



EGU25-21605

Status, issues and challenges with FAIRness of seismological waveform data and beyond

Florian Haslinger, Lesley Wyborn, Rob Casey, Helle Pedersen, Elisabetta D'Anastasio, Javier Quinteros, Jonathan Hanson, Jerry Carter

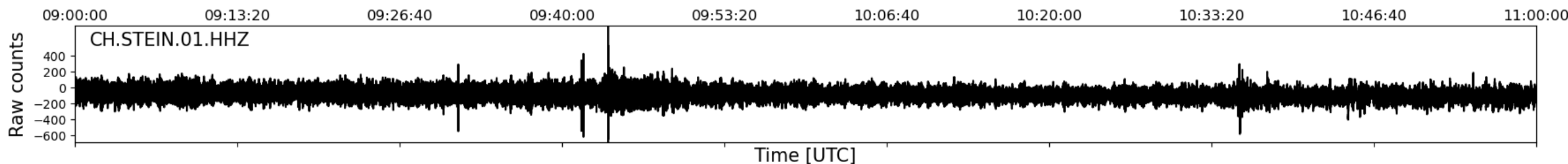


supported by EU
Horizon Europe
project 101058518



Standardisation in (waveform) seismology

2025-04-22 09:00:00.0 - 2025-04-22 11:00:00.0



SEED Reference Manual

Standard for the Exchange of Earthquake Data

- digital seismology starting in 1970s (before that recording on paper, film, ...)
- SEED as a standard adopted by seismological community (IASPEI) in 1997 defining the format (& content) for data and metadata
- further developments under IASPEI / FDSN over time → mSEED & stationXML

SEED *unique identifier*:

NETwork

CH

STAtion

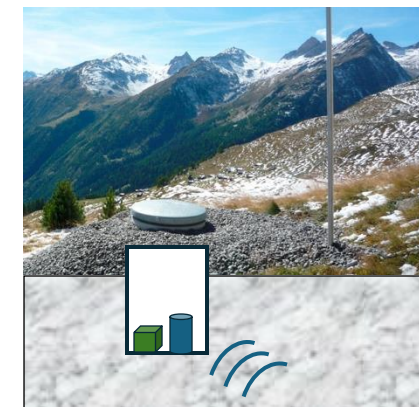
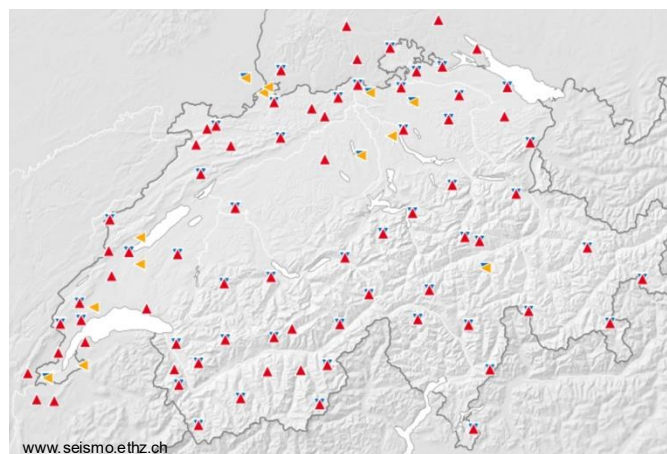
STEIN

LOCation (optional)

01

CHAnnel

HHZ



Status, issues and challenges with FAIRness of seismological waveform data and beyond

Standardisation in (waveform) seismology

Governance & Representation



CoSOI – Commission on Seismological Observation and Interpretation



www.fdsn.org

Current standards & their coverage (data, metadata, services)

