Orfeus

Observatories and Research Facilities for European Seismology

EIDA European Integrated Data Archive

Federating Data Centers within ORFEUS: The EIDA experience and future plans

The EIDA community

Presented by Angelo Strollo*

Prague

June 29. 2014

*Currently EIDA Management Board Chair and part of the interim ORFEUS Secretariat.

Outline

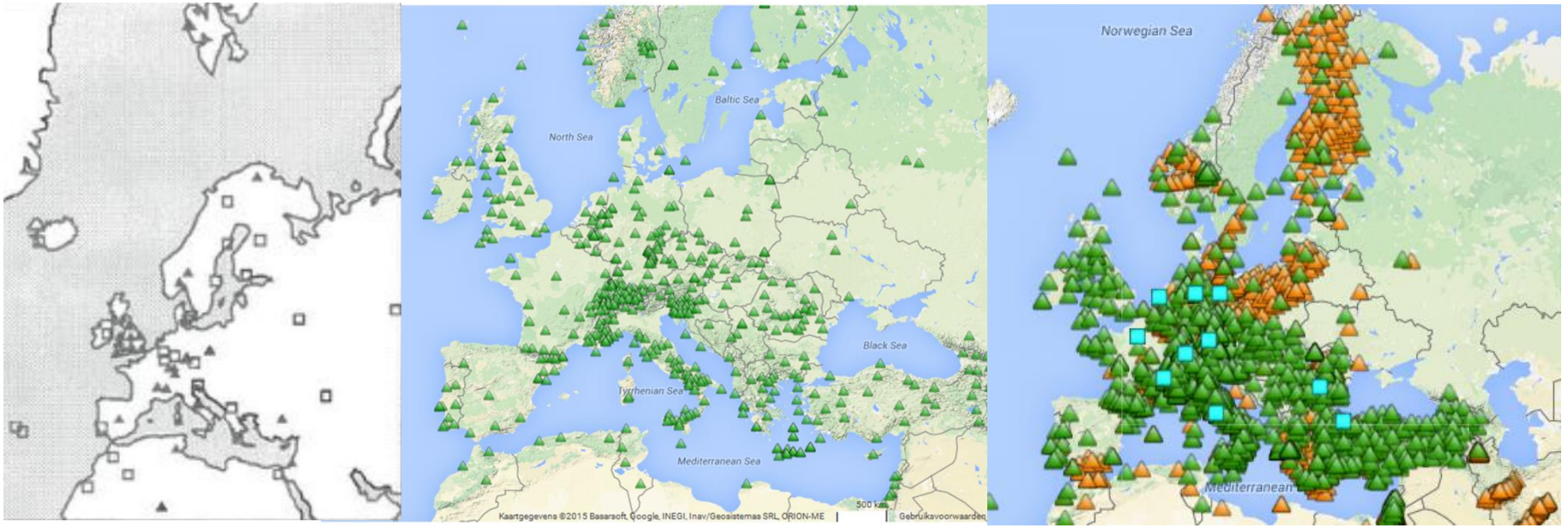
- ORFEUS evolution:
  centralized archive => EIDA => EIDA-NG
- The EIDA Next Generation
  - Routing service
- Quick demo using the routing service + fdsws
- Output and Input to the routing service
- How to maintain/feed federated routing services?
- Proposed actions within the FDSN based on the EIDA experience in federating DCs

ORFEUS

From the ORFEUS Data Center (single DC) to EIDA (distributed DC)

- 1986/87: ORFEUS plan launched/realized
- 2012/13: ORFEUS-VEBSN => ORFEUS-EIDA (EIDA = VEBSN + data holdings from 9 European DCs)
- 2015: ORFEUS-EIDA (10 nodes)

