EIDA
European Integrated Data Archive

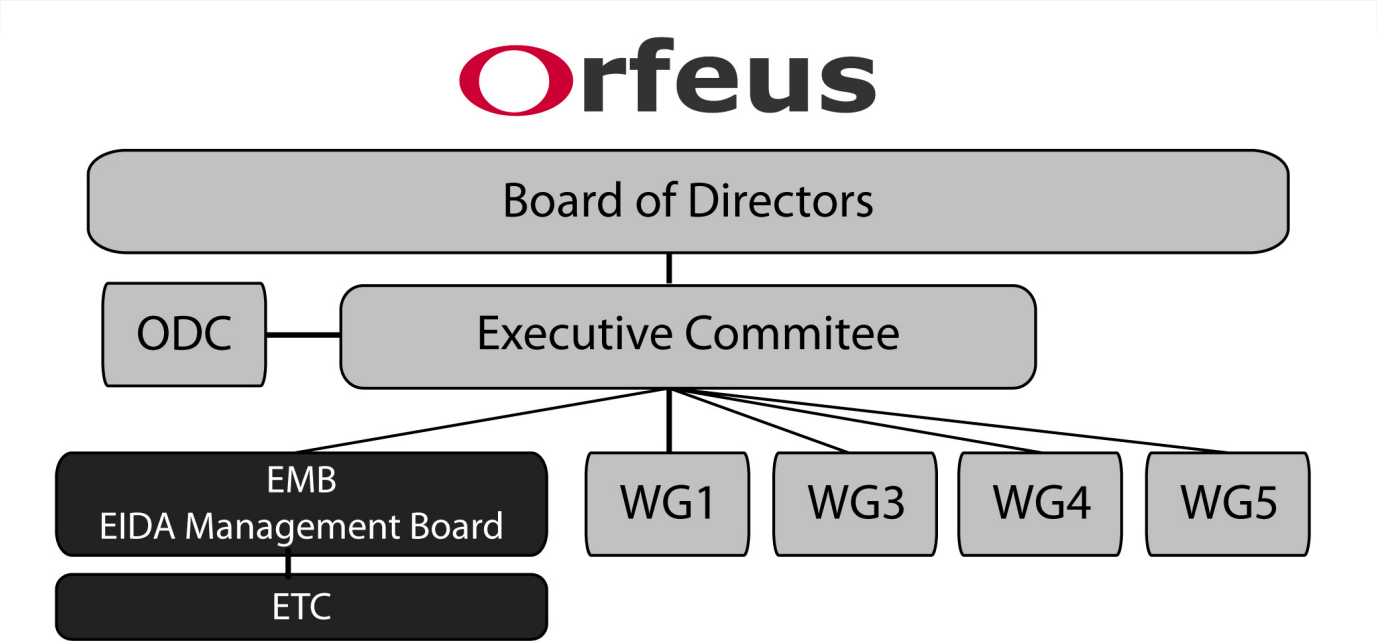
Javier Quinteros and the EIDA team

javier@gfz-potsdam.de

GFZ
Helmholtz Centre
POTSDAM

January 22th, 2015

What is EIDA?



- EIDA is a distributed data centre established to
  - **a)** securely archive seismic waveform data and related metadata, gathered by European research infrastructures, and
  - **b)** provide transparent access to the archives by the geosciences research communities.

Sharing data and metadata

- Current total archive size: ca. 350 TB
- Number of networks: more than 130
- Number of stations: more than 4500

Sharing this data can be real challenging considering two factors:

- Every country/institution has its own stations and interests
- Users would like to see everything integrated and get data in a transparent way
- On the other hand, network operators want credit for the invested resources.

