# INGV - EPOS

# Istituto Nazionale di Geofisica e Vulcanologia

# *ICT for Earth Sciences*

***Daniele Bailo***

*ICT@INAF, Nov 2017 Bologna*


ISTITUTO NAZIONALE DI
GEOFISICA E VULCANOLOGIA
Sismologia
Vulcanologia
Ambiente

# Infrastrutture e banche dati

- Realtà complessa
  - c.ca 120 Infrastrutture
  - c.ca 130 banche dati
  - Iniziative di calcolo in-house e con altri enti (CINECA)
- Varie iniziative di razionalizzazione
  - Progetti Europei → EPOS, EMSO)
  - Gruppo Polidat (data policy)
  - DPC -Protezione Civile

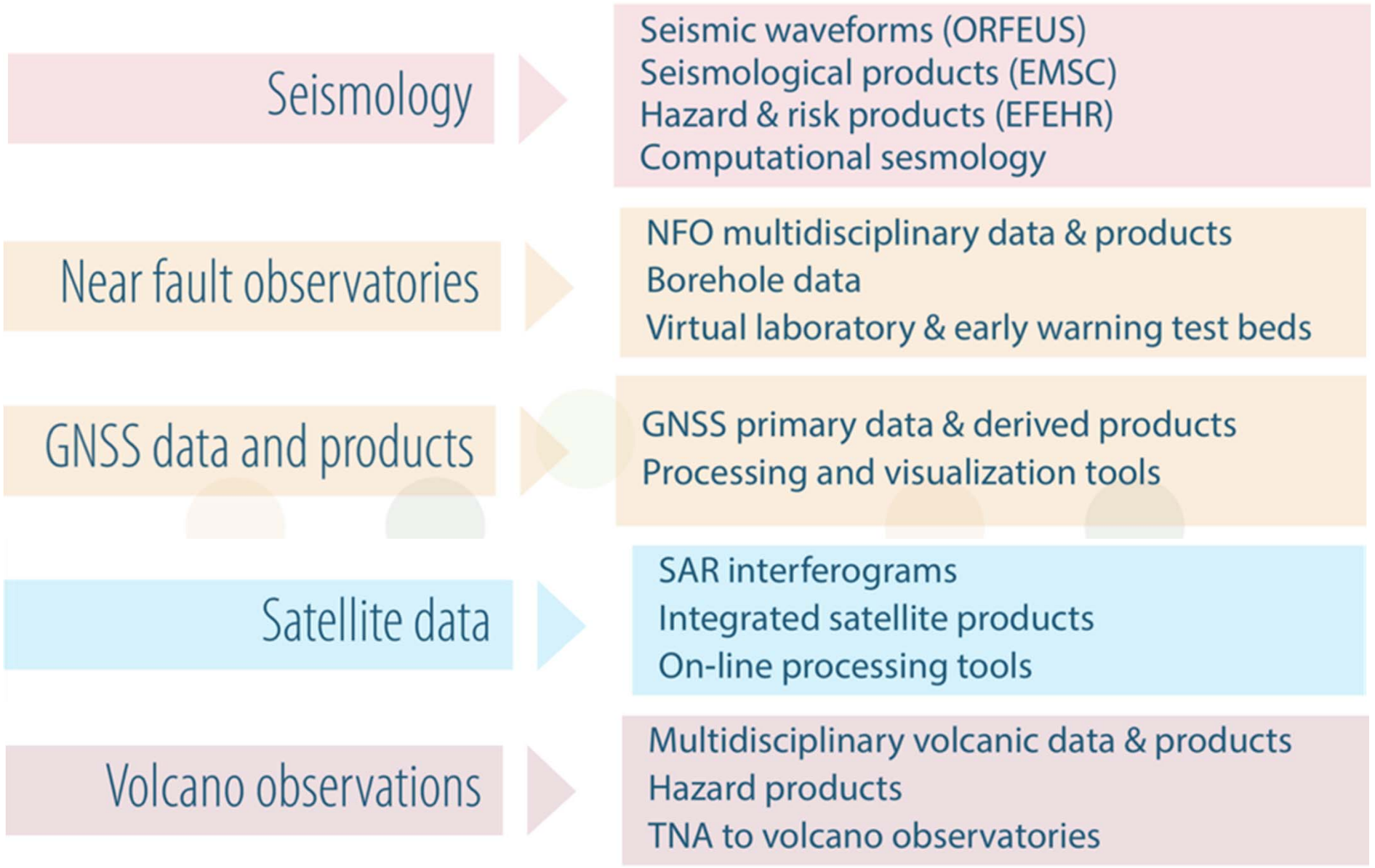

**Geology, geo-hazards, geo-resources, environmental processes in general do not respect national boundaries and scientific disciplines**

Seamless, trans-national integration of measurements and data is crucial for optimal research and related activities

To **understand** the Earth's chemical and physical processes
to **forecast** the events
to **assess** the hazard and **mitigate** the risk
to sustainably **exploit geo-resources**

**integrated multidisciplinary research approach is needed**

# EPOS TC Services 1/2



# EPOS TC Services 2/2

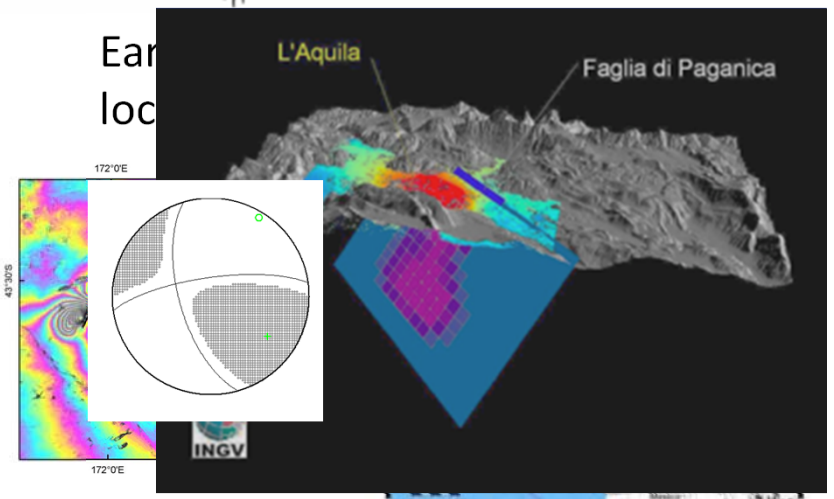
| | |
|---|---|
| Anthropogenic hazards | Data for AH episodes<br>Multi-hazard simulator - multi-risk assessment<br>AH data visualisation |
| Geomagnetic observations | Global and regional geomagnetic models<br>Magnetotelluric data |
| Geological information and modeling | Geological multi-scale data<br>Integrated geological maps<br>Borehole visualization |
| Multi-scale laboratories | Experimental & analogue data<br>TNA to experimental & micro-analytical facilities |
| Geo energy test beds for low carbon energy | Geo energy test beds<br>Access to in-situ GETB exsperiments |

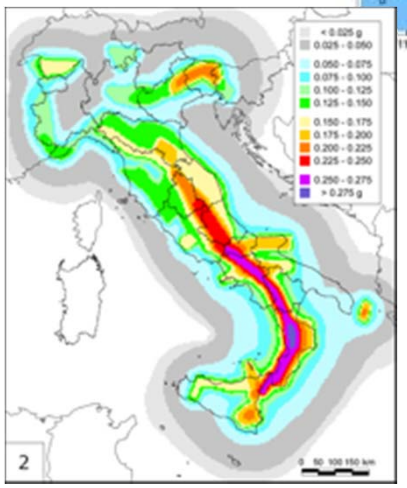
# Access to Data Products (Taxonomy)

- **Level 0**: raw data, or basic data
- **Level 1**: data products coming from nearly automated procedures
- **Level 2**: data products resulting by scientists' investigations
- **Level 3**: integrated data products coming from complex analyses or community shared products
- **Level 4**. Software, IT tools

seismograms

Ear
loc



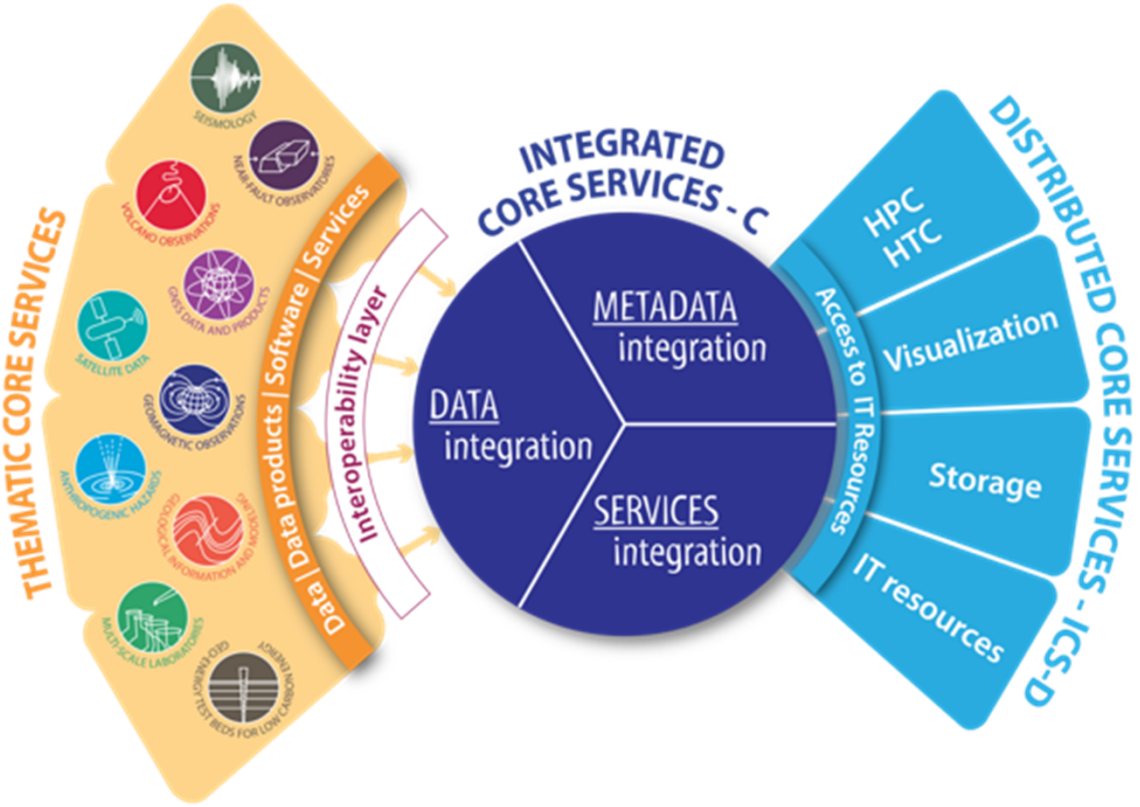
Seismic hazard map

# … summipg up EPOS


NRIs and scientific communities ensure the **competences and resources** for collecting and analysing data and for maintaining territorial observation systems

