 2023 when it merged with UNAVCO to become the EarthScope Consortium, which now operates the SAGE facility.

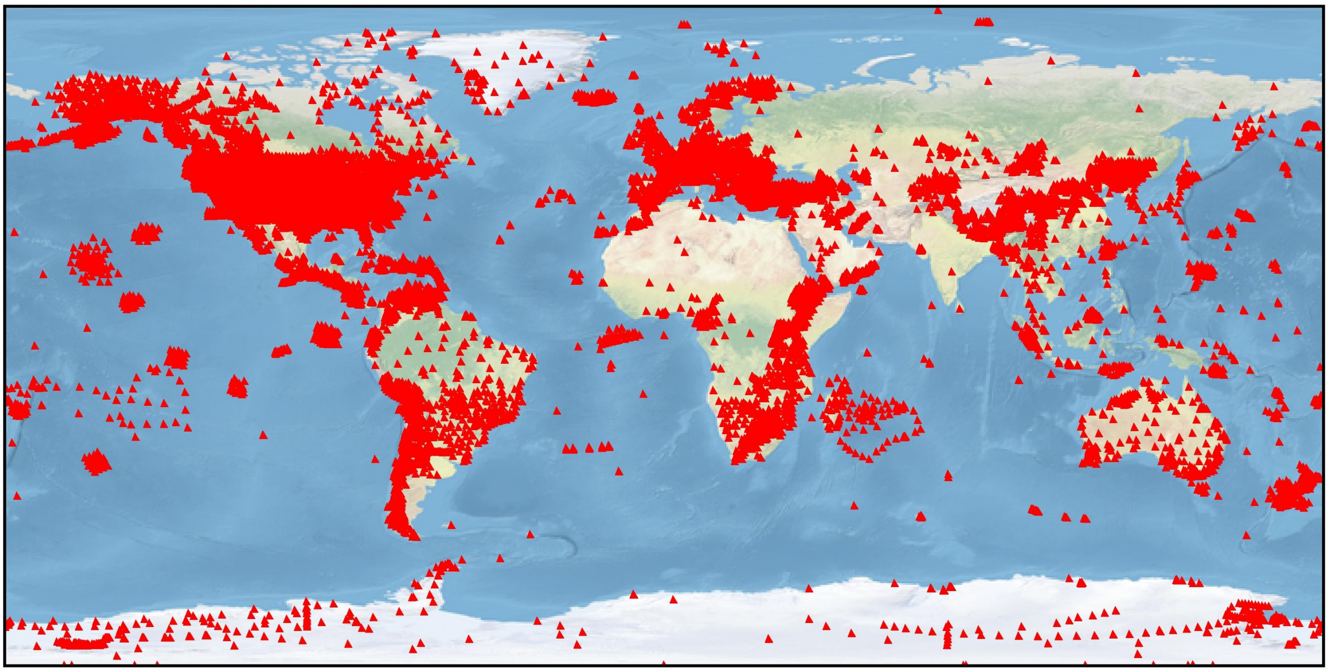

Figure 1 Stations of all FDSN registered networks. The map is limited to stations for which the station metadata can be obtained by the fdsnws-station webservice (see Section 2.3) using the FDSN datacenter registry (see Section 2.4).

in Section 3 must be seen in light of this varying rate of the weight of temporary networks, shown in Figure 2. Because FDSN keeps track of the network start year, both for temporary and permanent networks, we can track the degree to which DOIs have been attributed to networks created before FDSN adopted the DOI recommendation.

2.2 DOIs for seismic networks: granularity, scope, and implementation

For seismological waveform data, the DataCite DOI is a globally unique, persistent, and resolvable data set identifier, and, while useful for non-specialists, it is not essential for data findability for seismologists (see below). Rather, the DOIs fill important gaps in the original data distribution system. In the most recent (2023) FDSN recommendations for a DataCite DOI, it is stated that "[...] In our view, a seismic network is an entire collection of sensor data, including the seismic metadata associated with it, such as station details, instrument types, and response data." (International Federation of Digital Seismograph Networks, 2023).

It quickly emerged that the DataCite DOI metadata schema was well adapted to recognize the organizations and people contributing to the data sets, and is globally used across multiple disciplines beyond seismology. Therefore, FDSN (Evans et al., 2015) provided recommendations about harmonized DataCite metadata for seismological networks, originally based on the DataCite Metadata Schema v3.0. These recommendations were updated in December 2023 with improvements and based on a recent version of the DataCite metadata schema (v4.4). For example, https://api.datacite.org/dois/10.15778/RESIF.FR provides institutional information about the organizations contributing to the network data as 'DataManager', 'DataDistributor', 'HostingInstitution', 'DataCurator', 'DataCollector', and 'Sponsor'. FDSN recommends since December 2023 that contributors should be identified by means of persistent identifiers, such as, for example, ROR (Research Organization Registry, https://ror.org/), or ORCID (Open Researcher and Contributor ID, https://orcid.org/). Through the DataCite metadata schema and associated application programming interfaces (APIs) it is relatively straightforward to link citations to organizations or individuals that contribute to the network operation and/or data curation and management. Therefore, it is relatively easy to provide visibility of contributors in a collective data production and distribution effort.

The DataCite metadata schema v4.4 also offers the opportunity to provide standardized keywords for cross-domain data findability and applications. Finally, while FDSN presently has no recommendation about licences, the present version of the FDSN Recommendations (International Federation of Digital Seismograph Networks, 2023) indicates how to include licence information in DataCite metadata. Licence information is key for the reusability of data, and for communicating to users about their rights and obligations when they use the data.

The original motivation for introducing DataCite DOIs was the need to cite seismological waveform data (Evans et al., 2015), providing attribution to the network and, therefore, to the contributing organisations, people, and funders. To facilitate citation in downstream work, FDSN operates a service (https://fdsn.org/networks/citation/) that returns the citation for a given network code or a list of network codes, based on the information contained in the DOI metadata. With the fun-

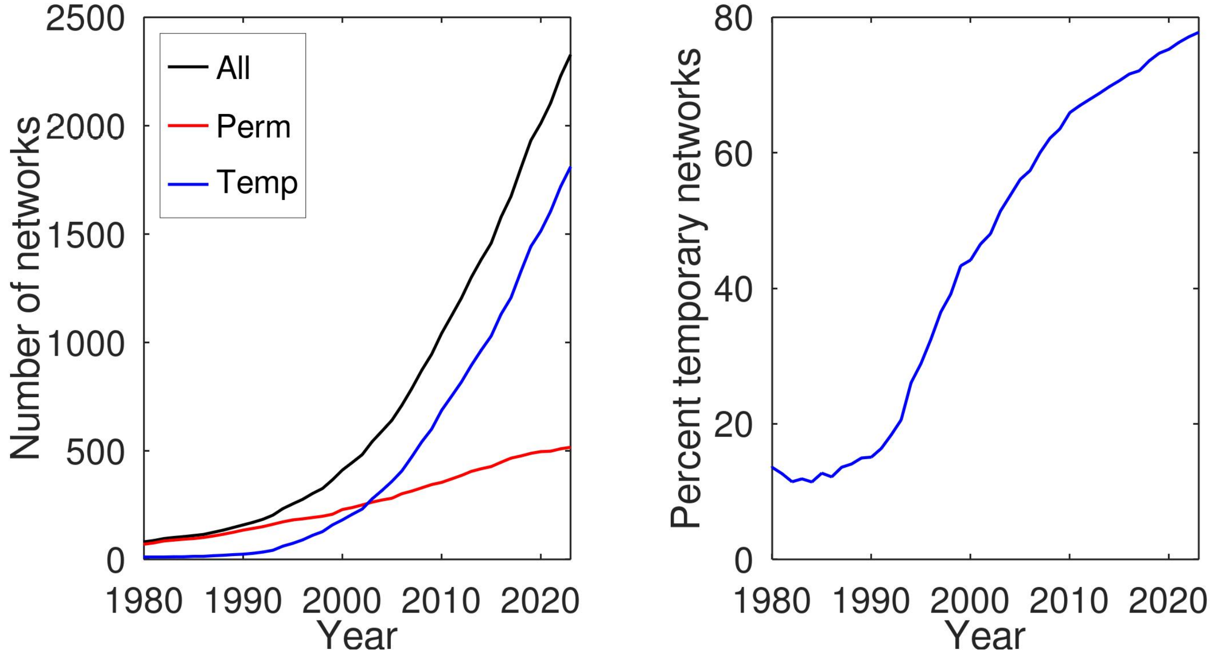

Figure 2 FDSN networks over time. Note that the number of networks created in a given year is given by the 'start year' of the networks as registered at FDSN. Left: Cumulative number of all FDSN network codes over time. Right: Ratio of cumulated temporary over cumulated all FDSN networks. Note that temporary networks only operate over a few years, so the total cumulated number also includes networks that are no longer operating but for which the DOI and data are still managed.

damental aim of promoting DataCite DOIs for data citation, FDSN recommends having a DOI for (waveform) data associated with a network rather than a scientific article describing the network. A network description in such a publication can be linked to the network DataCite DOI metadata as a 'related identifier'.

