

2015 FDSN Meetings
FDSN WGIII
2015.6.29
Draft Minutes
Prague, Czech Republic

Participants:

Tim Ahern – IRIS tim@iris.washington.edu
Chad Trabant - IRIS chad@iris.washington.edu
Kent Anderson - IRIS kent@iris.edu
Bruce Beaudoin - IRIS - bruce@passcal.nmt.edu
Paul Earle - USGS - pearle@usgs.gov
Catherine Pequegnat - EIDA - Catherine.Pequegnat@ujf-grenoble.fr
Michelle Grobbelaar - CGS, SouthAfrica - michelle@geoscience.org.za
Mark Chadwick - GNS Science, NZ - m.chadwick@gns.cri.nz
Angelo Strollo - GFZ strollo@gfz-potsdam.de
Florian Haslinger - ETH/SED - haslinger@sed.ethz.ch
John Clinton - ETH/SED - jclinton@sed.ethz.ch
Fabian Euchner - ETH/SED - fabian@sed.ethz.ch
Frederik Tilmann - GFZ - tilmann@gfz-potsdam.de
Nikolaus Horn - ZAMG/Vienna - n.horn@zamg.ac.at
Luca Trani - KNMI - trani@knmi.nl
Reinoud Sleeman - knmi/ORFEUS - sleeman@knmi.nl
Peter Voss - GEUS - pv@geus.dk
Seiji Tsuboi - JAMSTEC - tsuboi@jamstec.go.jp
Katrin Hafner - IRIS/GSN - hafner@iris.edu
Pete Davis - UCSD/IDA - pdavis@ucsd.edu
Eleonore Stutzmann - GEOSCOPE - stutz@ipgp.fr
Martin Vallee - GEOSCOPE - vallee@ipgp.fr
Hanna Silvennoinen - SGO - hanna.silvennoinen@oulu.fi
Ludek Vecsey - IG-CAS - vecsey@ig.cas.cz
Winfried Hanka - GFZ/GEOFON - hanka@gfz-potsdam.de
Alexey Malovichko - GSRAS - amal@gsras.ru
Kenan Yanik - AFAD/Turkey - kenan.yanik@afad.gov.tr
Eren Tepeugur - AAD/Turkey - eren.tepeugur@afad.gov.tr

Start 12:00

- 2013 meeting minutes are approved unanimously.
- Review of agenda and question to see if any other topics should be included
- Agenda includes the following main topics and areas of discussion
 - Review of Existing FDSN Services and possible changes
 - FEDERATING THE FDSN DATA CENTERS
 - FDSN QUALITY ASSURANCE
 - PRODUCTS
 - Digital Object Identifiers (DOIs)
- Review of Charter for WGIII
Ahern: Charge already includes Quality Assurance activities

Sleeman: there is some overlap with other WG activities

Ahern: suggested wording changes should be circulated and/or brought up at the end of the meeting

Clinton: WGIII recent focus has been on standards for services, perhaps services focus should be moved up

Ahern: OK, likes the idea of moving services up. Charges will be reordered.

- Ahern: do any existing service specifications need updating?
- fdsnws-event changes?
 - Clinton: an event type should be included.
 - Process suggestion: ETHZ should make a proposal (implicit: to mailing list)
 - Do we need full SEED? Catherine Pequegnot felt it was a good idea. Ahern felt that the concept of full SEED needs to change and we should be moving to StationXML for metadata and miniSeed services such as returned by dataselect. There seemed to be general support for the idea of separating the metadata (stationXML) and the timeseries (miniSeed). IRIS intends to move in this direction.
 - Hasslinger: event type needs clarification, use on an existing set of types. Earle: COSOI recommendations are already used in QuakeML, might evolve in current CoSOI meeting.
 - Hasslinger: some are offering “fdsn” services with differences.
 - Ahern: felt that when that happens those DCs should be contacted and let them know they are not FDSN compliant.
 - Should extensions be allowed: yes, but collisions are possible, e.g. new “format” values. Extensions to standard services should be implemented at an operator's own risk, the FDSN may claim the namespace of changes used at any future date.
 - Ahern: changes should be quickly brought to the FDSN for consideration as part of the standard.