TCS are responsible for **integrating data, metadata and services** from various infrastructures for each discipline

# … summipg up EPOS

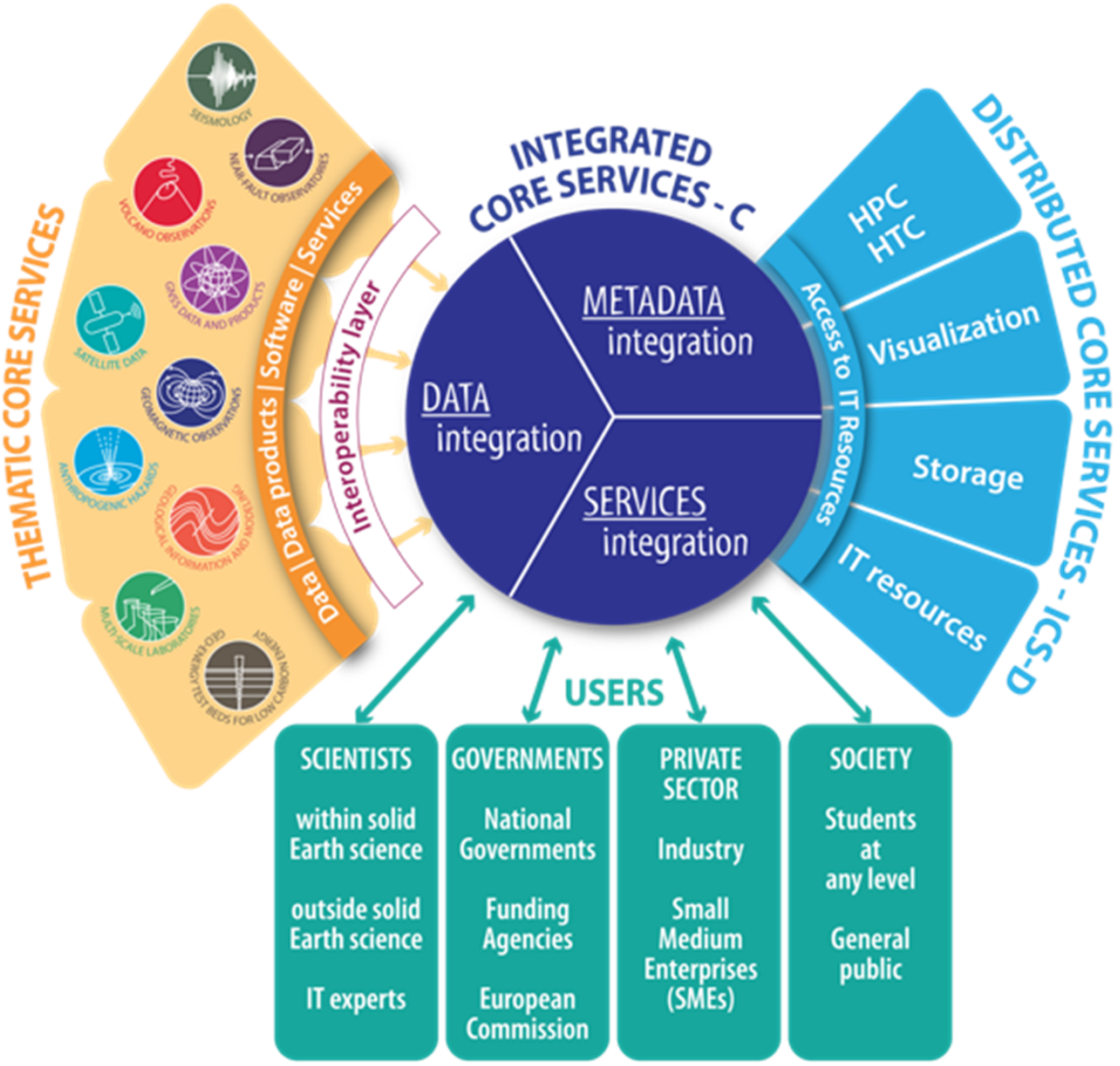



NRIs and scientific communities ensure the **competences and resources** for collecting and analysing data and for maintaining territorial observation systems

TCS are responsible for **integrating data, metadata and services** from various infrastructures for each discipline

ICS provide a **new interface** that

# … summipg up EPOS



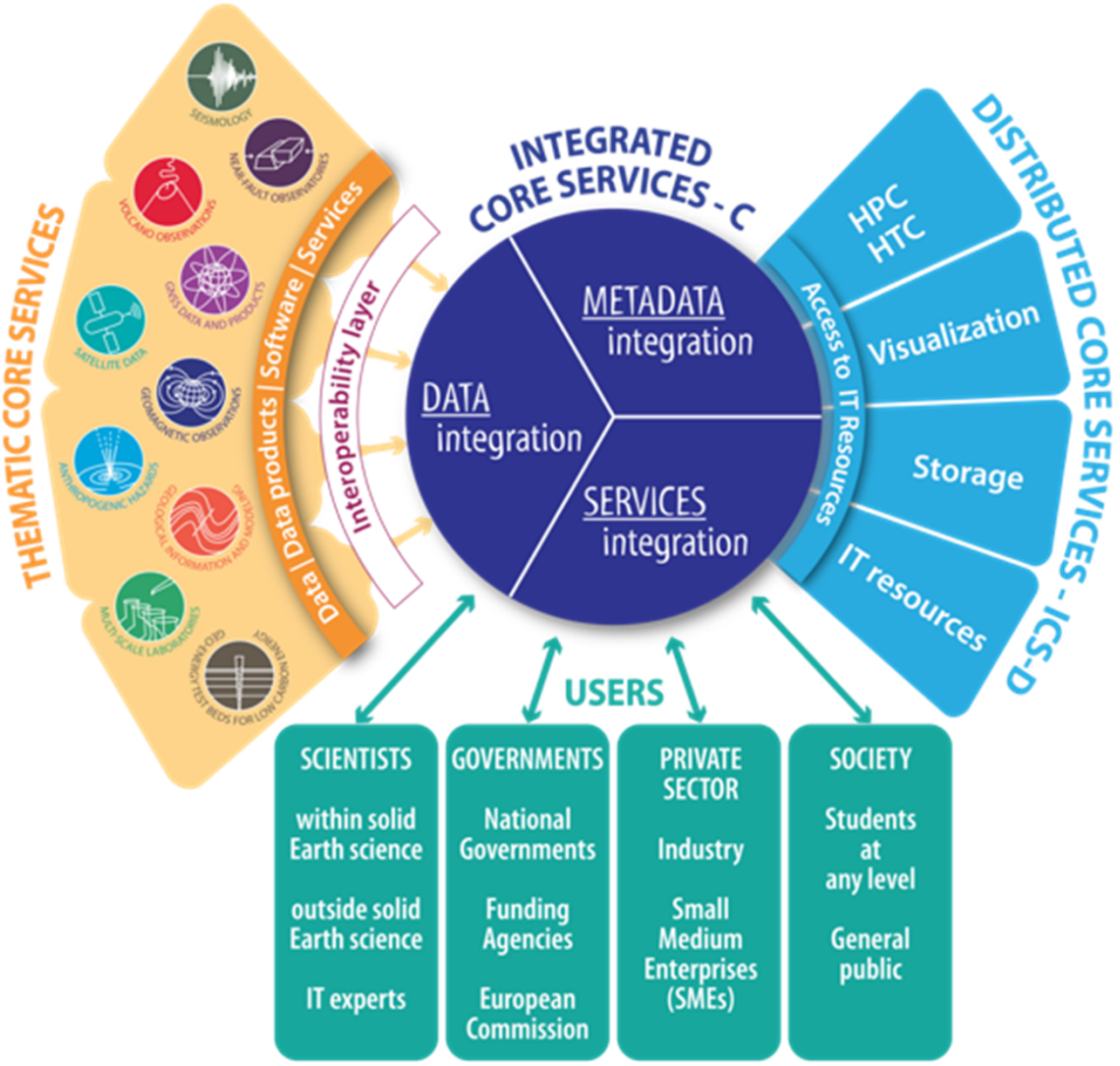

NRIs and scientific communities ensure the **competences and resources** for collecting and analysing data and for maintaining territorial observation systems

TCS are responsible for **integrating data, metadata and services** from various infrastructures for each discipline

ICS provide a **new interface** that by adopting data access policies aligned to **Open Science principles**, provides data and products in a **FAIR*** form for users

***Findable, Accessible, Interoperable, and Re-usable**

# … summipg up EPOS




NRIs and scientific communities ensure the **competences and resources** for collecting and analysing data and for maintaining territorial observation systems

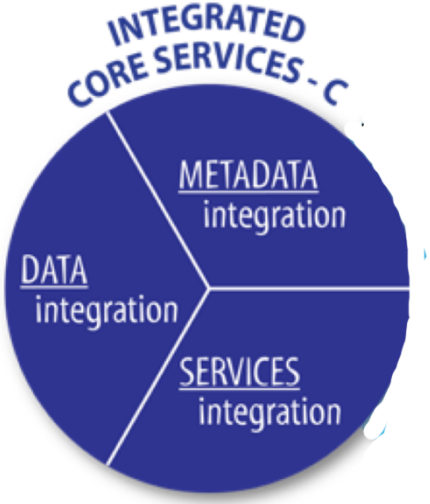
TCS are responsible for **integrating data, metadata and services** from various infrastructures for each discipline