EarthScope (for seismology, formerly IRIS) hosts the global FDSN network and data center registries and web infrastructure. It has a DataCite DOI minting service for the network operators who wish to have their DOI minted by FDSN rather than, for example, through their institution or from a national service. The FDSN DOIs point to a landing page hosted at FDSN and they have a minimum number of associated DataCite metadata fields. By default, anyone who requests a FDSN network code also obtains a DOI from the EarthScope minting service. At present, the FDSN recommends DOIs for network level only, so organisations who wish to have DOIs or other persistent identifiers at finer granularity, such as station level, need to manage those DOIs independently of the FDSN.

With the implementation of FAIR principles taken up across the world, the DOI minting services are now capable of producing millions of DOIs in an automated way. The FDSN recommendation is somewhat orthogonal to this evolution because the number of seismic network DOIs remains low (at the date of analysis, approximately 2400 networks were registered at the FDSN) and DOIs neither identify each data point nor each seismic station, for example. At present there seems to be no immediate advantage for the FDSN to move to a finer level of granularity for DOIs (see also Section 6): with a DOI at network level, the systems for data access are sufficiently mature so that the Source Identifier is exploited to easily guide the user to the correct data, while the DOI complements the non-seismology metadata information and provides ease of citation.

2.3 FDSN standards for services, data and metadata

FDSN offers a standard for seismological data (miniseed) and metadata (StationXML) through the standardized web services 'fdsnws-dataselect' and 'fdsnws-station', respectively. Here we focus on the metadata format and the service 'fdsnws-station'; the full set of standards and service specifications are available at https://fdsn.org/services/.

The metadata format is called StationXML (https://docs.fdsn.org/projects/stationxml/en/latest/) and contains all the relevant information that the user needs for interpretation and further processing at the different levels of granularity (with the highest level being network). The present version of StationXML (schema v1.2, 2019) has the option to include a persistent identifier such as a DOI, allowing proper attribution and citation by smart clients and tools in an automated way. Prior to 2019, the existence of a (network) DOI could only be included in a comments field; the latest version of StationXML fills the gap of the connection between StationXML and persistent identifiers (e.g. DOI, Handle, ARK). As an example, the request for metadata for the French national broadband network, code FR, yields (among others) information about the network and the persistent identifier:

Metadata request

```
https://ws.resif.fr/fdsnws/sta-
```

```
tion/1/query?network=FR&level=network
```

High-level Network information (extracted fields)


```
  <Network code="FR" alternateCode="RLBP"
startDate="1962-01-01T00:00:00" end-
Date="2500-12-31T23:59:59" restricted-
Status="open">
  <Description>RESIF and other Broad-band
and accelerometric permanent networks in
metropolitan France</Description>
  <TotalNumberStations>196</TotalNumber-
Stations>
  <Operator>
  <Agency>Réseau large-bande permanent
(RLBP)</Agency>
  <WebSite>https://rlbp.resif.fr</Web-
Site>
  </Operator>
  <Identifier type="DOI">10.15778/RE-
SIF.FR</Identifier>
```


As seen in this example the metadata request is carried out through the standardized web service fdsnws-station. For full documentation, see https://fdsn.org/webservices/fdsnws-station-1.1.pdf.

2.4 FDSN data center registry and data routing

FDSN has historically kept a registry of seismological data centers that distribute data with FDSN standard services and using FDSN data and metadata formats (https://fdsn.org/datacenters/). A seismic network can be registered at the FDSN without data distribution being in place. However, the majority of networks have data distribution through one or more data centers.

The most recent evolution of the data center registry is an API through which it is possible to obtain, for each data center, all the services that they support, the networks for which data is distributed, and an associated priority for each network. The priority establishes the 'authoritative data center', in case network data is hosted by several data centers. It is then possible to route data requests to the proper data center. For example, a network operator may allow several data centers access to real-time data streams for alert systems, while the highest-level quality data from that same network (gaps filled, metadata more frequently updated) may only be available in the authoritative data center. The routing also makes it possible for smart clients to seamlessly route data requests for large networks whose data are distributed across many data centers, with data distribution defined at station, or even lower, level.

2.5 The user experience: where to find and access data and metadata

The users have three main entry points for findability:

1. FDSN: users can interactively consult the data center registry, network descriptions, citations, etc. via the FDSN web pages. They can, for example, search for networks and obtain information about which data centers distribute the data. One can get this information either via the FDSN network webpage or via the FDSN data center registry API. FDSN also has an API to inform on past and presently used network codes.

2. Data center: users can request data and metadata directly from each data center, either through standardized FDSN web services or through data center specific web pages.

3. Federated level: Finally, the data center registry at FDSN makes it possible to build routing tools so that the user does not need to identify which data center distributes (all or parts of) a data set. There is also a strong collaboration between European data centers (EIDA, European Integrated Data Archive) that use data center registry information (EIDA Routing Service) to route the user data and metadata requests of European data to the relevant data center (Quinteros, 2017). Such tools can be either webpages (such as EIDA webdc3, https://orfeus-eu.org/webdc3/; EarthScope's MetaDataAggregator, https://ds.iris.edu/mda/); federated APIs (e.g. EIDA Federator, https://www.orfeus-eu.org/data/eida/nodes/FEDERATOR/; EarthScope's fedcatalog, https://service.iris.edu/irisws/fedcatalog/1/); or software built on top of these APIs, like Obspy routing clients (Beyreuther et al., 2010) and fdsnwsscripts (Heinloo, 2024).

At the time of writing, most users obtain metadata and data through programmatic tools either at the data center or the federated level.

2.6 Workflow of the assessment

To analyze the uptake of DOIs and licences, we developed a workflow that is fully automated and based on the building blocks described above. This analysis workflow is composed of the following steps:

1. Recover from FDSN all network codes and, if present, the associated DOIs.

2. For each network:

    - If there is a DOI indicated at the FDSN: recover licence information, if available, in the DataCite metadata.
    - Identify via the FDSN data center registry API the data center in charge of Priority 1 data distribution for the network and recover the URL of the fdsnws-station service.
    - If a data center is identified: test if the community metadata (StationXML) can be successfully retrieved.
    - If there is community metadata: test for the presence of DOI and retrieve if present.

Note that in the analysis for this study we used the EIDA routing tables for European data centers rather than the FDSN data center registry in the 2nd bullet of step 2, as we realized at that time that the relatively new FDSN data center registry still had some issues. At the time of writing, these issues are fixed, and EIDA routing and FDSN data center registry are coherent.

Some additional steps were carried out for consistency checks. This included keeping track of network start dates (start of permanent network or temporary deployment) from three different sources: FDSN, which indicates only the start year at FDSN; the FDSN data center registry, which indicates the start date of routed data; and StationXML metadata, which approximately indicates start date of actual data holdings. While those dates can be different, they need to be consistent. We also checked whether the DOI for each network was identical from the two different sources where this information is available: FDSN and in the network level StationXML. In case of temporary deployments, the end date is retrieved as well.

3 Results

In this section, we present the outcome of the assessment, carried out in April 2024, across all of the FDSN-registered seismic networks. We also subdivide the results into three subgroups (Earthscope, EIDA data centers, and 'Other'), to see if the geographical location of the data centers influences the statistics. We verified the consistency of results (see dedicated section) and concluded that only a small percentage of networks (~4%) had demonstrable errors or inconsistencies. Therefore, we consider that our analysis is qualitatively meaningful, and only has small quantitative errors.