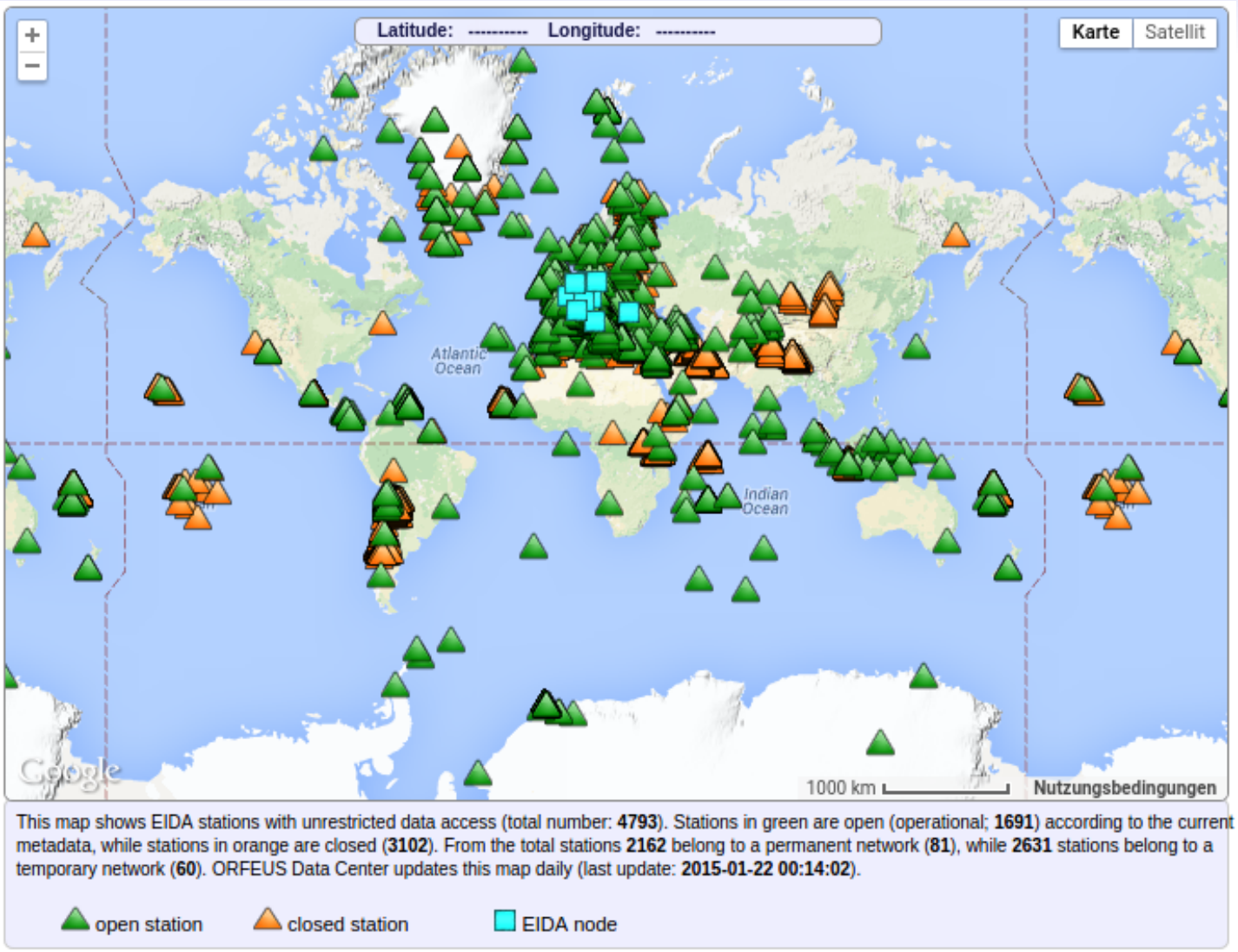
Sharing data and metadata

- Current total archive size: ca. 350 TB
- Number of networks: more than 130
- Number of stations: more than 4500

Sharing this data can be real challenging considering two factors:

- Every country/institution has its own stations and interests
- Users would like to see everything integrated and get data in a transparent way
- On the other hand, network operators want credit for the invested resources.

EIDA stations



How do nodes interact?

