Working Group 2
Data exchange

Session

Meeting

Agenda

- Review and Approval of 2021 Minutes from India
- Status of 2021 Action Items and summary of activities

- Metadata for Legacy datasets (T. Ahern)
- miniseed3: vision for the future and next steps after approval (C. Trabant)
- DAS metadata: the DAS RCN group and what can we expect at the FDSN level (J. Carter)

Agenda

- Proposals to be prepared:
  - Controlled Vocabularies for seismology (F. Haslinger/A. Strollo)
  - Best Practices and automated checks for metadata in StationXML
  - Reproducibility: what is the next step into this direction? (J. Quinteros)

- Any other business

Review and Approval of 2021 Minutes from India

https://fdsn.org/media/wg/II/2021/WGII_Meeting_Minutes_-_India_2021.pdf

Action Items from 2021

- **[Action Item – Metadata for legacy data]:** Send proposal to the WG with a Call to form an Proposal Review Team. (WG Chair)
  Proposal submitted, evaluated and accepted (June 2022).

- **[Action Item – miniseed3]:** Send the proposal to the WG with a Call to form an Evaluation Review Team. (WG Chair)
  Proposal submitted, evaluated and accepted (April 2023).

Action Items from 2021

- **[Action Item – GNSS integration]:** Identify the groups within the FDSN that are already working on this topic and together prepare a white paper about the GNSS data integration in Seismology. (GFZ and others)

- **[Action Item – GNSS integration]:** Develop guidelines to address the immediate needs (short term) and start discussion for long term options. (GFZ and others)

Action Items from 2021

Groups contacted and discussions are going on.
Special session following this Action Item during this Saturday 15th

JS05 Real-Time GNSS Data and Products Usage: Interoperability and Management Challenges (IASPEI, IAG, IAVCEI, IAPSO). Convener(s): Angelo Strollo (Germany, IASPEI). Co-Convener(s): Antonio Avallone (Italy, IAG), Yuhe Tony Song (USA, IAPSO), Clinton John (Switzerland, IASPEI), Giuseppe Puglisi (Italy, IAVCEI)

Mapping created in the context of this initiative:
https://www.seiscomp.de/doc/apps/seedlink.html?highlight=gdrt#seedlink-sources-gdrt-label

Action Items from 2021

- **[Action Item - StationXML]:** The Review Team work should start in the next few days. Issues will be evaluated and a Recommendation is expected as an output. Corrections to the standard, or documentation, but also new features for next release.

  Group formed, evaluation done, recommendation accepted, new schema (v1.2) released (April 2022).
- **[Action Item - QuakeML]:** ETH will open the discussion about QuakeML on the WG, including new functionality planned to be included in v2.0. (John Clinton)

  ETH intends to develop in the next 1-2 years, but resources are scarce and it's not a top priority. Other teams would be welcome to proceed, with our collaboration, if they would like.

Action Items from 2021

- **[Action Item - Event types]:** Dmitry Storchak has agreed to lead an effort that will span both FDSN WG2 and COSOI regarding Event Types. He will inform the community about the progress on this discussion and joint effort. (Dmitry Storchak)

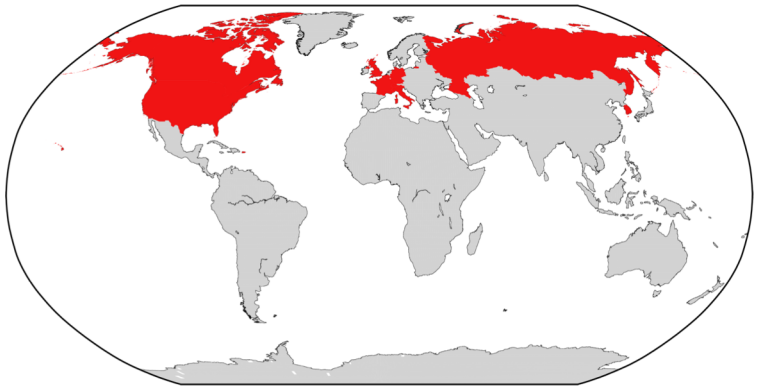
International Federation of Digital Seismograph Networks

IASPEI CoSOI-FDSN inter-commission Working Group on the Event Type Nomenclature

Introduction

Several relevant efforts were made over the years. The last attempt was made jointly by the ISC, NEIC, and EMSC during 2010-2012. It culminated with the nomenclature proposed for use in ISF and QuakeML formats (*Storchak et al., 2012*).

After 11 years, it was decided to update the effort by forming a working group of all interested parties. The WG has been formed based on the invitation circulated through IRIS and ISC bulk mail as well as by direct approach to potentially interested parties.