3.1 Overall network statistics

Table 1 shows the total number of networks registered with FDSN, and how they are distributed across the three subgroups, as well as the number of networks not associated with any data center as indicated in the FDSN data center registry API (see Section 2.4).

At the date of analysis (18th of April, 2024), 2368 seismic networks were registered at the FDSN. 544 (23%) of these networks are permanent, and 1824 (77%) are temporary (see also Figure 2). The network statistics for temporary experiments are strongly dominated by large national mobile instrument pools in the USA, Germany, and France, which have an open distribution policy of data from experiments that use these pools. The waveform data from these data sets are typically embargoed from open access for 2-3 years; however, the metadata is openly available during such embargo periods.

The majority of FDSN-registered networks distribute their data openly through an FDSN data center: Of the 2368 networks, 1676 networks (71%) are associated with at least one data center in the FDSN datacenter registry. For the remaining 694 networks (29%), we can identify three potential causes for why no data center is identified: a) data is not distributed; b) data is distributed through a data center that does not use FDSN standard services; c) information is missing in the FDSN data center registry.

1213 networks (51%) have EarthScope (IRISDMC) as the authoritative data center, while 394 networks (17%) declare an EIDA center as authoritative. Consequently, Earthscope (IRISDMC) and EIDA jointly cover 68% of FDSN networks, and 96% of all networks for which a data center was identified. The dominance of EarthScope (IRISDMC) and EIDA in the data distribution makes it meaningful to do a comparative study between them to assess the DOI and Licence uptake. This comparison can illustrate the consequence of different strategies and resources dedicated to FAIRification of data. Note also that EarthScope and EIDA, upon agreement with the network operators, distribute data from many networks outside the USA and Europe. The remaining data centers each distribute data from very few networks, so a statistical analysis over any of these remaining data centers may not be meaningful.

Of the 1674 networks with an associated data center, 96% had a successful response to a metadata request for a Network level StationXML file. The failed responses for the remaining 4% are linked to a technical issue at a single data center; this problem is presently being resolved. We, therefore, consider that the StationXML distribution at Network level is operating smoothly if a data center is indicated at the FDSN data center registry for that network. Testing the data distribution services goes beyond the scope of this paper, which focuses on the uptake of DOI and licences across FDSN. Based on our experience as EarthScope and EIDA node managers, however, we note that data distribution services are operating properly and are monitored closely.

We, therefore, confidently conclude that a majority of the 2368 (April 2024) FDSN-registered networks effectively distribute part or all of their data according to community standards.

3.2 DOI uptake

Table 1 also shows the number of FDSN-registered networks associated with a DOI. 1732 networks, i.e. 73% of the FDSN networks, have a DOI that is exposed at the FDSN. The coverage is approximately the same at EarthScope (IRISDMC) and across EIDA nodes, respectively 82% and 77%. Some of the remaining 636 networks (27%) may have a DOI, but we have no programmatic access to that information if the information is not available at the FDSN.

Figure 3 illustrates the increase of the number of networks, DOIs, and licences over the past 4+ decades and the percentage of networks with DOIs as a function of network start year. The percentage (right panel) is calculated over the cumulated number of networks, so the increase in DOI coverage in recent years is related to high DOI coverage for recently created networks. More than 80% of the networks created after 2014 have a DOI.

A plausible explanation for this high coverage is that network operators (which can be project PIs in the case of temporary experiments) who request a network code at FDSN are offered, by default, an FDSN DOI minted by EarthScope (IRISDMC), with landing page maintenance at the FDSN. This provides an instant solution for network operators to generate a DOI for their networks, temporary networks having dominated network code requests since the 2014 recommendation (Figure 2). Combined with the fact that Earthscope (IRISDMC)

Data center or organization	No of networks	No of networks with DOI	% data center networks with DOI	No of networks with licence	% data center networks with licence
Earthscope (IRISDMC)	1213	991	82%	10	1%
EIDA	394	303	77%	152	39%
Other	67	64	96%	6	9%
None	694	374	54%	20	3%
FDSN	**2368**	**1732**	**73%**	**188**	**8%**

Table 1 Summary of DOI and licence uptake across FDSN. The total FDSN numbers are in the bottom line. The four lines above show total number of networks, number of networks with a DOI, and number of networks with a licence across different organizations: Earthscope (IRISDMC), EIDA nodes, Other data centers. 694 networks did not have a data center identified in the data center registry ('None'). For each category, the percentage is calculated over the number of networks indicated in the same line. For example, 991 out of 1213 networks hosted by Earthscope (IRISDMC) have a DOI, corresponding to 82% of the Earthscope (IRISDMC) hosted networks.

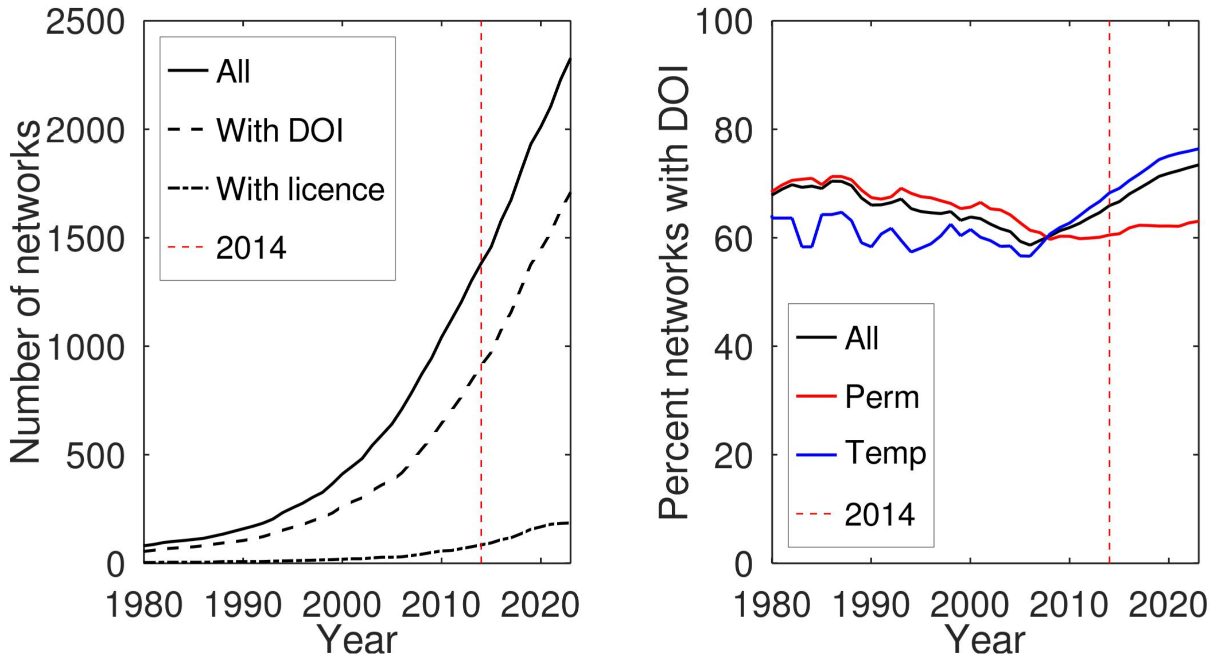

Figure 3 Evolution of DOIs and licences for FDSN networks as a function of network start year. The thin dashed red line indicates 2014, i.e. the year of the FDSN recommendation on DOIs. Left: total cumulated number of FDSN networks (solid line), networks with a DOI (dashed line), and networks with a licence (dotted-dashed line). Right: fraction of the cumulated networks with a DOI for all FDSN networks (black), permanent networks (red), and temporary networks (blue). The percentage is per category, i.e. number of DOIs for temporary networks divided by number of temporary networks. Note that the contribution of permanent and temporary networks to the statistics for all FDSN networks varies over time (see Figure 2).

hosts the largest number of temporary experiment data, the process is coherently handled internally. The additional two countries with large mobile pools and many temporary experiments (Germany, France) also have a strong DOI policy and management. This also explains the high uptake of DOIs for temporary experiments from before 2014, although associating past experiments with DOIs may mean protracted outreach to past project PIs or organizations in charge of the experiments. There are two main reasons for missing DOIs. The first one is that it is difficult to reach all former temporary network operators or PIs (some of them no longer active) to obtain permission for minting the DOI and to obtain relevant project information to provide relevant DataCite metadata. This raises the question of how to organise datacenter rights to avoid difficulties in the future, such as the update of DataCite metadata for an experiment with an unreachable or non-responsive PI. Second, new experiments that do not rely on FDSN (Earthscope) minted DOIs, may have a lag time after network code creation before minting a DOI, for example waiting until actual project start before creating a permanent identifier.