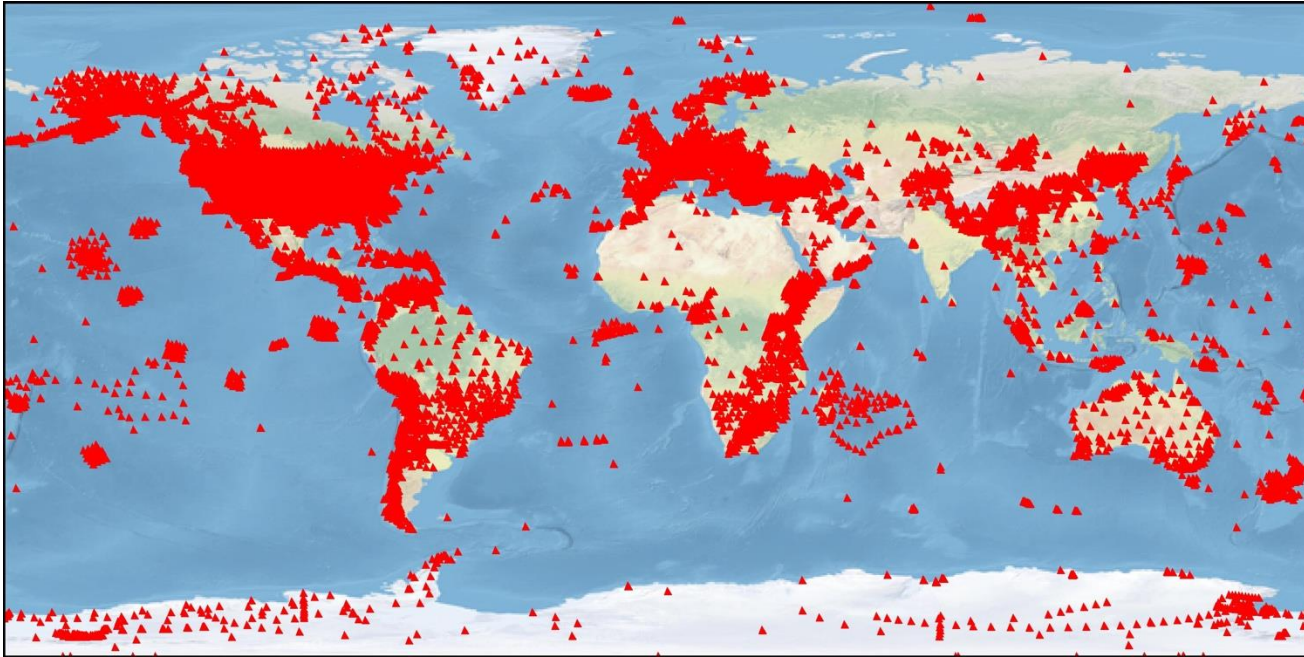
IANA registered media types	(mini)SEED mseed3	waveform data	adopted / IASPEI (1997) de-facto / FDSN (2023)
	stationXML	station information (metadata)	de-facto / FDSN (2012)
	FDSN standard web services	data- / metadata access	de-facto / FDSN (2013)
	QuakeML	event parameters (xml)	adopted / IASPEI (2015)
	IASPEI Seismic Format ISF	event parameters	adopted / IASPEI (2015)

FDSN recommendation (how) to use DataCite DOI on NETWORK level (2014)

community decision to keep identifier at that level
– main target: increase attribution of networks (operating institutions / PIs)



Standardisation in (waveform) seismology



from Pedersen et al., Seismica, 2025 DOI 10.26443/seismica.v4i1.1537

map of seismic stations worldwide (10'000s)

- that are part of networks registered with the FDSN
- for which station metadata can be retrieved by the standard FDSN webservice

includes

- permanent networks
(mostly for national earthquake monitoring purposes)
- temporary networks
(academic experiments for specific studies & targets)



centralised registry & data center, operated by (former IRIS)

in Europe: European Integrated Data Archive EIDA, by
(federated data centers / repositories)



How FAIR are seismological waveform data?

initial thinking: *sure, we have it (almost) all ...*

on closer inspection (testing network DOI on www.f-uji.net):

ouch – 30% FAIR only??

→ aha: *f-uji doesn't know our standards !*

dialog & improvements *(are still tricky...)*

→ **seismic networks score between 30% and 80%**

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	DOI, ORCID, ROR
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	DOI
F2	Which metadata schemas do you use for findability?	Metadata schema	Datacite, StationXML(?)
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	M->D: StationXML, Signposting in landing page
F4	In which search engines are your metadata records indexed?	Search engines	Handle(DOI) / FDSN
F4	In which search engines are your datasets indexed?	Search engines	Handle(DOI) / FDSN
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	FDSN StationWS, Datacite common API
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	FDSN Dataselect
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	No AuthN/AuthZ
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	HTTP Digest basic (if restricted)
A2	Which metadata longevity plan do you use?	Metadata longevity	DataCite DOI Policy
I1	Which knowledge representation languages (allowing machine interoperation) do you use for	Knowledge representation language	SEED?
I1	Which knowledge representation languages (allowing machine interoperation) do you use for	Knowledge representation language	SEED?
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	The ones specified in StationXML, GCMD, NAS
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	SEED standard
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	DataCite, StationXML
I3	Which models, schema(s) do you use for your datasets?	Data schema	SEED
R1.1	Which usage license do you use for your metadata records?	Data usage license	No license. Completely open, but not declared
R1.1	Which usage license do you use for your datasets?	Data usage license	Usually CC-BY. Europe: EPOS recommendation
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	StationXML, DataCite
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	StationXML, DataCite

recent f-uji results for 6 European datacenters
(average over all hosted networks / DOIs)



dashboard courtesy of E. Martinez & colleagues @EMSO

draft **FAIR Implementation Profile** FIP
for seismological waveform data
: input for community discussion & decision

How FAIR are seismological waveform data?