1987 - 30 **2013 - 613** **2015 ~2000 (5000+)**



Year - Number of open and operational stations

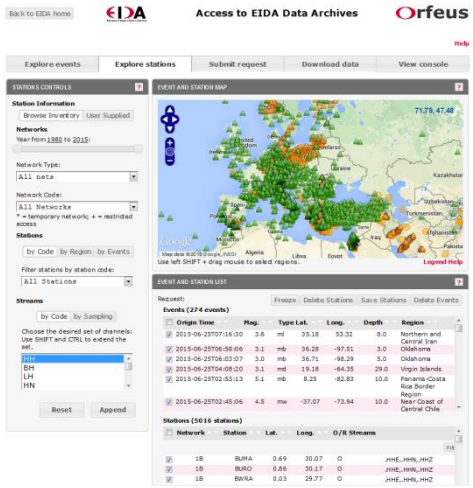
ORFEUS data infrastructure, services and products

EIDA => EIDA-Next Generation

More than just data and federated archive

- **Coordination of data holdings and software/strategic developments**
- **Provides quality control of data/metadata**
- **Helps define seismological center 'best practice' for ORFEUS community**

http://www.orfeus-eu.org/eida/eida.html



Users:
Geoscientists
Etc…

waveforms responses routing

Arclink routing table master table

> 5000 stations

INFP

New nodes

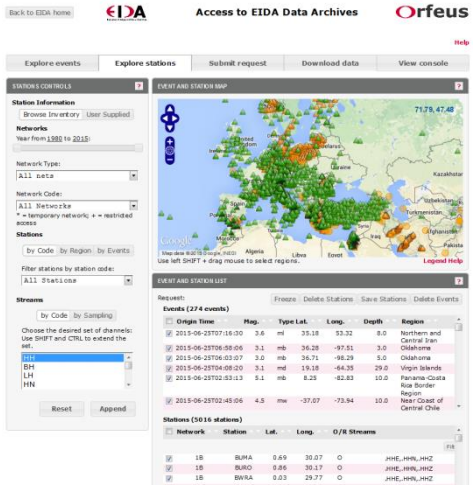
ORFEUS data infrastructure, services and products

EIDA => EIDA-Next Generation

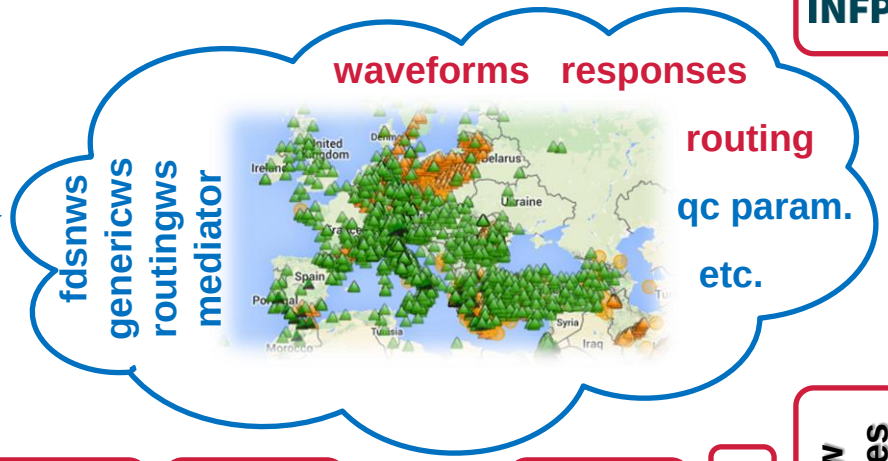
More than just data and federated archive

- **Coordination of data holdings and software/strategic developments**
- **Provides quality control of data/metadata**
- **Helps define seismological center 'best practice' for ORFEUS community**