Also, FDSN and the seismological community have made a substantial effort to create DOIs for permanent networks, in spite of challenges. Indeed, permanent network operators may have strong institutional, regional or national constraints to mint and maintain the network DOI. In addition, many permanent networks are operated by several organizations, or national consortia. Agreement from each of them is needed for the formalisation of their contributing role(s) within the

DataCite metadata, which can slow the process of creating a DOI. This might explain why new permanent networks have a lower percentage coverage of DOIs than older ones.

Considering all these challenges, the uptake of DOIs for seismic networks since the 2014 recommendation is clearly a success, being supported by the network operators and by the data centers hosting the network data.

Table 2 shows the distribution of DOIs in terms of minting organization. 99% of the minted DOIs are covered by only 5 organizations, and FDSN-minted DOIs are largely dominant. Overall, the strategy of FDSN to systematically offer a DOI together with the maintenance of the associated landing page is, therefore, an excellent strategy for DOI uptake.

Since the StationXML schema was updated in 2019 (https://docs.fdsn.org/projects/stationxml/en/latest/), there is a specific field for the DOI in the StationXML metadata. We, therefore, also checked whether the DOI was indicated in the StationXML in this specific field because including the DOI in the metadata directly informs the user about correct citation of the network data, and can easily be used by the data centers internally. This check could be carried out only for networks for which a data center was identified. We observe that 66% of FDSN-registered networks with a DOI have a link to that DOI in the StationXML metadata. This corresponds to 85% of networks with a DOI and for which a data center was identified. Therefore, the link between DOI and metadata is effective for a majority of networks with a DOI.

StationXML updates to include the network DOI are, however, significantly different between EIDA (71%) and EarthScope (94%). After further analysis, the low percentage for EIDA stems from missing DOIs in two of the EIDA nodes; the problem is presently being addressed. The missing links for EarthScope were almost exclusively data sets for which no metadata and data were in holding, some of them being ongoing experiments for which metadata and data have not yet been shipped.

3.3 Licence uptake

The latest version of the FDSN Recommendations includes detailed instructions about how to provide rights information in DataCite metadata. This is also required to meet the 'R' (reusability) of FAIR. In practice this means indicating a licence and/or embargo or other kind of restriction information. Licences may impose (admittedly without the community having the means to enforce nor monitor the respect of the licence conditions) the reuse, data citation, and user-defined conditions, and protect the data producers in terms of liability in case there are errors in the data or metadata. In principle, different licences could be applied to data and metadata, and discussions are ongoing whether metadata should be 'public domain'. At the moment, however, the FDSN standard for DataCite metadata concerns only the licence for the data, so we do not address metadata licences here.

Table 1 shows the number of networks for which a licence is indicated in the network DOIs. The licence information is recovered as part of the DataCite metadata. Only 8% of FDSN networks have a licence, which is 11% of FDSN networks with a DOI (i.e. with the possibility of retrieving a licence). This low number has several causes. First, FDSN has not yet implemented an easy-to-use tool to include licence information in the DataCite metadata. Second, there is no agreed community recommendation to have a licence, nor about the recommended licence(s) to use. Third, national legislation is diverse, and regional, national, and international institutions are progressively building their framework for international research data licensing. Finally, there can be a disconnection between the researchers producing the data and the institutional framework, both in terms of knowledge sharing and of actual rights that the researcher has in terms of indicating a licence for a data set.

There is a very significant difference in the statistics between the data sets hosted by EIDA and the rest. Only 36 networks (2.5%) of the networks hosted by a data center not belonging to EIDA have a licence. On the contrary, data from 152 networks hosted at EIDA are distributed under licence. This means that 81% (in terms of the number of networks) of data sets with a licence are hosted by EIDA. The successful licence uptake in Europe is related to a concerted effort across EIDA, and to the EPOS data policy. Based on this policy (https://www.epos-eu.org/sites/default/files/2024-03/EPOS%20DATA%20POLICY_July2018.pdf), EIDA has recommended European networks to use the Creative Commons CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/deed.en), which is also internationally emerging as the de-facto standard: it allows the users to freely use the seismological waveform data for any kind of further processing (e.g. mixing with other types of data), while legally protecting the data producers and requesting acknowledgment for data use in any downstream publication (any type of published material, not only scientific articles).

Of the 188 networks with a licence, 184, i.e. 98%, are using a CC BY 4.0 licence. The adoption of some other licences, like a 'share-alike' CC BY-SA, may be problematic because it implies a constraint on the adoption of a licence for derived products. The same could happen with national, institutional or service-specific licences. Scientists often use data from several networks and across national boundaries, and include data products (seismological, geodetic etc.) in the interpretation. If even a small part of the data used for the creation of a new data set had a CC BY-SA licence, the derived product must have the same licence, as it is the more restrictive (CC BY alone would not be allowed).

The experience in Europe has helped the seismological community gaining knowledge about licences, and, despite the difficulties mentioned above, the need to consistently apply licences is widely acknowledged in the community. Over the past years, the licence discussion within FDSN has matured, and FDSN is presently considering a recommendation for licences of seismological waveform data and metadata. FDSN

Minting organization	Number of DOIs minted	% of DOIs
FDSN (Earthscope/IRISDMC)	1481	86%
GFZ	118	7%
RESIF (now Epos-France)	74	4%
ETH	25	1%
INGV	19	1%
Others	15	1%
Total	1732	100%

Table 2 Number of seismic network DOIs from different minting organizations

is expected to recommend CC BY or CC0 (a 'public domain declaration', https://creativecommons.org/publicdomain/cc0/) as standards for seismic networks.

4 Observed inconsistencies

As indicated previously, there are three sources of information for metadata and data access: DOIs, StationXML distributed by the data centers directly, and federated access to multiple data centers using the FDSN data center registry. Consequently, the user has several paths to data discovery and access. Ideally, access paths and sources of information deliver consistent results. We, therefore, carried out some simple consistency checks, with the aim of solving errors when possible. The number of inconsistencies is small and corrective actions are already being taken for most of them. Below we highlight some of these inconsistencies for future reference in FDSN and for network operators, data centers, and other communities to be aware of.

4.1 DOI inconsistency between FDSN and StationXML

We found 16 data sets (<1%) with this type of inconsistency. The cases listed below demonstrate the variety of problems that can arise in consistently informing about the DOIs in different systems, let alone checking coherency between the DataCite DOI metadata and StationXML content. The observed inconsistencies should be used as input on how to further strengthen the FDSN DOI system to avoid accumulation of errors over time.

- Approximately half of DOI inconsistencies happen at the moment of request of a network code, where by default FDSN mints the DOI. Researchers can inadvertently request an FDSN-minted DOI for a seismic network even if a DOI already exists or will be minted by a national or local organization. In this case, the FDSN minted DOI then needs to be dereferenced and the network information at the FDSN updated with the correct DOI. The inconsistency is likely to not be detected by the person requesting the network code, so inconsistencies may endure until either the data center or network operator discovers the problem. This is a small downside to FDSN creating a DOI by default, and this effect is amply counterbalanced by the high DOI coverage.

- Typos in the DOI names indicated at FDSN, when other organizations mint the DOI. The DOI information to be uploaded to FDSN is manual, so typos are not detected.

- Two different network codes with the same DOI (2 cases). For these cases, two network codes were used because a temporary network operated for longer than foreseen in the initial FDSN temporary network code, and that code was already assigned to a different experiment in the following period. Hence a second network code was attributed to the continuation of the extended experiment. In this case, two network codes sharing a single network DOI makes sense from a scientific and citation point of view. However, the DOI information in this case does not lead the user to all the data because the landing page of the DOI is associated only with one of the data sets. FDSN should consider elaborating further recommendations on such cases.