Challenges and open issues

licenses (*particular for waveform data*)

- no established standard
FDSN working on a community recommendation
(likely CC-BY or CC0)
- depending on national legislations and policies
(can that type of data be licensed?)

(FAIR) vocabularies

- no comprehensive vocab for seismology (yet)
various (partial) attempts
needs governance & resources

new data types

- large (1000s of sensors) deployments or
data from optical fibers (DAS) don't fit
established formats (data & metadata)



Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

...

Accessible:

...

Interoperable:

...

I2. (meta)data use vocabularies that follow FAIR principles.

...

Re-usable:

...

R1.1. (meta)data are released with a clear and accessible data usage license.

R1.2. (meta)data are associated with their provenance.

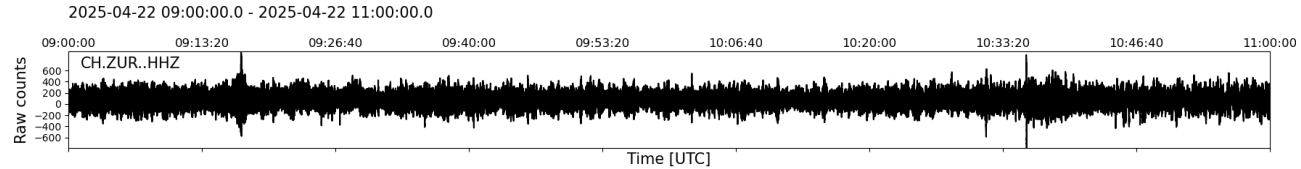
R1.3. (meta)data meet domain-relevant community standards.

other issues

- DOIs for networks may not be good enough(?)
- provenance information may be lacking (in metadata model)
- who defines what is a '**domain relevant community standard**'
(and how...)

FAIR seismology (beyond waveforms)

waveforms (ground motion recordings)
are the basic (raw) data of seismology



from those, seismologists (and computer programs) generate:

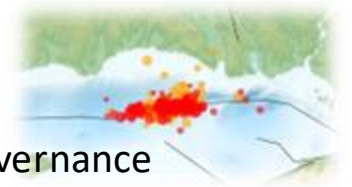
- earthquake information
- subsurface models
- seismic hazard assessments
- ...



earthquake information (products)

international / European organisations

- provide (some) coordination and governance
- promote best practices & standardisation



✓ **community standard formats** for (most relevant) products

? **identifiers** are rather 'local' than global
resolution / granularity ?

? **vocabularies** not existing and **licenses** a challenge

seismic hazard & risk assessment

international / European organisations

- provide (some) coordination and governance
- promote best practices & standardisation



✓ **community standard formats** (software driven)

? **identifiers** on 'whole model' level
resolution / granularity ?

✓ **vocabularies** & **licenses** partially covered

subsurface models (and other higher-level scientific work)

FAIR through academic publication (publisher policy)

- maybe?

... it depends ...

finally – conclusions, suggestions and questions for the way forward

Community (owned) governance is key!

40 yrs of community-internal standardisation provided a good start for FAIR
in particular **F** and **A**,
but still challenging for **I** (vocab) and **R** (licenses)

for seismology, utilising established (global) organisations under the
ISC / IUGG umbrella to govern standards is working reasonably well
how to utilise IUGG / ISC for cross-domain coordination?

granularity of (data) identifiers (network, device, file, dataset, ...)

NET level DOI for seismology may not be sufficient (*new: PIDs for instruments*)
different communities follow different approaches (for different reasons)

standards and (standardized) identifiers for higher level (derived) products?

Is there a difference (*should there be*) between general repositories for FAIR digital objects (research output)
and community-specific data infrastructures?



(among other great things...)
coordination of FAIR implementation and
assessment across ENV Research
Infrastructures in Europe:





EGU25-21605

Status, issues and challenges with FAIRness of seismological waveform data and beyond

Thank You

Florian Haslinger, Lesley Wyborn, Rob Casey, Helle Pedersen, Elisabetta D'Anastasio,
Javier Quinteros, Jonathan Hanson, Jerry Carter



*supported by EU
Horizon Europe
project 101058518*