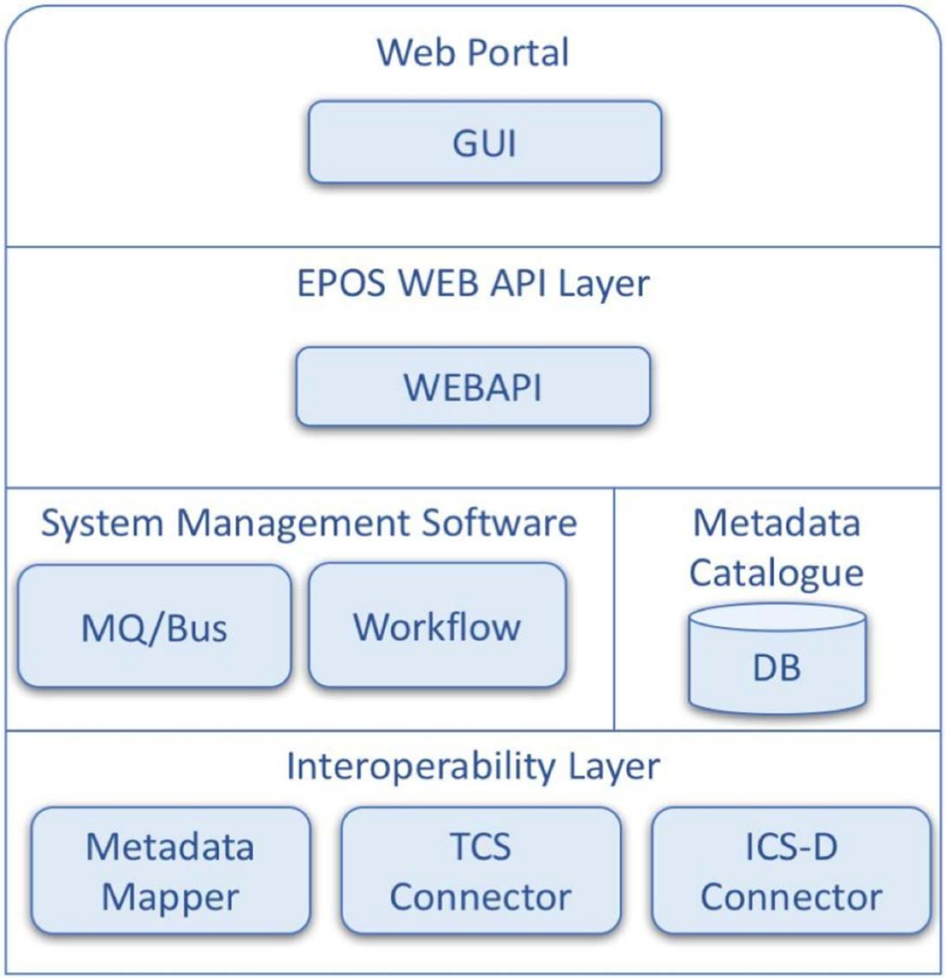
ICS provide a **new interface** that by adopting data access policies aligned to **Open Science principles**, provides data and products in a **FAIR*** form for users

Foster scientific, technological and ICT innovation for successfully **addressing global grand challenges in Earth science**

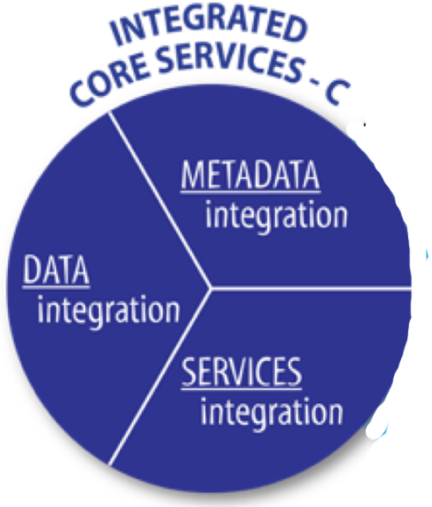
***Findable, Accessible, Interoperable, and Re-usable**

# Technical Challenges



# Technical Challenges



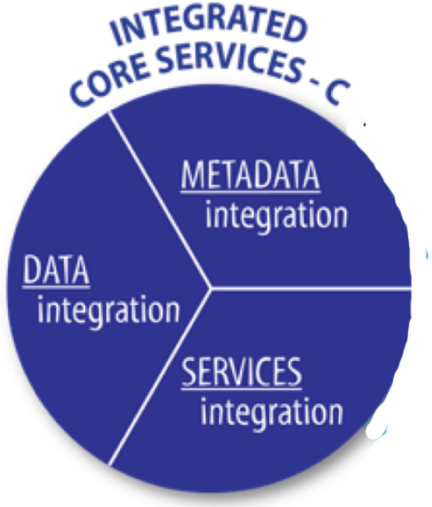

## METADATA

- Metadata model - CERIF
- Metadata description
- Mapping & ingestion

## DATA

- Harmonisation
- Access (web services)
- P.I.D. / DOI
- Provenance

# Technical Challenges



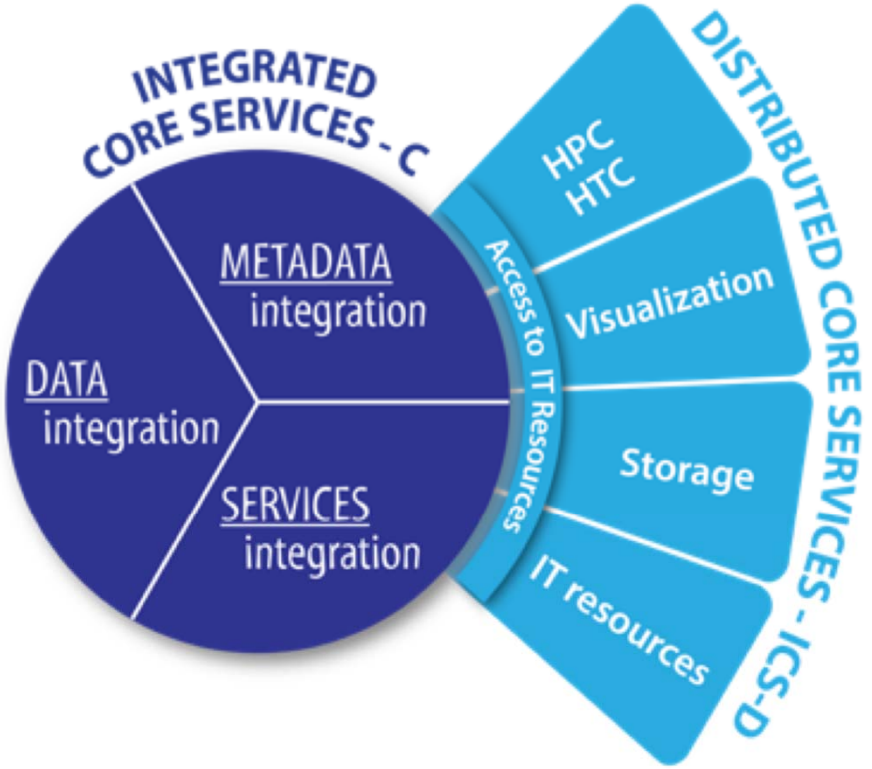

## INTEROPERABILITY

- Community Metadata Standards
- Community not ready
  (LONG WAY TO BE F.A.I.R. ?? 😱 )

## AAAI

- Data policy (EOSC, EPOS)
- AAAI technology (AARC)

# Technical Challenges



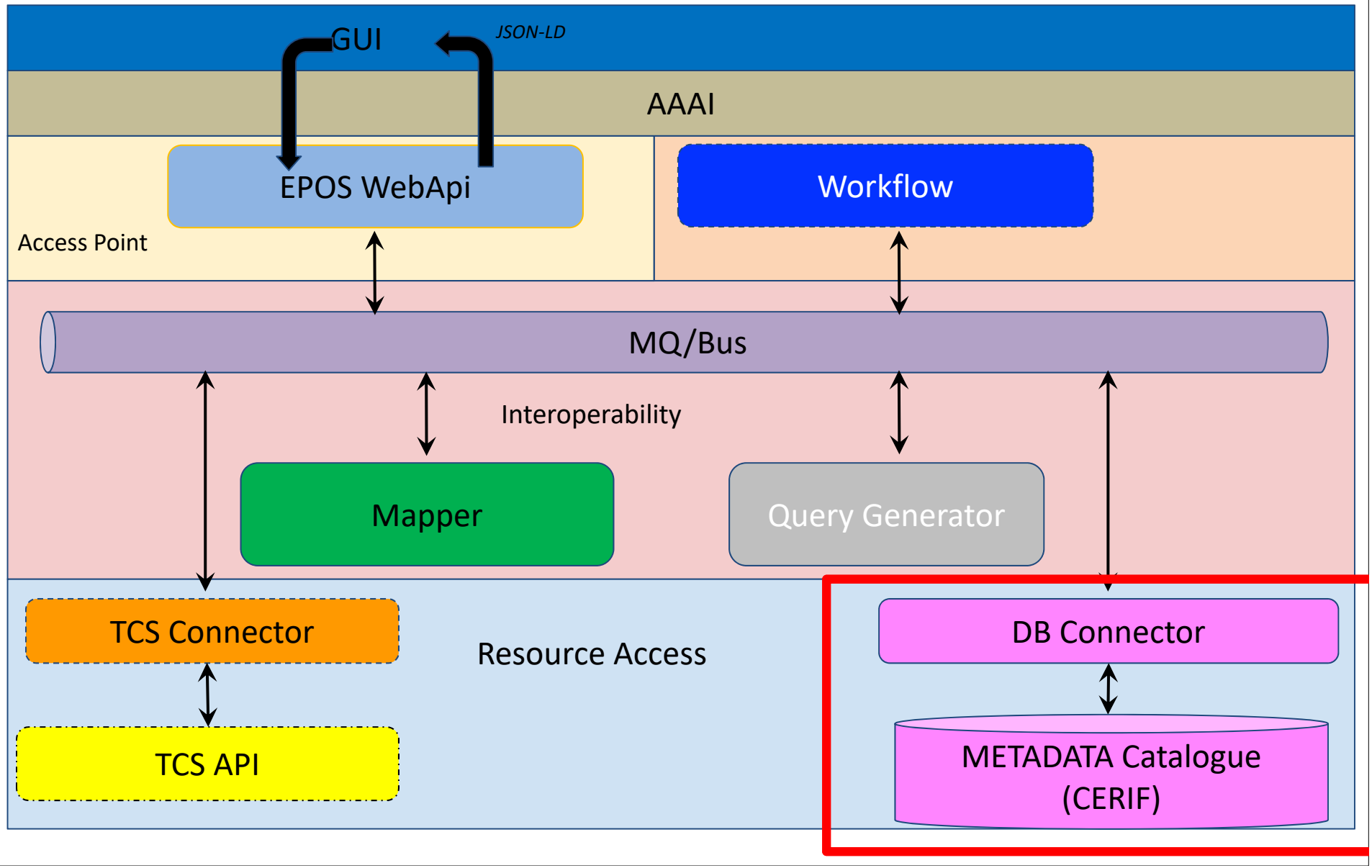

## CALCOLO E SERVIZI

- WORKFLOW MANAGEMENT
- CALCOLO SCIENTIFICO
- SUPPORTO E SOSTENIBILITA'
- **Quali procurement policies?**

# ICS – concepts & components

- Microservice-like architecture
- Main queue
- Metadata catalogue
- Orchestrator / workflow