- One temporary network declared at FDSN was a clone (all information, including DOI, being identical) of a national French network. A tentative explanation is that someone who wished to download data from the network by mistake requested a network code instead, and copy-pasted all the network information. It is an open question whether other such erroneous networks exist at FDSN, and how it would be possible to identify them.

4.2 Data center identified but StationXML request failed

The standard case for an identified data center but unsuccessful StationXML request would be if the data center was temporarily down. A couple of unexpected additional cases also appeared. These inconsistencies practically impact the accessibility (i.e. the 'A' in FAIR), so whilst inconsistencies are not unexpected in an operational environment, we shortly report on the causes to help minimize similar cases in the future:

- One data center had implemented non-standard URL naming for their web services, meaning that their data was effectively inaccessible for most programmatic tools. This problem was fixed within a few hours of notification, but had gone undetected, probably for months.

- For one data center, there were some routing issues, currently under investigation.

- Some networks had incoherent information between start and end dates in the StationXML file or in the FDSN data center registry. We also discovered some StationXML files with other errors.

4.3 Data center not identified but should possibly have been

For this category we identified 20 networks (around 1%) with a DOI and a (open) licence, but for which no data center was identified in the data center registry. That would violate the assumption that an open data licence is associated with effective data distribution. We analysed these cases to understand how such inconsistencies occur, but did not aim at a full search for all such inconsistencies across FDSN.

On this small subset of networks, we searched for data distribution information at FDSN (manually checking the network information at the website), which is independent of the FDSN data center registry, and we checked whether it was possible to obtain a StationXML file at network level directly from the data center indicated on the network page at the FDSN. This qualitative and manual search highlighted some causes of the apparent lack of data distribution:

- Errors in the FDSN data center registry (presently fixed)
- Inconsistencies in start and end dates in either the StationXML files or the routing as compared to the network start year at FDSN (many having been fixed since the analysis was carried out)
- Data center choices: one data center exposes StationXML at data center level for networks for which they don't have the right to distribute data, but do not include the networks in the routing. Other data centers choose to not expose any metadata for which they have no rights to distribute data. While either case can be argued, it is difficult to expose these subtleties and variability to the user who would gain from a harmonized strategy for these cases across FDSN.

5 Citation of seismic network data

Seismic networks and data centers are under pressure to justify their operational costs, for which demonstrating usage is a main criterion for funding. The initial objective of the DOIs for seismic networks, ease of identifying data usage and thus fostering proper attribution of data used in further studies and publications, is, therefore, as critical now as it was in 2014. Citing is not only essential for the network operators, project PIs, and data centers, but is becoming increasingly important as the proportion of data sets with a licence requiring attribution increases, thereby engaging the legal liability of authors and publishers. There is reciprocal high value for scientists because improved data set citation will be reflected in ORCID and other key citation trackers. FDSN and individual data centers have dedicated significant efforts to reach out to publishers and the scientific community.

This section first reminds the reader about how to cite seismic network data and then presents an initial analysis of citations of seismic networks and difficulties encountered.

5.1 How to cite data from seismic networks

Data from seismic networks that have a DOI should be cited in the same way as scientific research articles: short reference in the main text (in the main body of the text, or in a dedicated Data Section or Data Availability Statement) and long reference in the Reference list. The citation format of the scientific journal should be respected so there may be variants on the examples below.

For example, the use of data from the Seismic network of the Republic of Slovenia would be cited *Slovenian Environment Agency (1990)* as the short reference and in the reference list as: Slovenian Environment Agency. (1990). Seismic Network of the Republic of Slovenia [Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/SL.

The citation can be found either on the network page at FDSN https://www.fdsn.org/networks/detail/SL/ or through the FDSN service to request citations: https://www.fdsn.org/networks/citation/. This FDSN citation service provides multiple citations based on network codes. It is, therefore, easy to retrieve citation information for a large number of networks.

Examples of correct citation:

1. An article with numerous networks with short citations in the text, table and figure captions; in the Data Availability Statement; and full citations of each network is included in the References : Marignier et al. (2024)
2. An article citing many networks, with short citations in Table form (see Table 1 of the article), and citations included in the reference list: Pedersen et al. (2023)

5.2 Uptake of citations in publications

In the course of this work the authors started by comparing citations through different tools (Scopus, WoS, Cross-Cite, Google Scholar) across all FDSN minted DOIs (text string 10.7914/SN, covering 77% of FDSN network DOIs). The outcome was in all cases largely incomplete and incoherent between the different methods. We concluded that there is presently no way of reliably obtaining a comprehensive citation report for a large set of DOIs across peer-reviewed scientific journals. We therefore focus this section on some illustrative examples.

Figure 4 illustrates that network citations, whilst much lower than reality of network data usage, have been increasing since 2015 . The citations used in Figure 4 were obtained by searching the text string 10.7914/SN (covering 77% of FDSN network DOIs) across the full article texts of Geophysical Journal International and all the journals of the American Geophysical Union. The choice of these journals is partly due to feasibility, partly because many articles based on seismic network data are published in these journals, and partly linked to a particularly proactive approach from these journals and publishers concerning data citation (e.g. Stall et al., 2023). The full-text search was used because many articles indicate the DOIs in the main text or in dedicated Data Sections without including the citations in the reference list. It is possible on Geophysical Journal International to search specifically in the references. Overall, approximately two-thirds of the articles in Geophysical Journal International that indicate the DOIs in the main text also refer to the network DOI in the References. Geophysical Journal International has explicit author instructions since 2019 about how to cite seismic network data. Journals with less clear or more generic author instructions likely have a lower correspondence between DOIs in the full-text and in the references. This was tentatively confirmed by manual spotchecking in several other journals, including AGU journals. These manual checks also showed that a significant proportion of articles based on seismic network data miss the reference to the DOI altogether.

We also tried to use Scopus to search for references in Geophysical Journal International and in AGU publications. We only retrieved a partially overlapping subset of the references we identified on the journal websites, with numbers on Scopus being lower than those on the journal searches. Another option is to search for references via the DataCite services. This search yielded 1780 citations across all publication types (therefore citations in international peer-reviewed journals compose only a subset of this number) and without the yearly distribution. Looking into the details of the DataCite citation results, we observed a strong mismatch in many cases, where we had found significantly more citations for a network directly on the publisher sites than what appeared in the DataCite search. While we could not produce a citation report across all of the FDSN DOIs from Web of Science, the total number of citations with the string 10.7914/SN up to September 2024 yielded 2832 citations.

In spite of these various difficulties and overall low citations numbers, it is clear that citations are increasing both in quantity and in quality, demonstrating the, at least partial, success of the FDSN DOI recommendation for better data citation. One issue that is currently unresolved is the limit on the number of references currently imposed by some journals. This means that it may not be possible for some studies to cite all contributing datasets, for example when they are built on large and specific selections of data over space and time. Listing the references in Supplementary material or in Data Sections only is not an adequate solution presently, as the citations can not be found, and are not referenced in the citation services. While there is no obvious solution available yet, not citing the data is certainly not the solution. Relevant discussions are currently under way in various initiatives (e.g. RDA Complex Citations Working Group), and a number of alternative ways to address this issue may emerge.

6 Discussion

6.1 General outcome

Our assessment demonstrated the great success of the FDSN DOI recommendation from 2014: a large majority of FDSN-registered seismic networks are associated with a DOI. Significant success has been achieved in terms of attributing DOIs not only to new networks but also to networks predating the DOI recommendation. This means that the seismological community (organizations, network operators, data centers) has clearly understood and engaged in the process. Also, the level of DOI uptake is similar between Earthscope (IRISDMC) and EIDA-hosted network data.

The main causes of success reside in the capacity of FDSN to develop and update recommendations, the default attribution of DOIs by FDSN to new networks and internally hosting landing pages for FSDN (Earthscope) minted DOIs. Also, the choice of a high-level granularity (network) of the DOIs means that the resources required by individuals to create DOIs with rich metadata are manageable.

It is already challenging to maintain the internal coherency and updates of DataCite metadata. We question whether a finer granularity of the DOIs is possible without the use of a disproportionate amount of resources to maintain and continuously enrich the associated DataCite metadata.