Researchers from 12 countries have so far signed up to work as part of this working group

We plan to:

- expand those groups of events that have not yet received a detailed attention, such as **volcanic**, **infrasound**, **landslides**, **ice-quakes,** as well as events on **other planets**;
- extend the existing nomenclature for **induced** events, **explosions** and **collapses**;
- consider relevant needs of various **networks** (for example, *Ackerley et al., 2022. Procedures for event type discrimination at Canadian CHIS*);
- keep a hierarchy allowing the most general down to progressively more detailed type naming.

We aim to:

- publish a scientific article summarizing this work;
- build relevant webpages;
- compile a bibliography of relevant examples;
- encourage an update of the FDSN webservices and other event search mechanisms.

A Commission of IASPEI

Presentations

- Metadata for legacy datasets (Tim Ahern)
- Miniseed3: vision for the future and steps after aproval (Chad Trabant)
- DAS metadata: the DAS RCN group. What can we expect at the FDSN level? (Jerry Carter)

New activities (coming proposals)

- Controlled Vocabularies for seismology (F. Haslinger/A. Strollo)
- Best Practices and automated checks for metadata in StationXML
- Reproducibility: what is the next step into this direction? (J. Quinteros)

Best Practices and checks for StationXML

- StationXML documentation started in 2020 as an initiative supported by IRIS and ORFEUS
  https://docs.fdsn.org/projects/stationxml/en/latest/index.html

- First version: 2020-2021. (ISTI)

- Revision on April 2022. (P. Crotwell, R. Casey, P. Kolinsky, J. Schaeffer)

Best Practices and checks for StationXML

- We have detected guidelines written by many members of the community in a joint effort.

- We would like to do a fast review of those and include them in the official documentation. Some topics are related to:

    - Structural Monitoring
    - OBS data
    - Tools for automatic validation (StationXML validator)

XML to JSON conversion

- It is sometimes cumbersome to work with large XML files.
- Many members have expressed the need to be able to represent XML files in JSON format (e.g. Minutes from 2021).
- We will start working on the JSON version of our 2 XML standards:

StationXML and **QuakeML**

- Ideally a 1-to-1 mapping to be able to convert in both ways.
- We will come back to you all to ask your opinion and collaboration.

Reproducibility – Define a dataset

- Attribution for the data provider has already been addressed (see FDSN Recommendations for DOIs).
- **What happens when someone publishes a paper and the editor asks for a link to the dataset used?**
- It is almost impossible to provide, due to storage space limitations (e.g. the case of ML techniques).
- Even for small datasets could be difficult, as data requested is curated until considered final.
- Therefore, it is impossible for the reader to reproduce the results.

DataCite and dynamic datasets

- For datasets that are continuously and rapidly updated, there are special challenges both in citation and preservation.

- DataCite suggests 4 options to solve this. You can cite:
  - 1: a specific slice or subset (we cannot express the curation)
  - 2: a specific snap-shot (large storage space needed)
  - 3: the whole dynamic dataset with datetime of access (OMG!)
  - 4: **a query, time-stamped for re-execution against a database**

DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. https://doi.org/10.14454/3w3z-sa82

Reproducibility – A first step in this direction

1. Download data based on your own filters (streams, time windows)
2. Discard what you don't want (due to low quality, or other reasons)
3. Process your data and get results
4. Write the manuscript
5. **Scan the dataset in its final state and generate a precise list of streams and time windows (normal POST format for FDSN web services?)**
6. **Assign a permanent identifier to it (even a DOI from Zenodo would be enough)**
7. Submit that to the Journal

Reproducibility – A first step in this direction

- We can use the FDSN Data Centre Registry to determine the authoritative data centre.
- It is a good opportunity to foster the registration and declaration of datasets.
- Simple client can be provided by FDSN or even included in well known clients from the community (e.g. obspy, fdsnwsscripts, major IRIS tools).
- Easy to check for a reader that all data is available.

Reproducibility – A first step in this direction

- Pros:
  - Exact streams and time windows
  - It can be regenerated by the reader
  - Journals requirements are fulfilled
  - No extra storage needed for data centres

- Cons:
  - Data Centres do not always support versioning
  - Minor differences due to changes in (meta)data (e.g. gap filling)

Remarks

- Remember to include the **DOI in StationXML**. Since schema version 1.1 it is posible to include an identifier (e.g. DOI) at the network level. Also in other levels, but the network is the most important one.

- We need more participation of young researchers. Things change too fast and we need fresh ideas. How to reach them and invite them to participate?