# Architecture in practice - ICS



# TCS Generic architecture


ICS

API / web service

TCS system

Data/metadata catalogue

National network

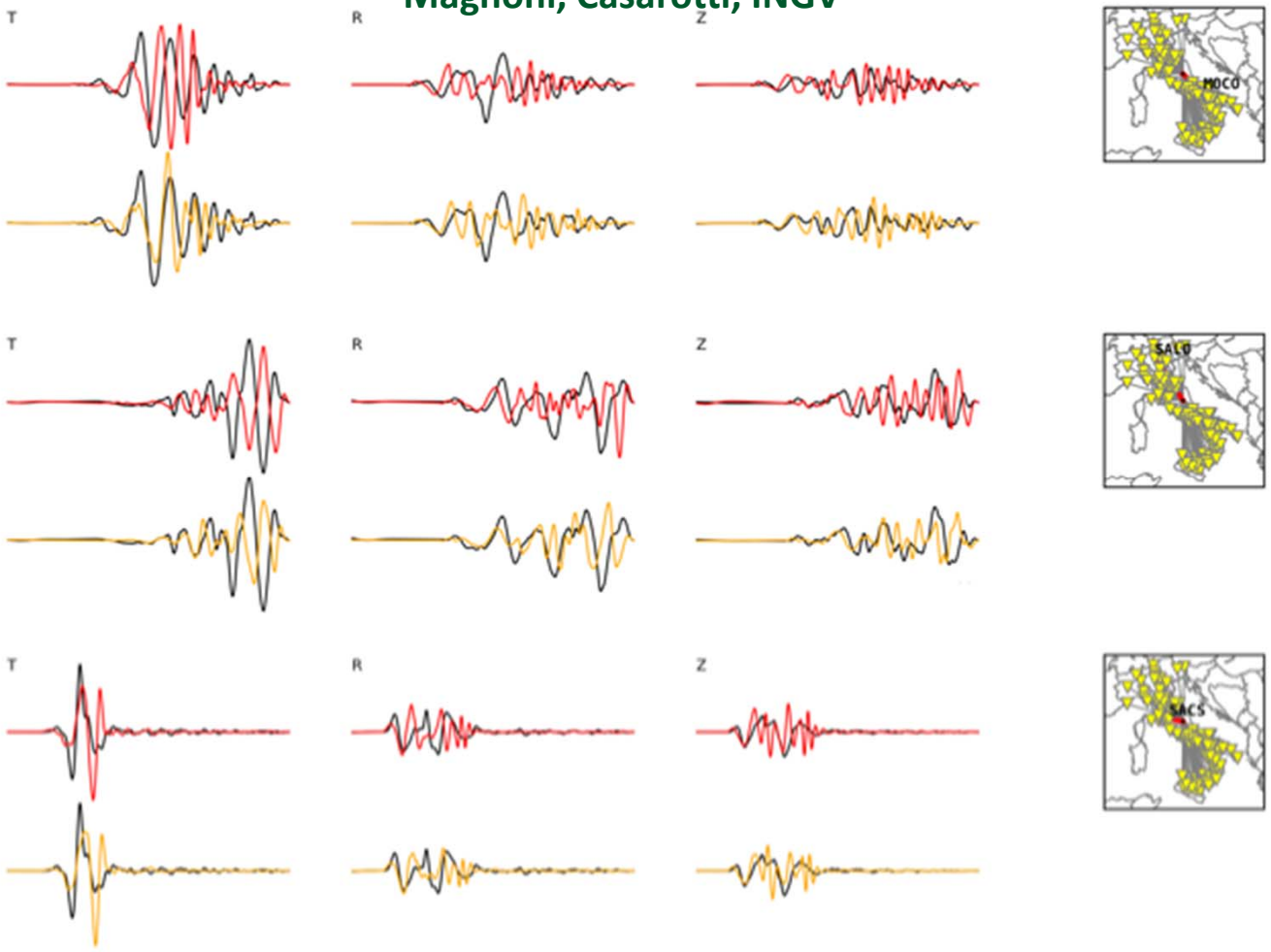
National repository

Local HPC

# VERCE Use case

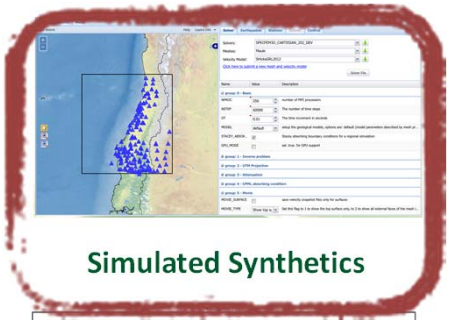
Virtual Earthquake and Seismology Research Community in Europe

**Comparison between SYNTHETICS and DATA.**
**Magnoni, Casarotti, INGV**

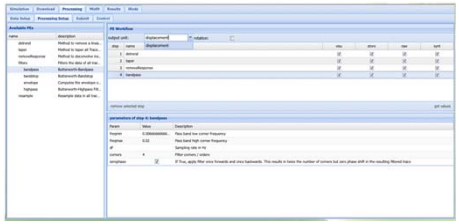


# Misfit Calculation

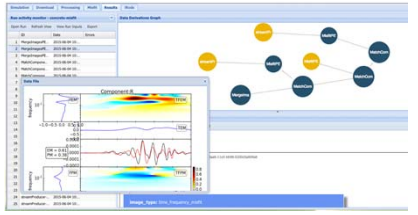
## Misfit between SYNTHETICS and DATA



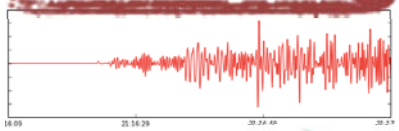
Simulated Synthetics

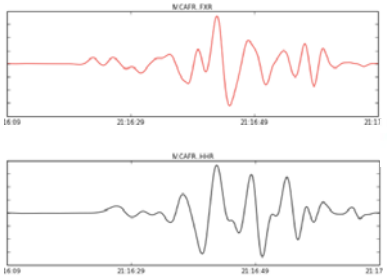


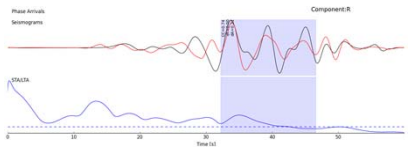
Data/Synt Processing

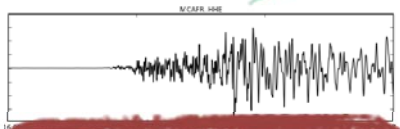


Misfit Analysis

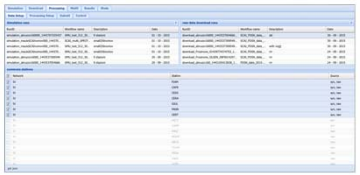


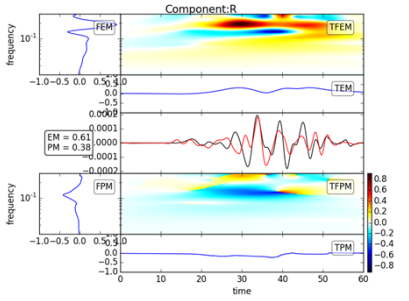






Data Download (FDSN)



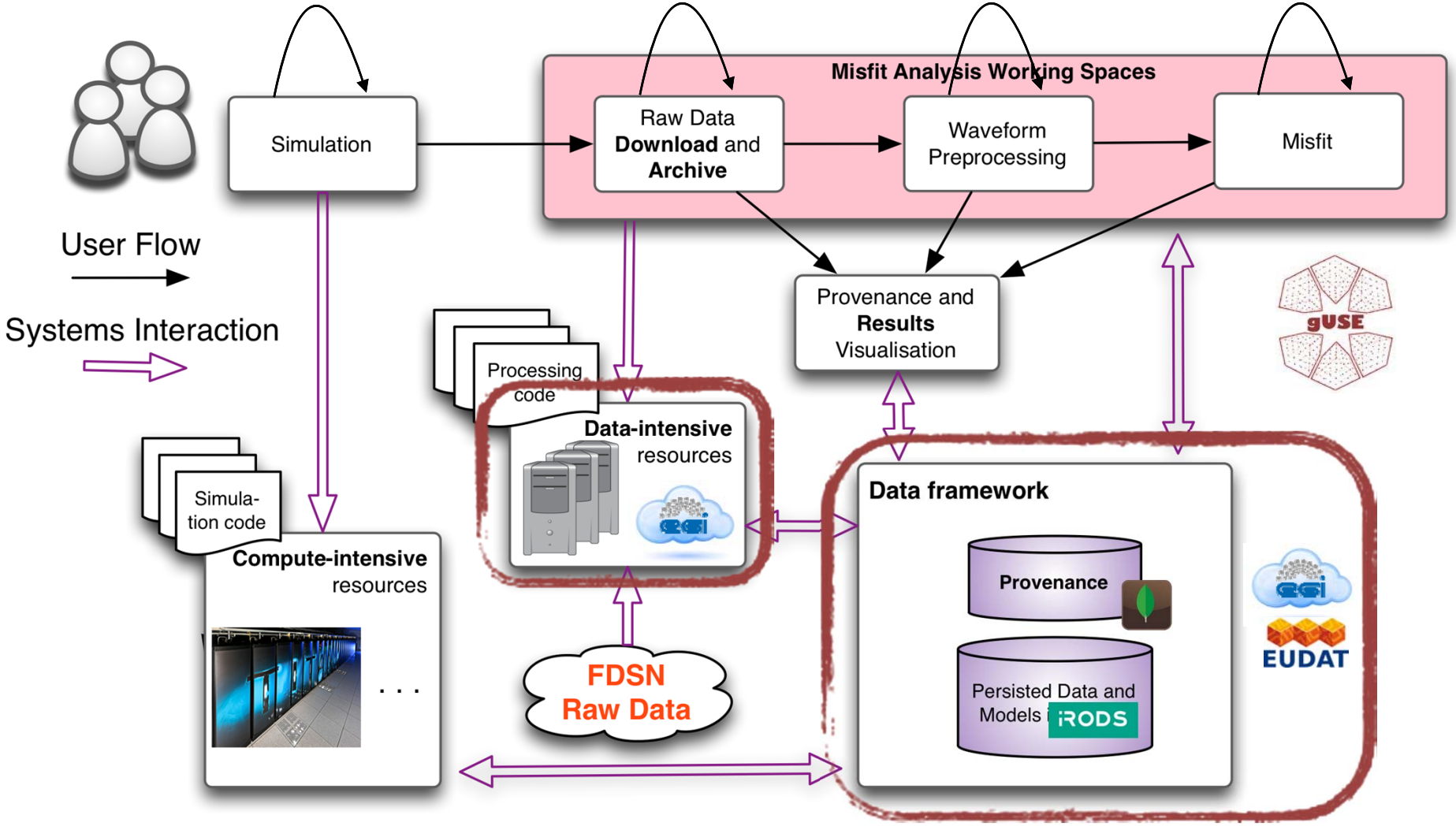



dispel4py

ObsPy

A Python Framework for Seismology

# Gateway's Interaction Flows



# Conclusion – lessons learnt

1. Co-development
2. Constant & continuous updates / discussions with communities
3. “Steering” & Synergy with other initiatives (EOSC, EGI, ENVRI, EUDAT, RDA…)
4. Accept the fact that the architecture and the challenge are complex
5. Procurement policies

# Thank You

**WebSite**


*www.epos-eu.org*

**Newsletter**


*www.epos-eu.org/newsletter*

**R.I.D.E.**


*www.epos-eu.org/ride*

**Epos Social**


www.epos-eu.org

# Key is the metadata

Researchers, research managers, innovators, media

**User Model**

User interaction with the system:
AAAI, interaction workflow, multlinguality

**Processing Model**

Describes functions of processing environment etc.