http://www.orfeus-eu.org/eida/eida.html



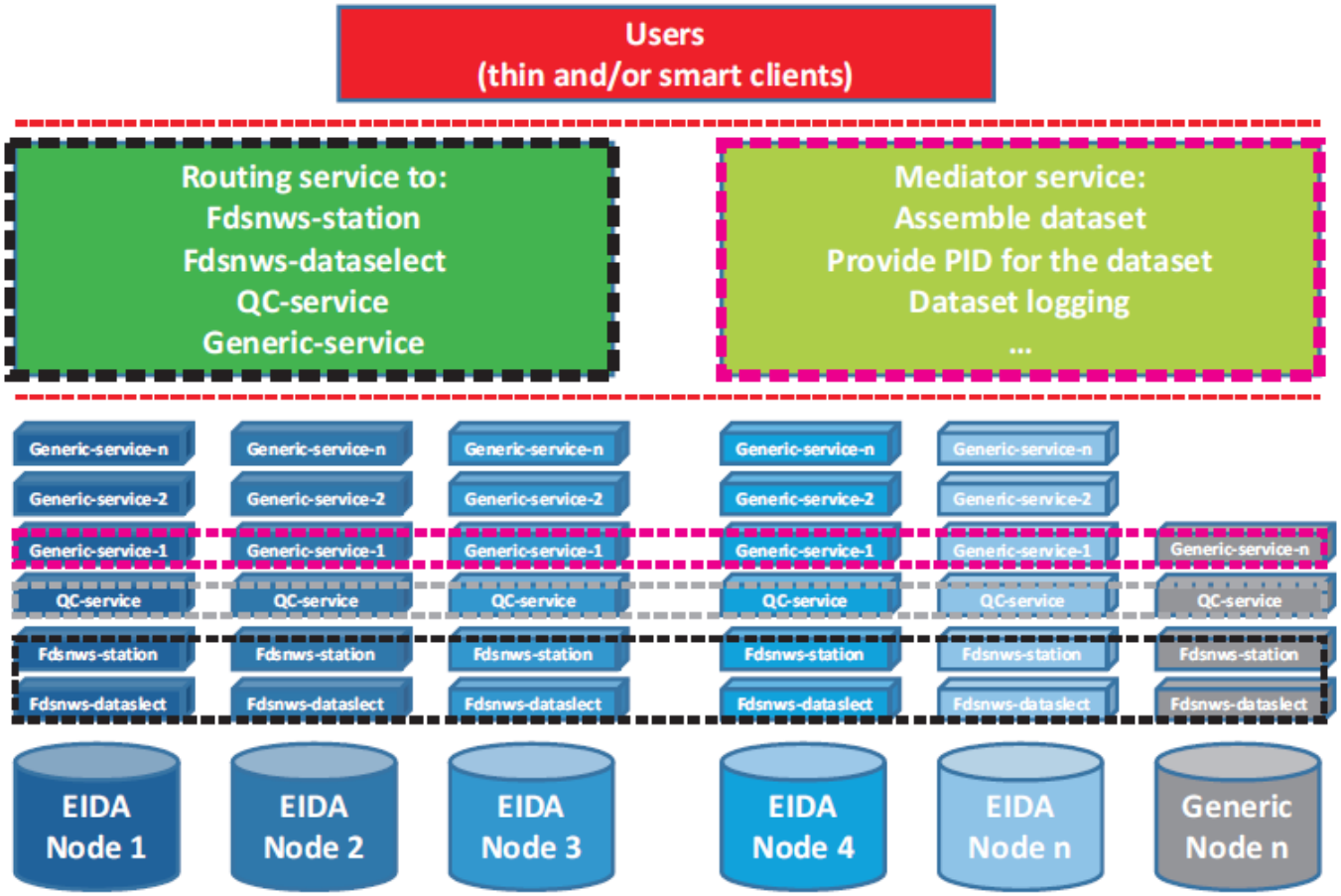
Users: Geoscientists Etc…



EIDA Next Generation

A federation of data and services

Service status:
In production/beta
Under development
Design phase



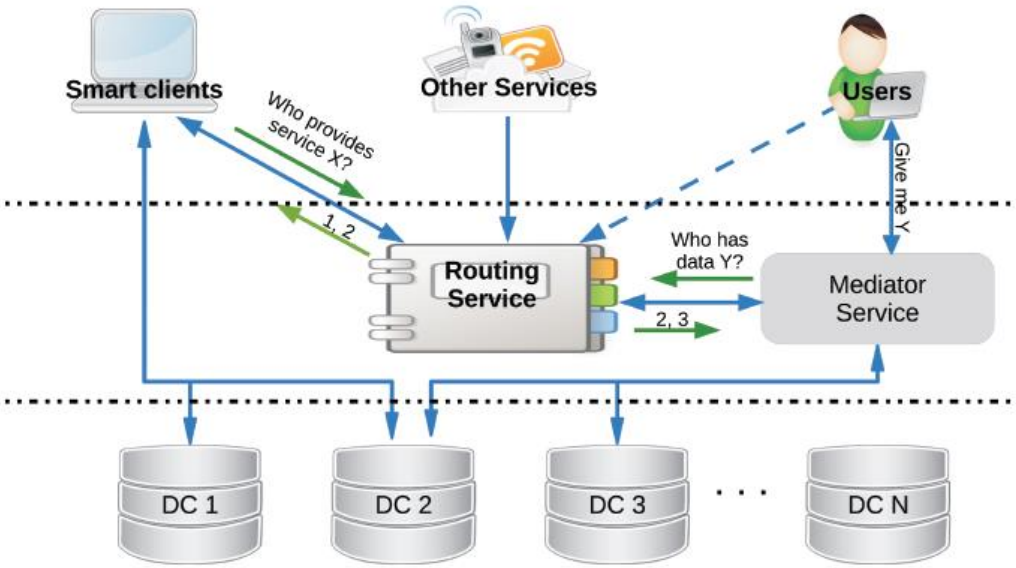
The EIDA routing service

What does a routing service?