The granularity of the DOIs also makes it easy for most studies using seismic waveform data to include network citations. The number of correct citations is effectively increasing, with the cumulated network citations in 2024 being counted in thousands. However, the network data usage is not yet fully reflected in citations, due to either lack of citations or incorrect citations by researchers. It is also still challenging to obtain a citation report for a large set of DOIs. A small number of seismological studies that use a very large number of networks may have difficulties to include all citations in the reference list, but these issues may be solved with ongoing work from, for example, DataCite and the International Research Data Alliance (RDA) working groups (e.g. the RDA Complex Citations Working Group, Agarwal et al., 2025).

Licensing of seismological data is emerging across FDSN networks, with a strong increase in Europe but not yet in the USA or elsewhere. The vastly dominating licence is Creative Commons CC BY 4.0. The European uptake and the use of CC BY 4.0 was also motivated by ORFEUS's connection to EPOS, because EPOS participating countries have agreed on a data policy, and because the scientific communities were engaged in the preparatory discussions for this data policy. Including licence information is a key element for data reuse, and, importantly for network operators and data centers, a licence can clarify (no) liability in case of errors in data or metadata. The licencing also means that the scientific journals have a strong legal incentive to improve seismic network citations, both in quantity and in quality.

We can easily measure the ongoing growth in both DOIs and data licences sampling over a short period.

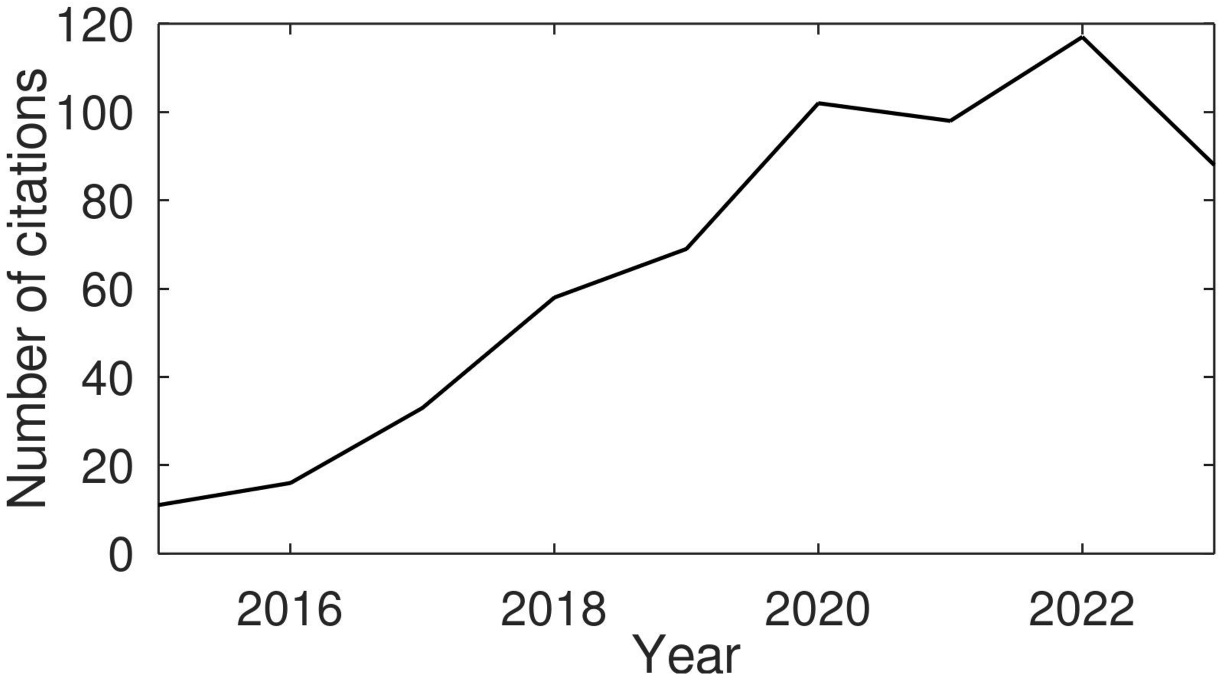

Figure 4 Number of citations for a subset of DOIs and a subset of scientific journals. We used a total of 1342 DOIs for which the prefix could be easily grouped as search strings with wildcards, therefore, covering 77% of the FDSN networks that have a DOI. We searched these strings on the website of Geophysical Journal International (https://academic.oup.com/gji) and across AGU publications (https://agupubs.onlinelibrary.wiley.com/). In both sets of searches, the strings were searched in the full-text rather than in references only. The citation numbers add up to 728 across that subset of DOIs for the two journal's full-text and all years.

Comparing the analysis from April 2024 with a repeat of the same analysis on August 16 2024 shows an increase in various metrics across FDSN (Table 3). The number of FDSN-registered networks has increased by 4%, Out of these networks, 98% are covered by a DOI and 20% by a licence. Overall, Table 3 demonstrates that the efforts are ongoing and still successful.

We assessed not just the DOI uptake, but also the coherency of information across the FDSN framework (data center registry, coherency between the DataCite DOI and the DOI in seismological metadata, effective metadata download etc). While the number of identified errors and inconsistencies was low, the analysis showed that significant efforts must be dedicated to the maintenance of data findability and accessibility, and to the coherency of DataCite DOI and seismological metadata and data distribution information at FDSN. Several of the issues identified in April 2024 were fixed during the preparation of this article, such as incoherencies with the EIDA routing and the FDSN data center registry, with DOIs and services. We also started quantifying and better understanding where different data center policies could be better harmonized.

6.2 Further analyzing the results in a FAIR setting

The discussion of introducing identifiers for seismological data started in the FDSN context around 2012, with the first version of FDSN recommendations for DOIs for seismic networks emerging in 2014 (Evans et al., 2015). The recommendation was mainly driven by the need to improve citations and attribution for seismic networks and their operators in the published scientific literature. In Europe, it was also motivated by the preparatory work to establish EPOS, where proper identification and licensing of assets was a topic from the beginning. DOIs for seismic networks, and also the first version of standardized FDSN web services (International Federation of Digital Seismograph Networks, 2013) thus predate the publication of the FAIR principles (Wilkinson et al., 2016) by a few years, but they clearly relate to the same objectives.

One core aspect of FAIR, machine-actionability (see definition by Wilkinson et al., 2016), was arguably not given too much conscious thought in those early discussions, likely due to the already well-established data and metadata standards for seismological waveform data that allowed easy Accessibility to and seamless Interoperability of data(sets) within the community. The StationXML metadata are however not machine-actionable in a formal FAIR framework as it contains mostly in-situ fixed content with few external references for a machine to expound further on sources and methods for that content. Overall, the current community standards and practices well support Findability (data center registry, routing service, StationXML metadata providing among others location and timeframe) and Reusability (licence, even though uptake across the community is challenging, and there is yet no community-endorsed recommendation or standard). Addressing other FAIR elements will require further work, for example, to standardize provenance information and community

No of networks April 2024	No of new networks Apr-Aug 2024	% increase new networks	No of DOIs for new networks Apr-Aug 2024	% New networks with DOI	No of new networks with licence Apr-Aug 2024	% New networks with licence
2368	95	4 %	93	98 %	19	20%

Table 3 Increase of different metrics between April and August 2024

Recommendation	Who: Network Operators	Data Centers	FDSN Federation	Journals	DataCite	Citation Search	Scientists	Why: Findability	Accessibility	Interoperability	Reusability	Improves Citation	Attribution
Distribute data through an FDSN datacenter	X	X							X	X			
Request/Mint DOI and provide complete information for datacite metadata	X	X	X					X		X		X	X
Support network operators for management of datacite metadata		X						X	X		X		X
Monitor correctness and completeness of DOI metadata		X	X					X	X	X	X		X
Ensure consistency of information available from different sources		X	X					X	X		X		
Elaborate recommendation on licences			X								X		X
Support network operators to decide on licence		X									X		X
Community outreach to promote recommended citation practices	X	X	X									X	X
Interact with publishers on strategies to increase and improve citation		X	X									X	X
Include data citation practices in the review process				X								X	X
Improve author guidelines				X								X	X
Improve technical editorial checks				X								X	X
Ensure that network citations are included in the reference section				X			X	X				X	
Ensure that networks are cited in Data Availability Statements or equivalent				X				X				X	
Develop tools for citation of a large number of networks				X	X			X				X	
Avoid limitations on number of citations of datasets				X				X				X	X
Make it possible to create citation report, with choice of type of publication, for a large set of DOIs				X	X	X							X
Cite network data as recommended by the FDSN (see explanation Section 5)							X					X	X
Avoid depositing copies of network data in generalist repositories				X			X			X		X	X
Train and inform scientists about correct citations		X	X				X					X	X

Figure 5 Recommended actions. The left part identifies the activity, the central part the target audience, and the right part the improvements that the action will support

governed FAIR vocabularies for seismology. Local implementations of vocabularies for passive seismic data are emerging (e.g. https://geofon.gfz.de/cv/seisdata/), but there is yet no agreed global standard.