GFZ	European, Global, temporary deployments
ODC	European-Mediterranean area (VEBSN)
ETH	Switzerland
INGV	Italy, European-Mediterranean (MedNet)
RESIF	France + Global temporary deployments
BGR	Germany
NIEP	Romania
IPGP	France (volcanological observatories) + Global (GEOSCOPE)
LMU	Germany (BayernNetz)

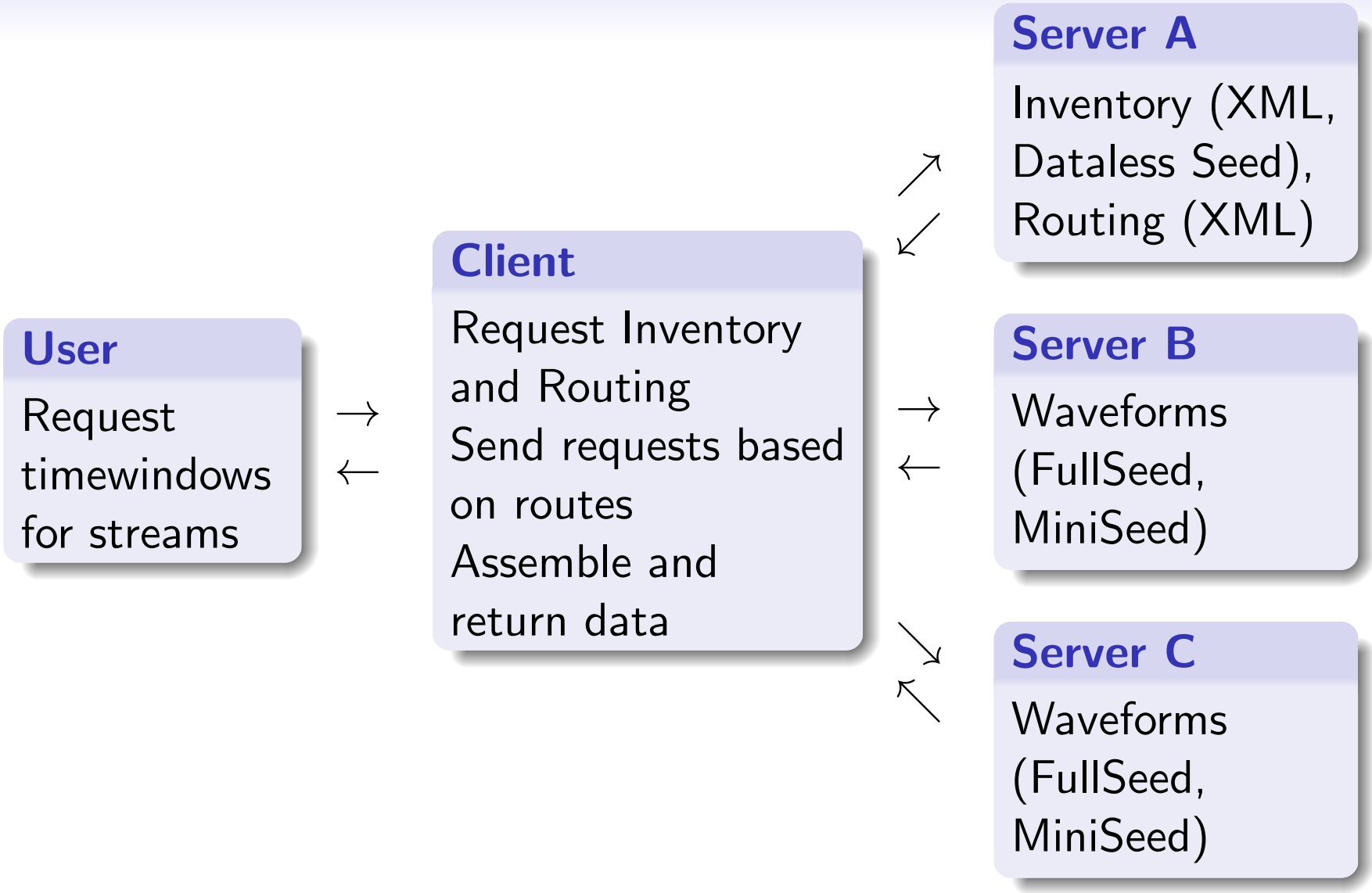
- Technically, EIDA is based on an underlying architecture developed by GFZ to provide transparent access to all nodes' data. Data within the distributed archives are accessible via the **ArcLink** protocol.
- Each node has an Arclink server that fulfills the request
- Each node synchronizes the stream metadata daily

How do nodes interact?