Action Item: ETHZ will submit a proposal to include event type in fdsnws-event specification.

- Trani: Data availability does not necessarily belong in StationXML/fdsnws-station service. Should be part of a separate service, perhaps format so that the concept can be expanded to other details such as quality metrics.
- Trabant: time series data availability is so low-level and commonly needed to be matched with station metadata that ‘matchtimeseries’ and simple notation of availability in StationXML/fdsnws-station is worthwhile.

Ahern: Recommendations for new services:

Timeseries (a processed data service):

Earle: this may slow down adoption due to being more complex. Ahern mentioned that the capability exists within IRIS and the code could be shared.

Should core functionality of timeseries be prioritized? Should we identify a few key new capabilities (downsampling, more formats supported as output, etc.)

Hasslinger: Data centers might all need some rate limiting capability to keep from being overloaded. Ahern: IRIS already has built infrastructure to control/throttle individual access to services to address this, other centers might need to do the same thing.

Tilmann/Vallee: could focus on data-volume reducing functions (decimation, only radial, etc.) might be a good focus.

GEOSCOPE: all-in-one-file still has advantages. GEOSCOPE could continue supporting this. IRIS will likely not do so as there are also many disadvantages to keeping the metadata with the timeseries data.

Most focus: format conversion and data volume reducing operations.

Stutzmann: need standard processing algorithms (minimal change). Ahern agreed that the code to do these things would be best if it is a common implementation across centers

Action Item: Ahern to send email to request added capabilities to a processed data service.

SAC P&Z and RESP

Ahern: any objection to proposing these for adoption

Clinton: could these simply be output options for fdsnws-station?

Ahern: Not completely inline with the service oriented architecture model where services perform specific functions in a modular service style. Trani: but they are all descriptions of a station, so it may fit in that scope.

Rotation

Strollo: can this be an output option for other time series services (e.g. dataselect, timeseries)?

Trabant: non-trivial number of special parameters needed for rotation.

Martin Vallee: one trace in a set could be bad, rotation hides this from the user.
email dialog needed

Metadata change service

Trabant: provided a description of metadata change tracking system. It is driven from stationXML files and a method of detecting significant differences

Sleeman: through multiple changes, does a user get different answers? Trabant: No, changes are accumulating, so the set is always added to.

Strollo: tracking is not linked with when the operator (and time series) actually changed, this is also important, possibly addressed in the metadata versioning.

Ahern: this could be federated across multiple federated centers if such a service is adopted as an FDSN standard.

Trabant: there are two cases: one user-focused, when they got metadata from a DC is important and not necessarily the same as when an operator changed parameters. Two: metadata/timeseries focused, changes when the actual data changed.

StationXML versioning is needed to track any sort of metadata focussed change tracking. An element and support is needed in StationXML, a WGII issue.

Action Item: need to propose metadata versioning *task force* that can make recommendations to WGII (DMC?). Trabant offered to lead the effort.

Federating Data Centers within ORFEUS (Angelo Strollo):

History from centralized archive

The EIDA Next Generation
Routing service
Demo using routing service and fdsnws
Output and input to the routing service
How to maintain
Proposed actions within the FDSN to federate DCs

Early federation was done with ArcLink based on a master table and routing table. Next generation is similar approach but based on FDSN web services.

1. Routing service - knows where endpoints are, directly accessible by users for use with Smart clients.
2. Mediator service - uses routing service to discover, then collects data and assembles for user.

Demo: loading waveforms from IRIS+EIDA in SC3 using the routing service

Initially the ArcLink routing table was used as input for the routing service. New input is for a DC to declare what they want to expose (data and services) with what level of priority and who are the other DC's belonging to the federation. The routing service can handle different level of federation (local, institutional, etc.).