A draft FAIR Implementation profile (FIP, Schultes et al., 2020) that summarises our understanding of the current status of FAIR for seismological waveform data as standardized by FDSN is included in Supplementary Material.

6.3 Consideration of the CARE principles for seismic waveform data

The scientific community needs to consider the CARE (Collective Benefit, Authority to Control, Responsibility and Ethics) Principles for Indigenous Data Governance (Carroll et al., 2020) which apply to the collection, management, curation and publication of research data collected on Indigenous lands. The CARE principles have synergies with the FAIR principles and, over and above FAIR, CARE can provide criteria for cultural metadata, provenance, Indigenous governance, Indigenous ethics, transparency, integrity, and equity (Carroll et al., 2021). The diverse range of needs and requirements of Indigenous communities around the world makes it difficult to define a globally consistent set of practices that the science community can follow to uphold Indigenous rights.

When CARE is considered in the seismology waveform data life cycle, it is often and mostly addressed when placing the instruments in the field. Where planned network sites are on Indigenous lands, local traditional owners must be consulted well before starting new field acquisition. For example, the Transportable Array of the USArray project (http://www.usarray.org/public) handled these issues through a dedicated Permits team and the NSF (National Science Foundation) supported dedicated workshops (see, for example, Semken et al., 2007). The local authority structure and the level of formalisation of agreements for installation are varied and focus foremost on respect for the land itself as well as the sovereignty and agency of the communities that govern it. Agreements may be dependent on experience from prior engagement with science projects or broader land users. Well prepared handouts with place-based information, briefings to councils and accessible web information about the project are very helpful aides to a dialog about sharing land use.

Data Sovereignty and Governance may or may not be a preoccupation for the local authorities, and is often overlooked in seismology. The CARE framework would help to clarify the data management policy. When new data collected on Indigenous lands are placed in a repository, the local communities may need to be consulted about the management of, and access to, such data; and repositories may need to adjust their practices to comply with CARE. O'Brien et al. (2024) have provided guidelines on actions that Earth Science Data repositories can take to better adhere to the CARE Principles. Where Indigenous communities are asserting greater control over the use and application of data collected on their lands, there are increasing attempts to develop better data management practices, including when data are open. For example, seismic experiments in New Zealand are increasingly conducted following CARE principles, and recent temporary deployments have been fully designed and co-developed with the local indigenous community (Iwi Ngati Tūwharetoa and Te Arawa) following protocols and sharing decisions, knowledge and results with landowners, community leaders, schools, and young people (Mestel, 2023). Despite data from such experiments being distributed through FDSN and that the local indigenous authorities are acknowledged in the seismic metadata, Mestel (2023) recognizes that this is not sufficient to adhere to CARE principles. Similar co-development practices are generally hard to apply to nation-wide, continuous seismic network or dataset. In New Zealand, government agencies and research institutes are starting a long and complex journey to develop data governance best practices to uphold Māori Data Sovereignty rights (Kukutai et al., 2023; Kukutai, 2024). Some groups are promoting the use of Local Contexts Notices and Labels (https://localcontexts.org) as part of the metadata. With that, Indigenous communities can express their own definitions on access constraints. Through the use of Local Contexts Notices and Labels, a researcher can apply a Notice to indicate that there are Indigenous interests in a dataset and then the Community can provide Labels that express their perspectives on access and benefit sharing (e.g., Liggins et al., 2021).This approach is not endorsed by all Indigenous communities, highlighting again the need of co-developing the best approach with local authorities.

The scientific community should consider CARE principles for existing historical seismic data. If those were collected on indigenous lands, the relevant traditional owners should be contacted. Where historical data was collected with Indigenous permission, these protocols should be linked to the data set metadata (Indigenous Data Lab, 2024).

7 Conclusions and recommendations for future actions

The FDSN implementation of DataCite DOIs has overall been very successful and is supported by FDSN, data centers, and network operators. A positive effect of the present work was identifying and fixing a number of errors and inconsistencies that had accumulated over the past decades. Maintaining high quality and coherence between seismological metadata, DataCite metadata, and all aspects of FAIR needs continuous effort, and we do not at this stage believe that high quality and coherency is possible with a finer granularity level of DOIs.

Given the significant number of seismic network DataCite DOIs and the growing experience in the community with DataCite metadata, licences, and FAIR issues overall, we believe that it is now a good moment to reflect on the way forward and take appropriate action, and we hope that our analysis can contribute to doing that. We have assembled recommendations for short to intermediate-term actions in Figure 5.

Further FDSN efforts will also be dedicated to improving FAIRness, in an evolving landscape on both data types and data management in general. For example, FDSN is currently pursuing activities related to the integration of large volume data sets (Distributed Acoustic Sensing, DAS, in particular), and developing a community standard solution for Authentication and Authorization Infrastructure. Discussions are also being initiated concerning keywords and controlled vocabulary for inclusion at different levels of the FDSN channel stream identifier or and in the DataCite DOI metadata. Future actions should also include ethical concerns, such as the CARE principles (Carroll et al., 2020), which are relevant not only for countries with indigenous populations, but also for deployments in areas with indigenous populations.

FDSN needs to tackle challenges on citations of many networks, for example, for global studies, where the number of citations may go beyond the publisher's accepted number of citations per article. We argue that there should be no limitations on the number of data citations in any scientific journal. Additionally, there are ongoing initiatives to develop tools for two (or multiple) layer referencing, such as the RDA (Research Data Alliance) Complex Citation group (see recommendation by Agarwal et al., 2025). One key requirement to any solution is that they must allow finding and reporting on all referenced datasets, independent of where the dataset reference actually is. Other needs might arise in the future, for example citation of specific equipment and instruments, and other hierarchical discoverability and access needs. In all these evolutions, the network operators, data centers and FDSN as a whole will need to carefully weigh the benefits for each stakeholder to prioritize the actions that will have the highest impact for scientific users of the data, and for the data producers in terms of visibility and ease of sharing the data. We hope and believe that the present manuscript can contribute to FDSN decisions and services over the years to come.

To address the remaining challenges, the seismological community benefits from an established community governance structure with global reach, FDSN, and other organisations which are commissions of or hold commission status in the International Association of Seismology and Physics of the Earth's Interior (IASPEI, http://iaspei.org). IASPEI is an association of the International Union of Geodesy and Geophysics (IUGG, https://iugg.org/), which is a member organization of the International Science Council (ISC, https://council.science/). This hierarchical governance structure enables discussion and definition of community standards and best practices, and the curation and governance of their future development, not only within seismology itself, but also across other geophysical associations and ultimately across all scientific disciplines.

Acknowledgements

The authors are thankful to the organizations and personnel that operate seismic networks across the world and who share their data openly for science, and their funders for backing them. We also thank the scientific users of the data who continuously provide incentives and feedback for improving FDSN and data center services. We thank the FDSN community for developing and sharing open standards and technologies that enable seismology to be an open science. Special thanks go to those scientists, reviewers, and editors who follow the FDSN recommendations for network data citation.