- **Provides routing to data (streams)**
- **Routing to services**
- **Routing priorities**

Additional parameters information being discussed:

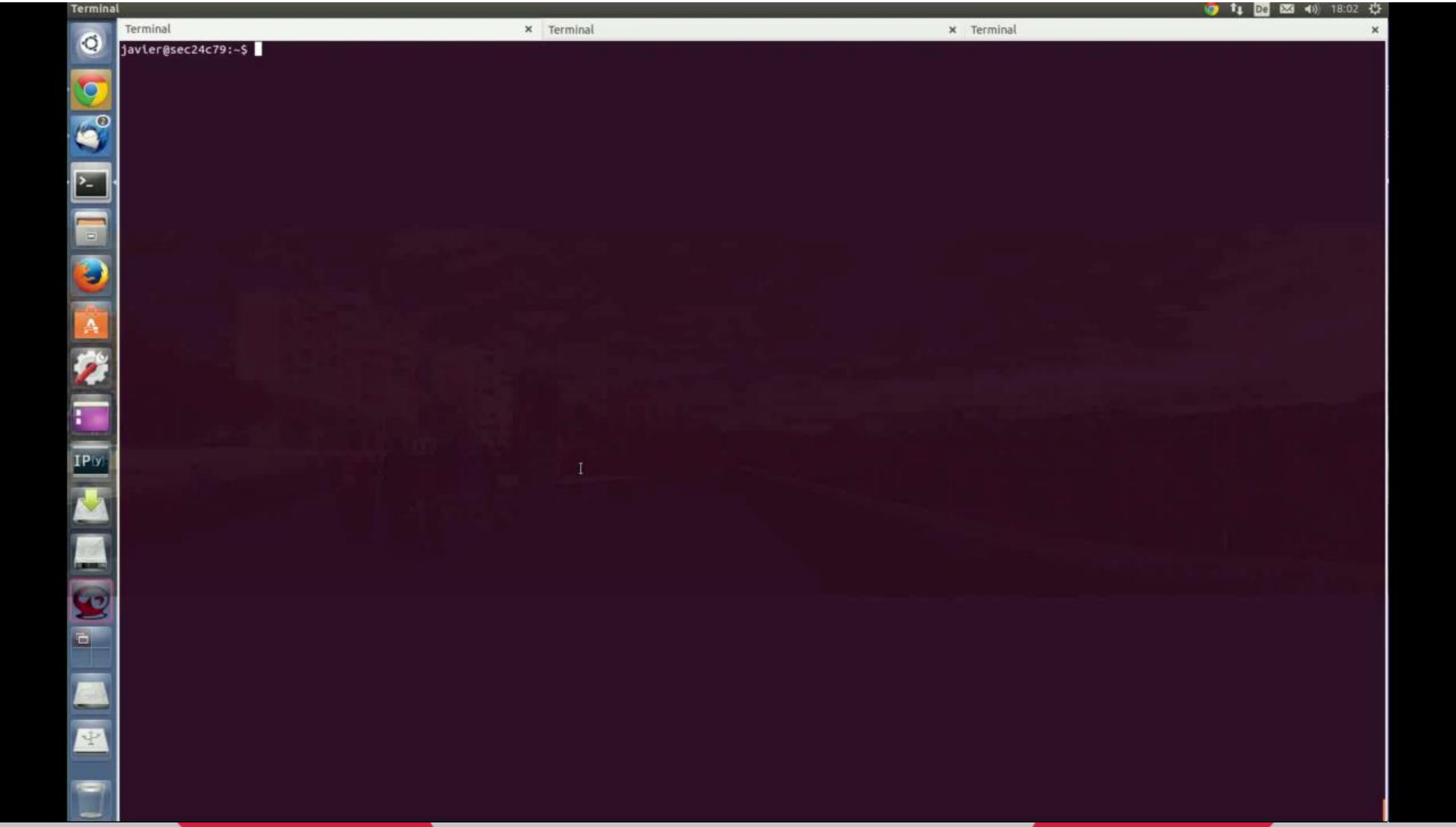
- **Geolocation of data and services**
- **Type of archive (master, validated/non validated copy, etc)**
- **Contact information**