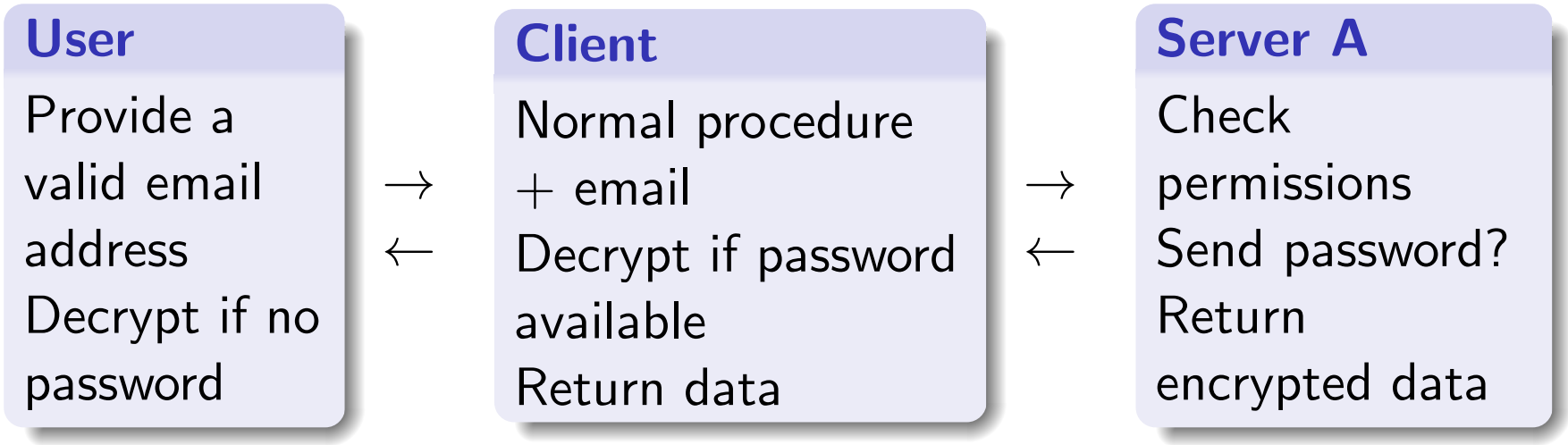
GFZ	European, Global, temporary deployments
ODC	European-Mediterranean area (VEBSN)
ETH	Switzerland
INGV	Italy, European-Mediterranean (MedNet)
RESIF	France + Global temporary deployments
BGR	Germany
NIEP	Romania
IPGP	France (volcanological observatories) + Global (GEOSCOPE)
LMU	Germany (BayernNetz)

- Technically, EIDA is based on an underlying architecture developed by GFZ to provide transparent access to all nodes' data. Data within the distributed archives are accessible via the **ArcLink** protocol.
- Each node has an Arclink server that fulfills the request
- Each node synchronizes the stream metadata daily

How does Arclink work?



In the case of restricted data...