This work was partly supported by Geo-INQUIRE, funded by the European Union under project number 101058518 within the HORIZON-INFRA-2021-SERV-01 call (https://doi.org/10.3030/101058518). EIDA services are integrated into and supported by the European Plate Observing System (EPOS). EarthScope Consortium is the current SAGE Seismic Major Facility, funded by the National Science Foundation (USA) CSA Award No. 2310069. GNS' GeoNet programme is sponsored by the New Zealand Government through its agencies: Natural Hazards Commission Toka Tū Ake (NHC), GNS Science Te Pū Ao, Toitū Te Whenua Land Information New Zealand (LINZ), the National Emergency Management Agency (NEMA) and the Ministry of Business, Innovation and Employment (MBIE). AuScope and NCI are supported by the Australian National Collaborative Research Infrastructure Strategy (NCRIS). Helle Pedersen (ISTerre and OSUG) and Jonathan Schaeffer (OSUG) are part of Labex OSUG@2020 (ANR10 LABX56). Frank Vernon and Bob Busby provided experience on seismic installations on indigenous lands in USA.

We thank Barbara Romanowicz and Tim Ahern for constructive and insightful reviews, and take the opportunity to thank them for their immense contributions to FDSN and to the seismological community.

Data and code availability

We used network information hosted at FDSN (https://www.fdsn.org/), and from the data centers indicated in the FDSN datacenter registry available through the data center api (https://www.fdsn.org/ws/datacenters/1/query). The following FDSN datacenters operate the fdsnws-dataselect service and were accessed for the analysis: The Australian Passive Seismic Server at ANU (http://auspass.edu.au), BGR Data Centre (https://www.bgr.bund.de), ETH Data Centre (https://www.ethz.ch), GEOFON Program (https://geofon.gfz.de), New Zealand National Seismograph Network (https://www.geonet.org.nz), ICGC Data Centre (https://www.icgc.cat/en/terratremols), Institute of Earth Sciences, Academia Sinica, Taiwan (http://batsws.earth.sinica.edu.tw/fdsnws), Instituto Geográfico Nacional (http://www.ign.es), INGV Data Centre (https://www.eida.ingv.it), IPGP Data Center (http://datacenter.ipgp.fr), NSF SAGE Data Services operated by EarthScope Consortium (https://ds.iris.edu), KOERI Data Centre (https://www.koeri.boun.edu.tr), Ludwig Maximilians University Munich (https:

//www.geophysik.uni-muenchen.de), Northern California Earthquake Data Center (https://ncedc.org), NIEP Data Centre (https://www.infp.ro), NOA Data Center (https://www.noa.gr), Canadian National Data Centre (CNDC) (https://earthquakescanada.nrcan.gc.ca/stndon/CNDC/index-en.php), ORFEUS Data Centre (https://www.orfeus-eu.org/), Raspberry Shake Seismic Network (http://raspberryshake.org), RESIF (now Epos-France) Data Centre (https://seismology.resif.fr), Southern California Earthquake Data Center (https://scedc.caltech.edu), UIB-NORSAR EIDA node (http://eida.geo.uib.no), Centro de Sismologia da Universidade de São Paulo (http://www.moho.iag.usp.br). The software developed in this study is openly available (Petrakopoulos and Schaeffer, 2024).

Competing interests

The authors have no competing interests

References

Agarwal, D., Ayliffe, J., Buck, J., Damerow, J., Parton, G., Stall, S., Stockhause, M., and Wyborn, L. Complex Citation Working Group Recommendation Version 3. *Zenodo*, 2025. doi: 10.5281/zenodo.14106602.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., , and Wassermann, J. ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, 81(3):530–533, 2010. doi: 10.1785/gssrl.81.3.530.

Carroll, S., Herczog, E., Hudson, M., Russell, K., and Stall, S. Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, 8:108, 2021. doi: 10.1038/s41597-021-00892-0.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., and Parsons, M. e. a. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(43):1–12, 2020. doi: 10.5334/dsj-2020-043.

Dziewonski, A. M. The FDSN: history and objectives. *Annals of Geophysics*, 37(5), 1994. doi: 10.4401/ag-4191.

Evans, P. L., Strollo, A., Clark, A., Ahern, T., Newman, R., Clinton, J. F., Pedersen, H., and Pequegnat, C. Why seismic networks need digital object identifiers. *Eos*, 96, 2015. doi: 10.1029/2015EO036971.

Heinloo, A. fdsnws_scripts: a collection of next generation distributed data request tools based on FDSN web services and the EIDA routing service. https://doi.org/10.5880/GFZ.2.4.2024.002, 2024. doi: 10.5880/GFZ.2.4.2024.002.

Indigenous Data Lab. Indigenous Metadata Bundle Communique. https://doi.org/10.6084/m9.figshare.24353743.v1, 2024. doi: 10.6084/m9.figshare.24353743.v1.

International Federation of Digital Seismograph Networks. FDSN Web Service Specifications Version 1.0. https://fdsn.org/webservices/FDSN-WS-Specifications-1.0.pdf, 2013. Accessed: november 2024.

International Federation of Digital Seismograph Networks. FDSN recommendations for seismic network DOIs and related FDSN services. International Federation of Digital Seismograph Networks (FDSN). https://doi.org/10.7914/D11596, 2023. doi: 10.7914/D11596.

Kukutai, T. How Indigenous communities in New Zealand are protecting their data. *Science*, 384, 2024. doi: 10.1126/science.ado9298.

Kukutai, T., Campbell-Kamariera, K., A., M., Mikaere, K., Moses, C., Whitehead, J., and D., C. Māori data governance model., 2023. https://www.kahuiraraunga.io/_files/ugd/b8e45c_803c03ffe532414183afcd8b9ced10dc.pdf.

Liggins, L., Hudson, M., and Anderson, J. Creating space for Indigenous perspectives on access and benefit-sharing: Encouraging researcher use of the Local Contexts Notices. *Molecular Ecology*, 30:2477–2482, 2021. doi: 10.1111/mec.15918.

Marignier, A., Eakin, C., Hejrani, B., Agrawal, S., and Hassan, R. Sediment thickness across Australia from passive seismic methods. *Geophysical Journal International*, 237(2):849–861, 2024. doi: 10.1093/gji/ggae070.

Mestel, E. *Seismicity, sub-surface structure and partnership with Tangata Whenua of Taupō Volcano*. PhD thesis, Te Herenga Waka-Victoria University of Wellington, 2023.

O'Brien, M., Duerr, R., Taitingfong, R., Martinez, A., Vera, L., Jennings, L., Downs, R., Antognoli, E., ten Brink, T., Halmai, N., and David-Chavez, D. Earth Science Data Repositories: Implementing the CARE Principles. *Data Science Journal*, 23(37):1–29, 2024. doi: 10.5334/dsj-2024-037.

Pedersen, H. A., Mattern, F., Poli, P., , and Stehly, L. Imaging with seismic noise: improving extraction of body wave phases from the deep Earth through selective stacking based on H/V ratios. *Geophysical Journal International*, 232(2):1455–1467, 2023. doi: 10.1093/gji/ggac388.

Petrakopoulos, V. and Schaeffer, J. find-net-board, 2024. https://hal.science/hal-04820268v1.

Quinteros, J. Routing Service: A data centre federation for the seismological community. GFZ German Research Centre for Geosciences, 2017. doi: 10.5880/gfz.2.4.2017.001.

Romanowicz, B. and Dziewonski, A. Toward a federation of broadband seismic networks. *EOS*, 81(67):541–542, 1986. doi: https://doi.org/10.1029/EO067i025p00541.

Schultes, E., Magagna, B., Hettne, K. M., Pergl, R., Suchánek, M., and Kuhn, T. Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In Grossmann, G. and Ram, S., editors, *Advances in Conceptual Modeling*, pages 138–147, Cham, 2020. Springer International Publishing. doi: 10.1007/978-3-030-65847-2_13.

Semken, S., Fouch, M., Garnero, E., Zah, P., and Lippert, D. Meshing American Indian concerns with goals of EarthScope's USArray. *EOS*, 88(31):309–310, 2007. doi: https://doi.org/10.1029/2007EO310001.

Stall, S., Bilder, G., and Cannon, M. e. a. Journal Production Guidance for Software and Data Citations. *Scientific Data*, 10:656, 2023. doi: 10.1038/s41597-023-02491-7.

Wilkinson, M., Dumontier, M., and Aalbersberg, I. e. a. The FAIR Guiding Principles for scientific data management and stewardship. *Science Data*, 3:160018, 2016. doi: 10.1038/sdata.2016.18.