Can also be deployed as a standalone router to run on the client side to create virtual DCs

First stable release 10.2014

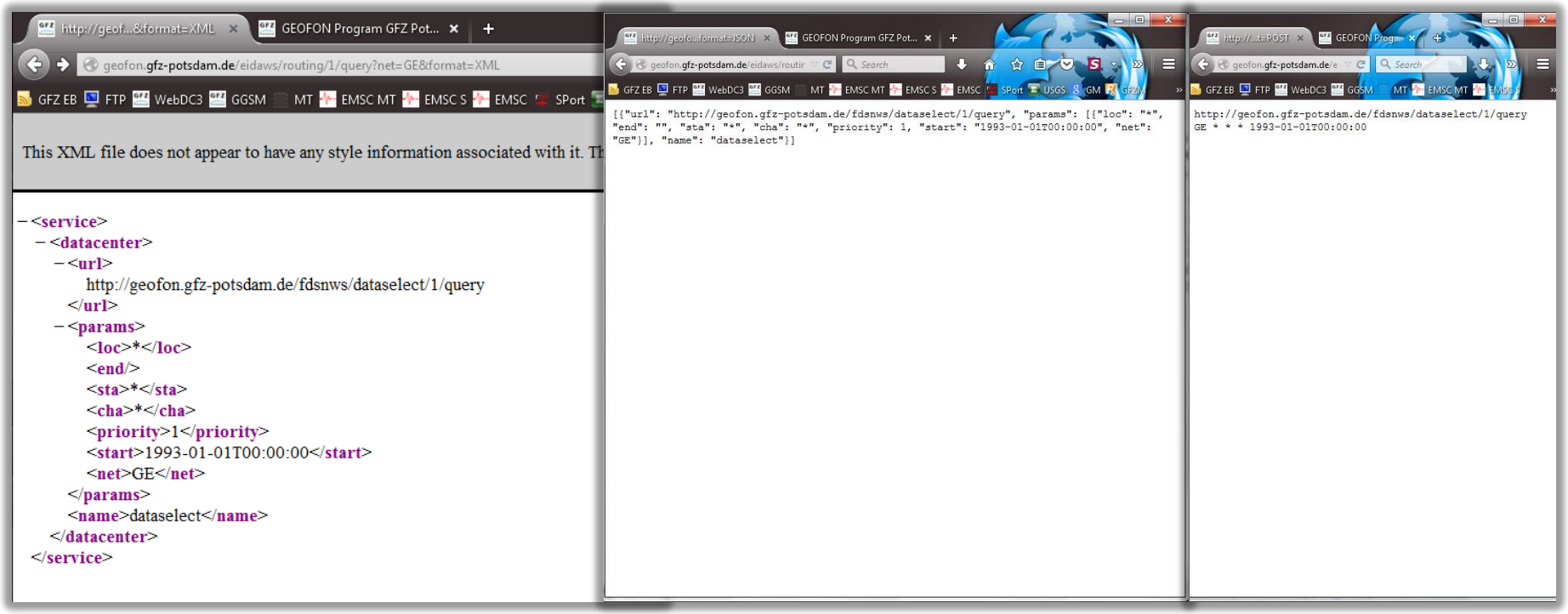
SeisComP3 demo Arclink => EIDA routing + fdsnws