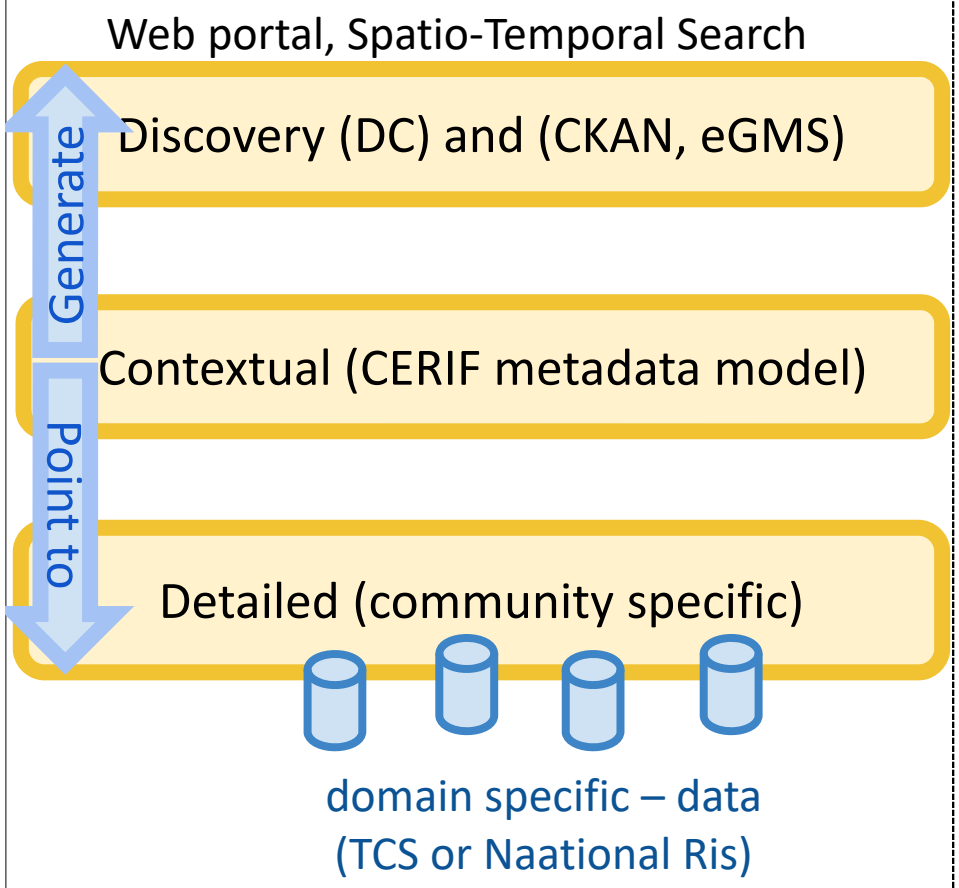
Describes research data and ancillary info (instituions)

**Data Model**

Representing ICT resources

**Resource Model**

Complete ICT environment for research




# Metadata model

- *3-layers*
- *Manage the complexity (see previous slide)*
- *CERIF (formal syntax, declared semantics)*

***Functions***

*-Discovery, selection*

*-Impact assessment*

*-Manage interoperability*

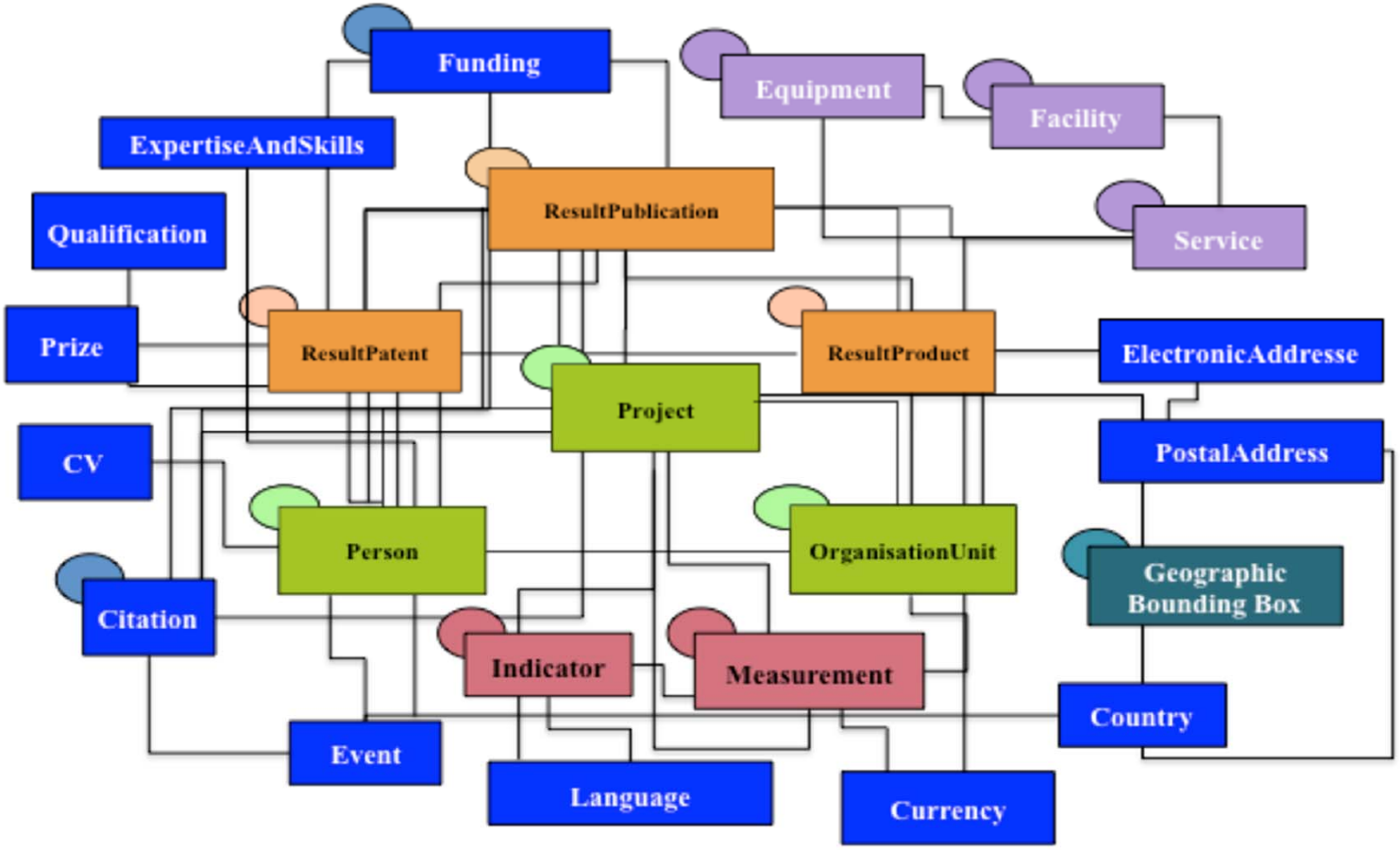
***Issues (selection of)***

- *heterogeneity*
- *Many standards to be mapped from TCS*
- *Ontologies*
- *Complex work*

Contextual (**CERIF** metadata model)

(http://www.eurocris.org/)

**Common European Research Information Format**



# Definition: AAAI

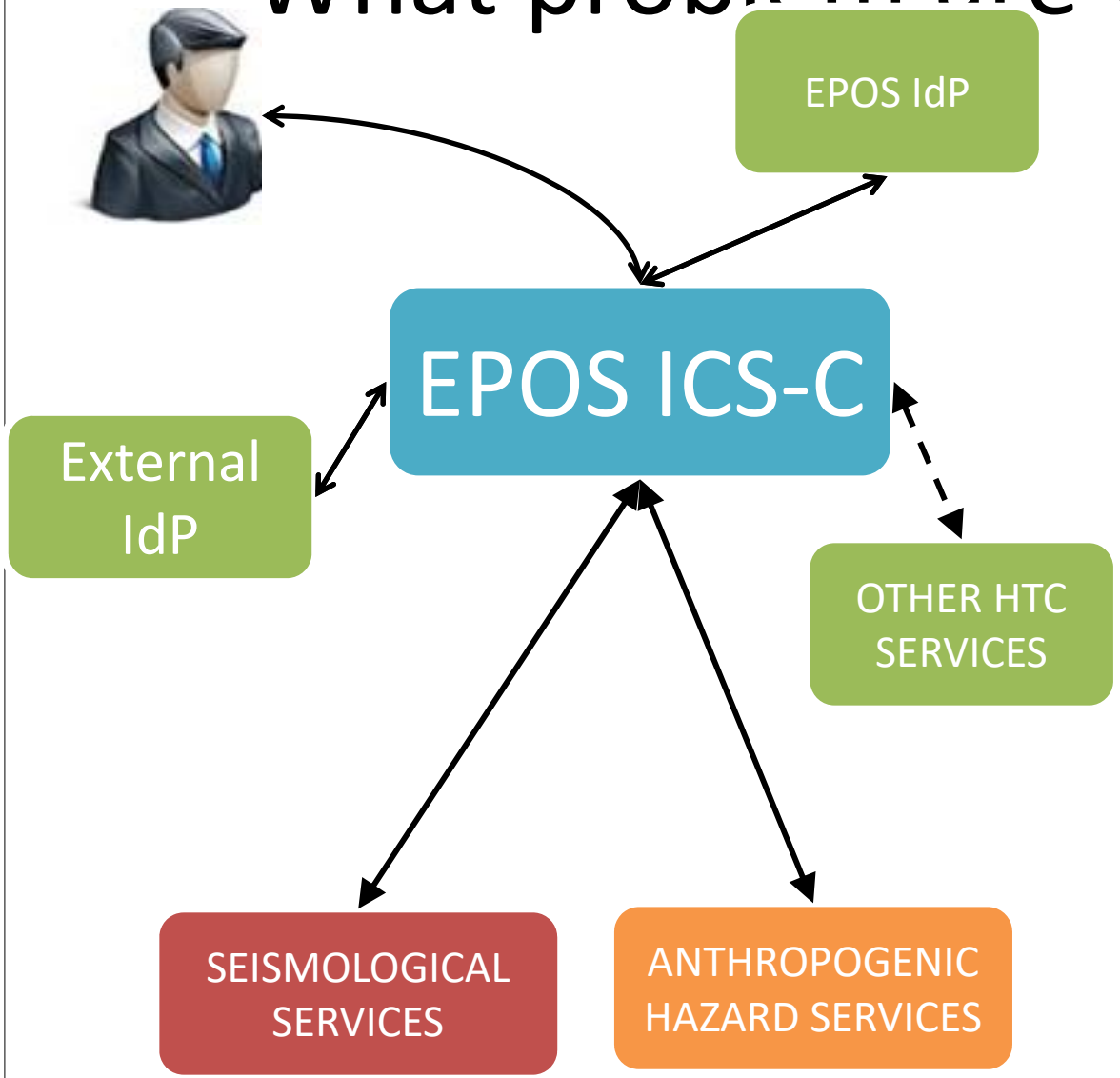
Authentication → “who are you?”