Some important issues considered:

simple harvesting of fdsnws-station services will lead to ambiguities: questionable priorities in the routes, may assume fdsnws-dataselect and/or other services where fdsnws-station is running (may wrongly assume additional services), may need to interpret mismatching station locations.

Proposed approach: should allow the joining of a DC to a federation and exposing services instead of harvesting metadata and aggregating them to a central database. Trabant mentioned that performance is a factor and best served by a central database harvested from exposed services.

EIDA would welcome the formation of a small task force to discuss further the design and implementation of a FDSN federator .

Discussion:

Ahern: Is the foundation in SC3? Strollo: no dependencies on SC3 although will be implemented also in SC3.

Catherine: feedback from an EIDA node querying the IRIS federator and the EIDA routing service. <http://www.fdsn.org/wgIII/2015/2-WGIII-RESIF-DC-Feedback.pdf>

Catherine felt that data producers: should be allowed to designate where the authoritative data center is for their data. Ensure the authoritative data reach the user, obtain statistics, and be able to improve/correct data sets.

Examples: for IRIS Federator with some problems found, one seemed to be a real bug and several were just related to incomplete business rules that are still evolving.

Chad: IRIS Federation efforts

- started by harvesting all the metadata - quicker
 - bare minimum to be able to identify locations

- started with data centers that are contributing now and are listed on the FDSN WGIII site
- Two components
 - composite catalog of time series that is updated nightly
 - Service for queries
 - Is being integrated into other IRIS DMC user tools and community developed toolkits
- Rules:
 - used to remove duplicates and direct data request to primary data centers if possible
 - greater granularity than Net DC and
 - supports Network, Station, Channel, Location, time
- DEMO:
 - fedcatalog service
 - patterned after FDSN WS - extensive help
 - DEMO 1: Example URL builder (FR RUSF)
 - output (request format) shows data centers and request you would submit to receive data
 - BUG: returns duplicates at this point (both ORFEUS and RESIF) - should have only returned RESIF
 - should rank and only return the highest ranking data center
 - Angelo: how is priority calculated? Not really clear the rules and how the ranking is calculated.
 - Chad: ranking provided for each rule and only the highest overall ranked DC is presented
 - Tim: This is how it's being implemented at this time at IRIS, but FDSN may standardize the rule interpretation and can set priorities.
 - DEMO 2: outputs generate visible datacenters attribution
 - Ahern: This may be different from EIDA's approach
 - Strollo: the originating data center is actually there. Also on the client side
 - Ahern: It didn't seem clear to an end-user in the demo even if the information might be available.

Ahern

Developing an FDSN Federator

- Comparing the two proposed approaches (EIDA routing service and IRIS federator)
- Envisaged functionalities
- Considerations
 - reliability
 - politics
 - proximity
 - scalability
- Developing the Rules
- Secondary Data centers;

Action Item: Ahern: If you're interested in being on a FDSN Federated System task force- send Tim an email and it would be wonderful to get participation from other than US/Europe. The intent is to form a *task force* for this activity.

Ahern:

Overview of MUSTANG - automated data quality assurance system

Most internal activities are web services based, all external interfaces are web services. Some MUSTANG clients exist: MUSTANG data browser, LASSO and scripts that are effectively clients are being developed.

Automated text reports (operator/analyst focused):

- Created by script that harvests multiple MUSTANG metrics

- These scripts attempt to group MUSTANG metrics into problem types

- Quickly focus on problem stations rather than human review of many/all stations.

Network operator reports can be produced, based on network and virtual networks.

These are still labor intensive and IRIS hopes to make them more automated.

Challenges: at IRIS there are huge amounts of metrics that still need to be calculated.

The MUSTANG system for large parts of the IRIS holdings is totally operational including GSN, FDSN, and PASSCAL data held at the DMC.