The EIDA routing service output

Provides the routing to data and service according to priority level requested by the user.

- **Formats: XML, JSON, GET, POST**
- **Collapsed (useful in case of simple routes , the majority)**



Input to the EIDA routing service 1

For the initial deployment at EIDA nodes the Arclink routing table has been used as configuration file.

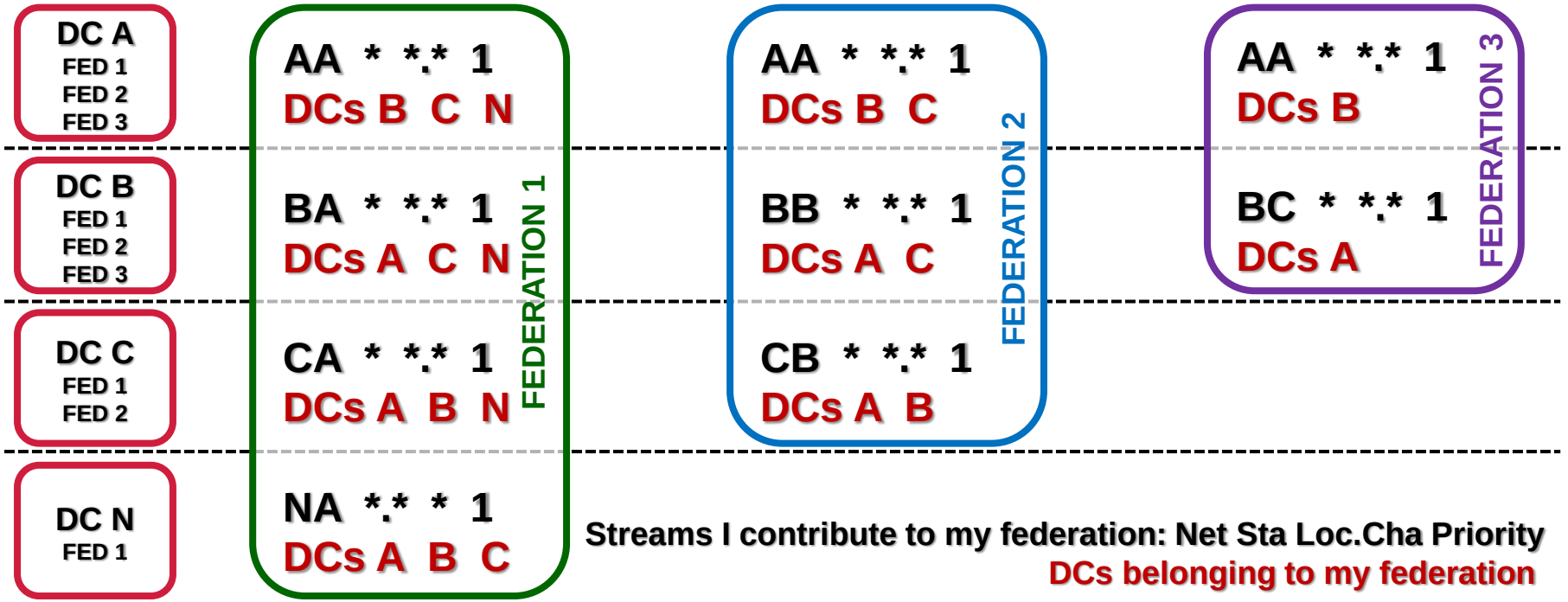
routing-2.xml

```
<?xml version="1.0" encoding="utf-8"?>
<ns0:routing xmlns:ns0="http://geofon.gfz-potsdam.de/ns/Routing/1.0/">
    <ns0:route locationCode="" networkCode="BE" publicID="Route#20140417144427.719649.65791" stationCode="" streamCode="">
        <ns0:arclink address="bhlsa02.knmi.nl:18002" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="BA" publicID="Route#20141104135225.609435.44505" stationCode="" streamCode="">
        <ns0:arclink address="eida.rm.ingv.it:18002" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
        <ns0:seedlink address="eida.rm.ingv.it:18000" priority="1" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="WM" publicID="Route#20130513164519.547084.63954" stationCode="" streamCode="">
        <ns0:arclink address="eida.gfz-potsdam.de:18002" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="WI" publicID="Route#20140819100547.025038.8390" stationCode="" streamCode="">
        <ns0:arclink address="eida.ipgp.fr:18001" end="" priority="1" start="2008-01-01T00:00:00.0000Z" />
        <ns0:seedlink address="eida.ipgp.fr:18000" priority="1" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="JS" publicID="Route#20140110193637.095613.69344" stationCode="" streamCode="">
        <ns0:arclink address="eida.gfz-potsdam.de:18002" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="BW" publicID="Route#20130204144021.621364.1185" stationCode="RWMO" streamCode="">
        <ns0:arclink address="141.84.11.2:18001" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
        <ns0:seedlink address="141.84.11.2:18000" priority="1" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="BW" publicID="Route#20130204144021.62504.1207" stationCode="RTKA" streamCode="">
        <ns0:arclink address="141.84.11.2:18001" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="BW" publicID="Route#20130204144021.624036.1203" stationCode="FURT" streamCode="">
        <ns0:arclink address="141.84.11.2:18001" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
        <ns0:seedlink address="141.84.11.2:18000" priority="1" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="BW" publicID="Route#20130204144021.619343.1168" stationCode="RTSP" streamCode="">
        <ns0:arclink address="141.84.11.2:18001" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
    </ns0:route>
    <ns0:route locationCode="" networkCode="BW" publicID="Route#20130204144021.621703.1187" stationCode="RTSW" streamCode="">
        <ns0:arclink address="141.84.11.2:18001" end="" priority="1" start="1980-01-01T00:00:00.0000Z" />
    </ns0:route>
```