Authorization → “what are you allowed to do?”

Accounting → log actions -> provenance

Infrastructure → …Infrastructure...

# What problem are we facing?




*User wants:*

1 single authentication (SSO)

to access all resources

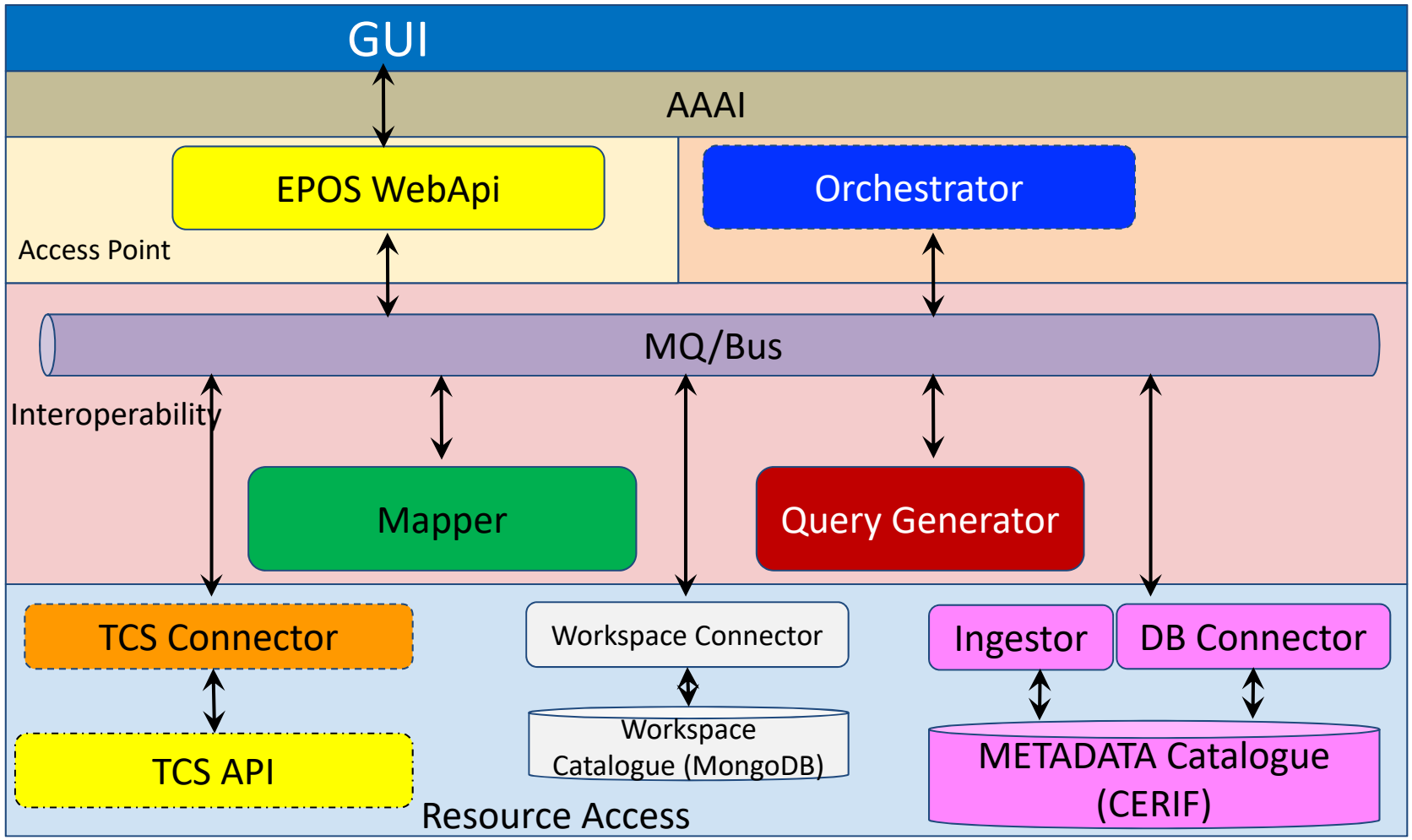
*System must:*

be delegated

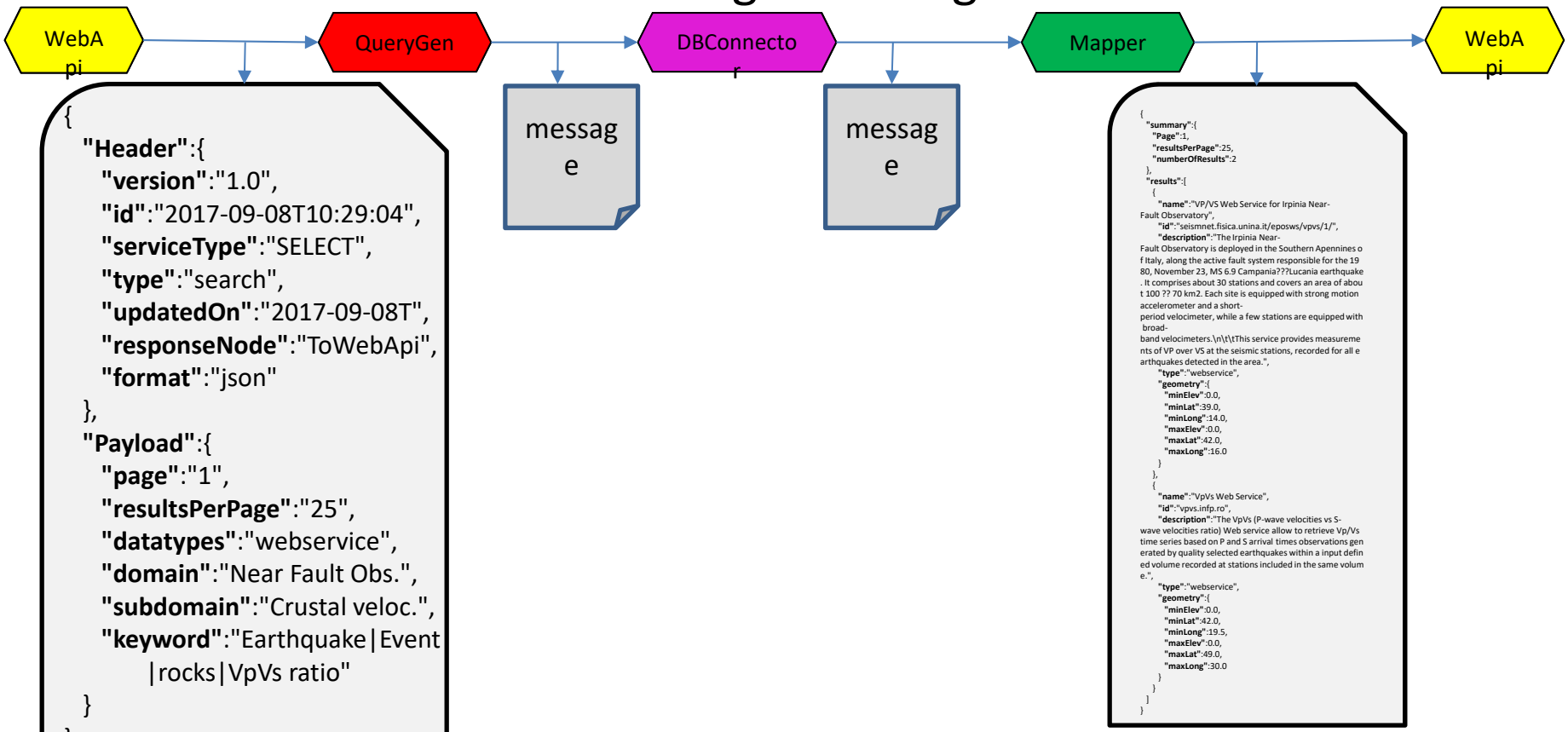
to act on user behalf

*TCS must know*

Who's querying it


GUI
AAAI
EPOS WebApi
Orchestrator
Access Point
MQ/Bus
Interoperability
Mapper
Query Generator
TCS Connector
Workspace Connector
Ingestor
DB Connector
TCS API
Workspace
Catalogue (MongoDB)
METADATA Catalogue
(CERIF)
Resource Access