EIDA in the near-future

FDSN web services

Dataselect Service providing waveforms

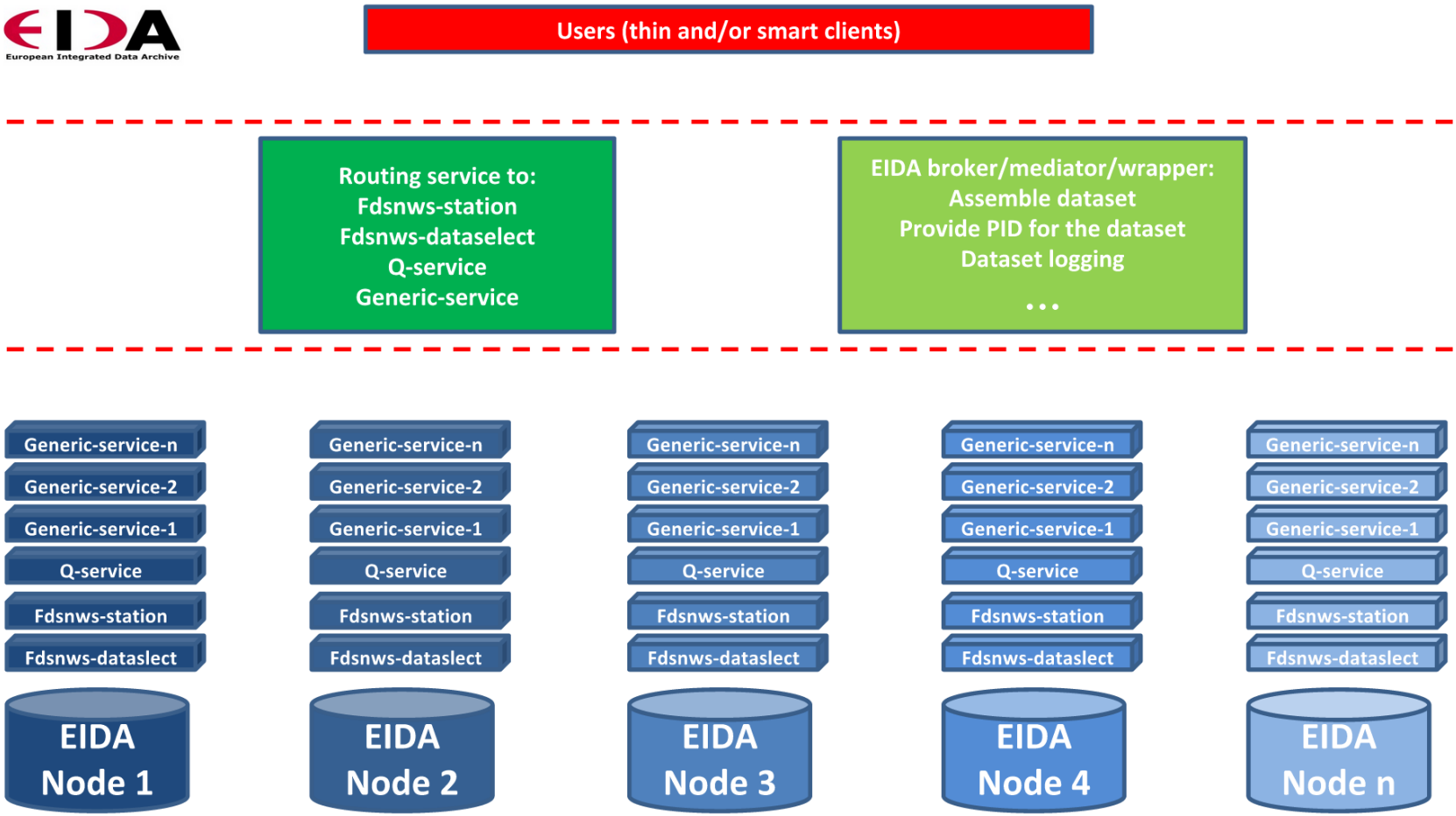
Station Service providing metadata

Event Service providing earthquake information

Security for restricted data

- Basic HTTP digest authentication is the solution proposed in the standard.
- Implemented only by one data centre (RESIF).

EIDA components



Wishlist for our AAI system

Federated approach No usernames and passwords should be exchanged/synchronized.

Non-interactive login Most of our users request the data via scripts or from the command-line. Webisoget could be a solution, but there are drawbacks.

Add users easily Some users do not belong to our institution and a few are not related to any Identity Provider.

Permissions only locally Every data centre (or network PI) is responsible to assign permissions to the data hosted locally.

Validity of credentials The digitally signed credentials should be valid for any site within EIDA. Of course, respecting the time restrictions.

Add new nodes New data centres should have a reasonable way to have an IdP and implement this solution.

What about the compatibility?

Which compatibility should we expect between AAI services/approachs oferred/adopted by GEANT, EUDAT, EPOS?

Cloud services

As part of the EUDAT H2020 project we plan to use some of the EUDAT services.

- Replication and Data Management policies via B2SAFE with KIT (Karlsruhe) as partner
- B2FIND
- PID minting for the datasets
  - Possible relation to DOI assigned to networks?
  - We need to tackle the problem of reproducibility in scientific experiments.