eXtensible Markup Language file | length : 326125 lines : 3565 | Ln : 1 Col : 2 Sel : 0 | Dos\Windows

Input to the EIDA routing service 2

The DC to join a federation should declare what He want to expose (data and services) and who are the other DCs belonging to the federation and with what level of priority. This can be done at different levels: Personal /Institutional, European, Global, …



Some important issues we considered

Simple harvesting of fdsnws-station services will lead to the following ambiguities:

- **Questionable priorities in the routes (the decision stays with the data owner or network operator)**
- **May assume that fdsnws-dataselct and/or other services are running also where fdsnws-station is running**
- **May wrongly assume additional services**
- **May need to interpret mismatching station locations**

A declarative approach would remove all these ambiguities. Disadvantage? Not trivial to maintain as an harvested catalog but definitely providing the correct information to the user.

Proposed approach for the FDSN

A federator should be a service/functionality that allows a DC to join a federation of DCs rather than a service harvesting metadata and aggregating them to a single central db.

Information about data/services locations should be gathered from the federator rather than contacting the fdsnws-station directly.

EIDA would welcome a small task force within WGIII to discuss these issues further together with additional services similarly to the approach taken for the DOIs. In particular the following points:

- **Definition of standard output of a federator/routing service**
- **How to input and maintain the routing information**
- **All technical and political implications**