# Use Case: generic search with paramenters
# Message exchange





```
{
  "Header":{
    "version":"1.0",
    "id":"2017-09-08T10:29:04",
    "serviceType":"SELECT",
    "type":"search",
    "updatedOn":"2017-09-08T",
    "responseNode":"ToWebApi",
    "format":"json"
  },
  "Payload":{
    "page":"1",
    "resultsPerPage":"25",
    "datatypes":"webservice",
    "domain":"Near Fault Obs.",
    "subdomain":"Crustal veloc.",
    "keyword":"Earthquake|Event
        |rocks|VpVs ratio"
  }
}
```

```
{
  "summary":{
    "Page":1,
    "resultsPerPage":25,
    "numberOfResults":2
  },
  "results":[
    {
      "name":"VP/VS Web Service for Irpinia Near-
Fault Observatory",
      "id":"seismnet.fisica.unina.it/eposws/vpvs/1/",
      "description":"The Irpinia Near-
Fault Observatory is deployed in the Southern Apennines o
f Italy, along the active fault system responsible for the 19
80, November 23, MS 6.9 Campania???Lucania earthquake
. It comprises about 30 stations and covers an area of abou
t 100 ?? 70 km2. Each site is equipped with strong motion
accelerometer and a short-
period velocimeter, while a few stations are equipped with
 broad-
band velocimeters.\n\t\tThis service provides measureme
nts of VP over VS at the seismic stations, recorded for all e
arthquakes detected in the area.",
      "type":"webservice",
      "geometry":{
        "minElev":0.0,
        "minLat":39.0,
        "minLong":14.0,
        "maxElev":0.0,
        "maxLat":42.0,
        "maxLong":16.0
      }
    },
    {
      "name":"VpVs Web Service",
      "id":"vpvs.infp.ro",
      "description":"The VpVs (P-wave velocities vs S-
wave velocities ratio) Web service allow to retrieve Vp/Vs
time series based on P and S arrival times observations gen
erated by quality selected earthquakes within a input defin
ed volume recorded at stations included in the same volum
e.",
      "type":"webservice",
      "geometry":{
        "minElev":0.0,
        "minLat":42.0,
        "minLong":19.5,
        "maxElev":0.0,
        "maxLat":49.0,
        "maxLong":30.0
      }
    }
  ]
}
```

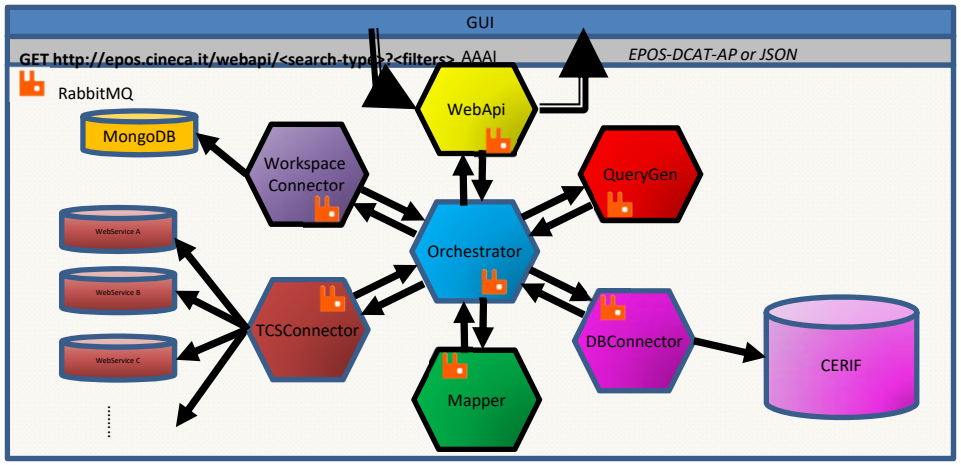
# System Configuration

*Centralised* Configuration (current)

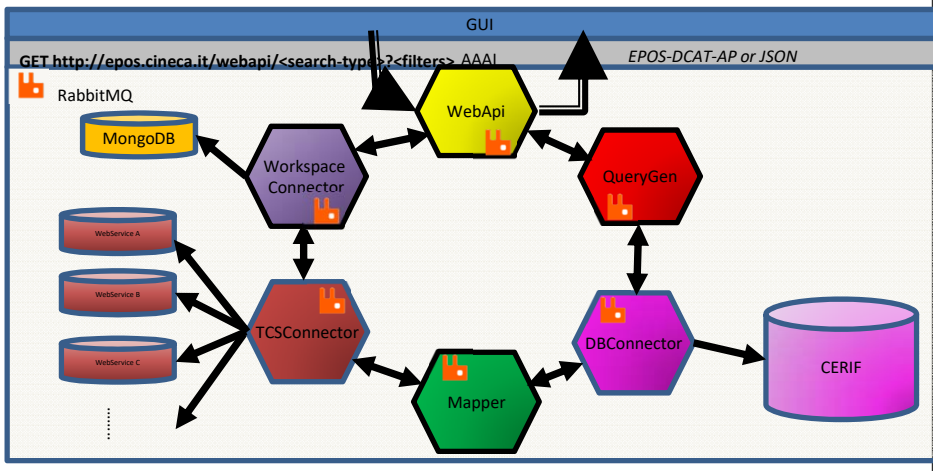
- AT EACH STEP Orchestrator receives and sends messages
- according to pre-defined message routes
- Each use case is associated to a message route
- *Complexity*: *O(2*n)*, where *n* is number of components in the route

*RoundRobin* Configuration

- Orchestrator doesn't exist as central component
- Each component "X" dispatch messages to "Y" according to pre defined routing tables
- *Complexity*: *O(n)*, where *n* is number of components in the route




*Centralised*

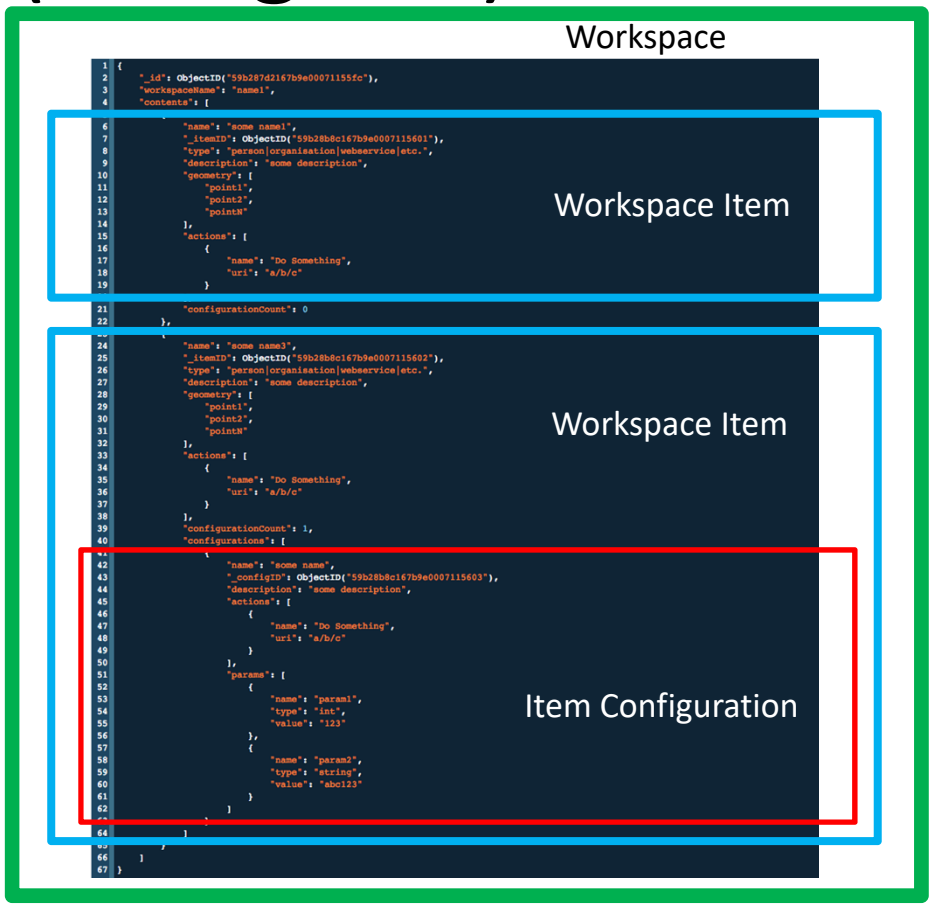



*Roundrobin*

# Workspace (MongoDB)

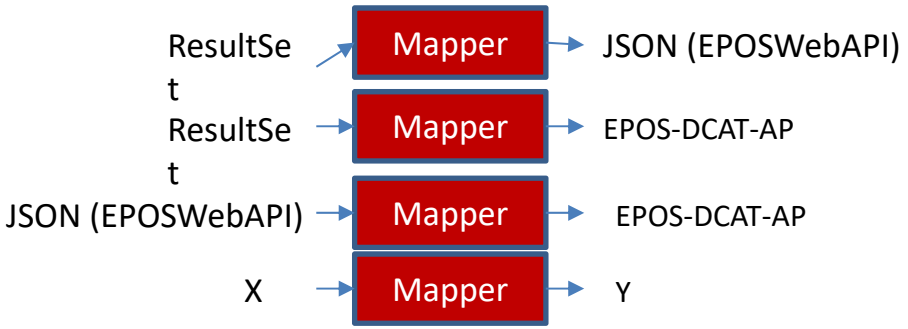
MongoDB:

- Quicker to implement (WEB APIs have JSON output)
- noSQL database
- each workspace is a JSON object
- Stores user items within the workspace



# Mapper

*Converts format "X" to format "Y"*




**ResultSet**


```
"ResultSet_webservice_with_param":[
    {
      "temporalenddate":"2005-01-08 00:00:00",
      "temporalstartdate":"2005-01-08 00:00:00",
      "spatialrepresentation":"Vector",
      "elevation_max":null,
      "elevation_min":null,
      "longitude_max":"16",
      "longitude_min":"14",
      "latitude_max":"39",
      "latitude_min":"42",
      "publisherid":"PIC:999976590",
      "contactid":"https://orcid.org/0000-0002-4553-2380",
      "documentation":null,
      "version":null,
      "operation":"TBDc",
      "keyword":"NFO, rocks, seismic waves propagation, seismology",
      "subdomain":"Crustal velocity parameters",
      "domain":"Near Fault Observations",
      "created":"2017-06-08 00:00:00",
      "webserviceid":"seismnet.fisica.unina.it/eposws/vpvs/1/",
      "spatialreferencesystem":"EPSG:4326",
      "publicaccesslimit":"open data",
      "format":"JSON",
      "uri":"http://seismnet.fisica.unina.it/eposws/vpvs/1/query?",
      "accessanduserestriction":"Creative Commons for data, Open Source licences for software",
      "modified":"2017-06-08 00:00:00",
      "published":"2017-06-08 00:00:00",
      "description":"The Irpinia Near-
Fault .\n\t\t\tThis service provides measurements of VP over VS at the seismic stations, recorded for all earthquake
s detected in the area.",
      "title":"VP/VS Web Service for Irpinia Near-Fault Observatory"
    }
  ]
}
```