EIDA waveform quality overview (WFCatalog), Luca Trani

<http://www.fdsn.org/wgIII/2015/3-WGIII-EIDA-QC.pdf>

WFCatalog:

- provides a well-defined API to query for seismic waveform metadata (including QC information)

- enables continuous waveform discovery based on metadata => no unnecessary downloads

- preliminary analysis and processing of seismic waveforms moved to data centers

=> big advantage for users who can quickly browse data and related features

- more than just QC...

Metrics are from a harmonization from NERA EC and discussed in EIDA and are present in MUSTANG.

EIDA provides a complete package to extract and manage QC and waveform metadata.

Specification principles:

- maximize compliance with FDSN standard services

- facilitate compliance/interoperability with existing systems, e.g. MUSTANG

- provide flexibility to address different use cases

- enable future extensions, e.g. number of metrics and diff granularities

- enable integration within broader EIDA NG arch.

scalable and flexible architecture based on a few key components. Luca feel it is "big data" ready.

Concept of data filtering based on quality parameters included. This can be effective way to reduce the data volume sent to end-users.

Tech: settled on MongoDB.

Requirements: performance, scalability, efficient data handling, flexibility and extensibility.

Action Item: Ahern to circulate an email to form a *task force* to discuss standardized quality assessment approach.

Hasslinger: metric definition depends on desired usage. CTBT, for example, creates metrics for masking/filtering data for downstream processing.

MUSTANG approach started first to improve data quality by reporting to network operators, but a known second use case is to support data filtering for users Research Ready Data Sets (RRDS). RRDS effort will begin early in 2016 and will be completed by September 2018.

Data Product Efforts

IRIS - Trabant

- High level data products
 - some based on reducing data sets for researchers or for EPO purposes
- Visualizer Tools
 - Earth model repository
 - 3D visualizations
 - Python based
 - works with general netCDF volumes
 - would like to release this to the community to have add-ons produced by the community that IRIS would monitor
- Data products usage
 - bot stats are biasing actual usage and so are routinely removed from statistics
 - peaks in product usage cluster around earthquakes (event based products)
 - broad usage
 - Nepal example
 - sub peaks in usage around weekends
 - product usage was distributed across several products (more than just GMV's)
 - Global CMT access much higher than every other product.

Strollo:

Overview of EIDA data products (primarily time series data and related description) and Strong Motion data base and related products. There were no higher-level products mentioned. http://www.fdsn.org/wgIII/2015/4-WGIII-ORFEUS-EIDA-Prodsvr_20150629.pdf

FDSN Digital Object Identifiers

Ahern gave a summary of the new presence of DOI minting capability on the FDSN web site. IRIS has revamped FDSN Network pages significantly.

The new system is no longer in Oracle, but is now in PostGress

FDSN web presence is totally isolated from DMC operations through a dedicated virtual machine

Requesting a permanent network code (<http://www.fdsn.org/networks/request/perm/>) is completely FDSN branded. Examples were shown.

The application now includes Network Citations details, including 3 DOI options:

1. FDSN should mint and managed DOI
2. This network has a DOI already
3. Do nothing right now

New FDSN Network pages (<http://fdsn.org/networks/>) show either FDSN-minted DOI or network operator supplied DOI.

Network station map comes directly from FDSN data center running federated services whenever possible. Publications can be included in network pages. It is straightforward for a network operator to update this information. The older method of providing citations for networks will be replaced by this new system.

Important steps for DOIs:

- Awareness is needed that the capability exists. So WGIII members are encouraged to let other network operators know.
- IRIS DMC will include DOIs in quarterly reports sent to users receiving data from the IRIS DMC
- IRIS DMC requests help contacting operators to get them to mint DOIs for their networks and register them with the FDSN by updating the information for Networks at <http://www.fdsn.org/networks/>, selecting the relevant network and clicking on the Update information link such as <http://www.fdsn.org/networks/request/XX> where XX is the actual network code.
- DOI support is needed in StationXML and this information has been forwarded to WGII.

Florian: perhaps Working Group Chair and Vice Chair should be discussed in working group?

If there are other nominations, they should be submitted to the Secretary (currently Michelle) or other members of the ExCom.