Mapper

**EPOS-DCAT-AP**


```
<?xml version="1.0" encoding="UTF-8"?> <eposap:Epos xsi:schemaLocation="http://www.epos-
ip.org/terms.html https://raw.githubusercontent.com/epos-eu/EPOS-DCAT-AP/master/schemas/EPOS-
DCAT-AP.xsd" xmlns:schema="http://schema.org/" xmlns:adms="http://www.w3.org/ns/adms#"
xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:cnt="http://www.w3.org/2008/content#"
xmlns:eposap="http://www.epos-ip.org/terms.html"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:vcard="http://www.w3.org/2006/vcard/ns#" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:dct="http://purl.org/dc/terms/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#" xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:http="http://www.w3.org/2006/http#" xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:locn="http://www.w3.org/ns/locn#" xmlns:dcat="http://www.w3.org/ns/dcat#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"> <eposap:WebService> <dct:title>VP/VS Web Service for
Irpinia Near-Fault Observatory</dct:title> <dct:description> The Irpinia Near-Fault Observatory is
deployed in the Southern Apennines of Italy, along the active fault system responsible for the 1980,
November 23, MS 6.9 Campania???Lucania earthquake. It comprises about 30 stations and covers an
area of about 100 ?? 70 km2. Each site is equipped with strong motion accelerometer and a short-period
velocimeter, while a few stations are equipped with broad-band velocimeters. This service provides
measurements of VP over VS at the seismic stations, recorded for all earthquakes detected in the
area.</dct:description> <dct:issued>2017-06-08T00:00:00.000Z</dct:issued> <dct:modified>2017-06-
08T00:00:00.000Z</dct:modified> <dct:license>Creative Commons for data, Open Source licences for
software</dct:license> <foaf:page>
<foaf:primaryTopic>http://seismnet.fisica.unina.it/eposws/vpvs/1/query?</foaf:primaryTopic>
</foaf:page> <dct:format> <dct:MediaTypeOrExtent>JSON</dct:MediaTypeOrExtent> </dct:format>
<dct:rights> <dct:RightsStatement>open data</dct:RightsStatement> </dct:rights>
<dct:conformsTo>EPSG:4326</dct:conformsTo>
<dct:identifier>seismnet.fisica.unina.it/eposws/vpvs/1/</dct:identifier> <dct:created>2017-06-
08T00:00:00.000Z</dct:created> <eposap:domain>Near Fault Observations</eposap:domain>
<eposap:subDomain>Crustal velocity parameters</eposap:subDomain> <dcat:keyword>NFO, rocks,
seismic waves propagation, seismology</dcat:keyword> <eposap:operation>TBDc</eposap:operation>
<dct:hasVersion>null</dct:hasVersion> <eposap:parameter>
<http:paramName>param_mettype</http:paramName> <rdf:label>None</rdf:label>
<dct:type>None</dct:type> <owl:versionInfo>1.0</owl:versionInfo> </eposap:parameter>
.....
```

# System: issues

## Issues before validation:

- Paging
- Free text search (*solr* like | SQL *OR*)
- Other?

## Issues after validation

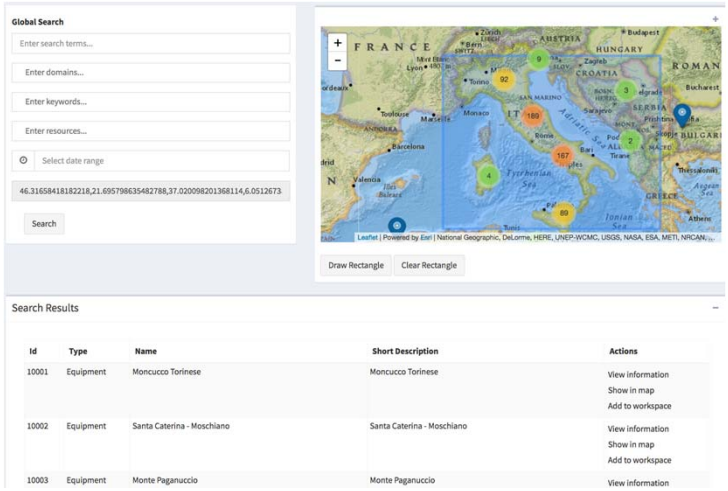
- Roundrobin configuration
- rabbitMQ “as orchestrator”
- Mongo and workspace issue

# WEB APIs

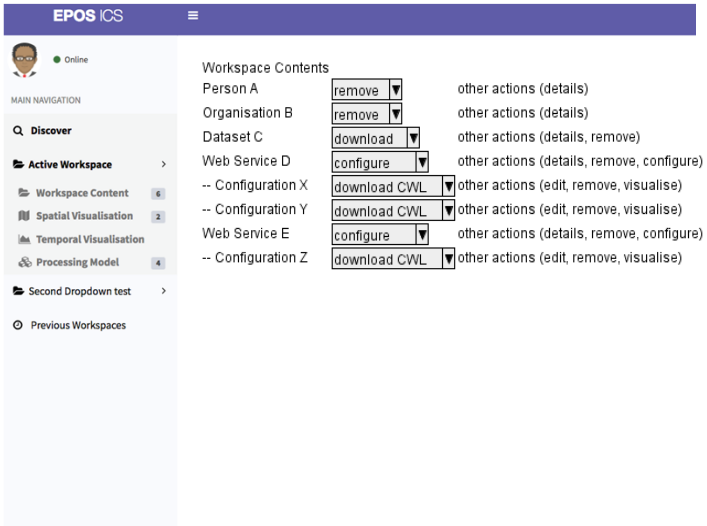
Two step search:

1. Search for catalogues
2. Add catalogues to workspaces and configure them for data selection

SEARCH PAGE




WORKSPACE PAGE



# WEB APIs

**URL http://epos.cineca.it/webapi/swagger-ui.html#/**

| DESCRIPTION | URI |
|---|---|
| *Search boxes autocompletion* | ***GET /domains*** **API** call to get list of domains and subdomains<br>***GET /keywords?param*** API call to get list of keywords<br>***GET /datatypes API*** call to get list of data types |
| *two-step search* | ***GET /search?freetext=value&keywords=value&***…. Generic search: returns a collection of objects, each one containing a minimal set of attributes<br>***GET /getdetails?id=identifier*** Returns EPOS-DCAT-AP XML description of *webservice, person, organization* objects. |

# WEB APIs

OBJECTS TO MANAGE:

1. Workspaces
2. Workspace items (file, notes, webservices .. And other)
3. Configurations (apply to webservices)

| DESCRIPTION | URI |
|---|---|
| Workspace management | ***GET /workspace?id=value&page=value&resultsPerPage=value*** returns the full list of the workspaces if no id is provided otherwise returns the workspace with the defined id value<br>***POST – PUT – DELETE /workspace*** creates/update/delete workspace with a given name/id<br><br>***GET /wsitem?id=value&configurationsDetail=true\|false***<br>returns the content of the workspace item with input id.<br>***POST – PUT – DELETE /wsitem*** creates/update/delete wsitem with a given name/id<br><br>***GET /configuration?id=value*** returns the details of the configuration with input id.<br>***POST – PUT – DELETE /configuration*** creates/update/delete configuration with a given name/id |

# Resources allocation and plans (tomorrow)

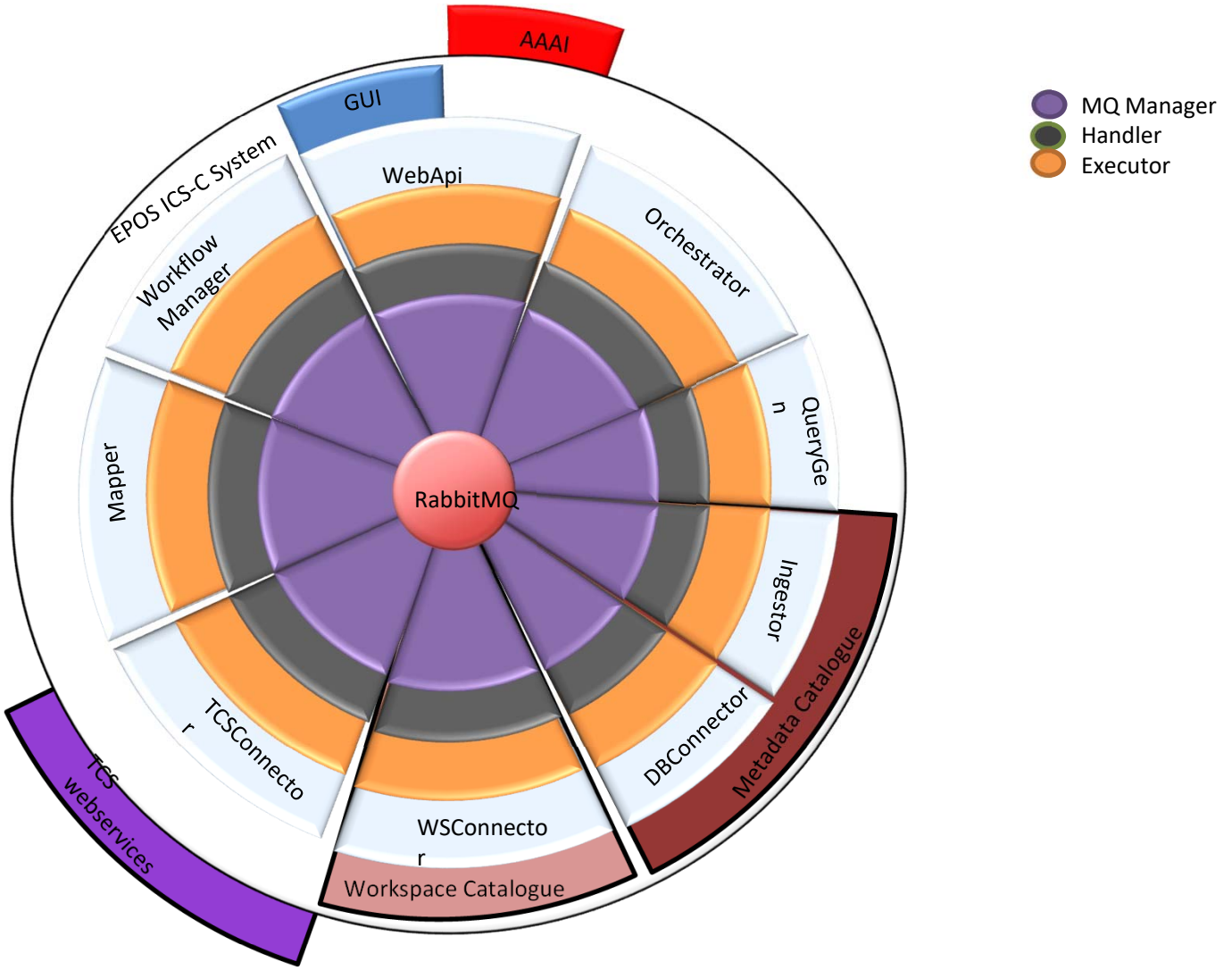
# Groupwork topics

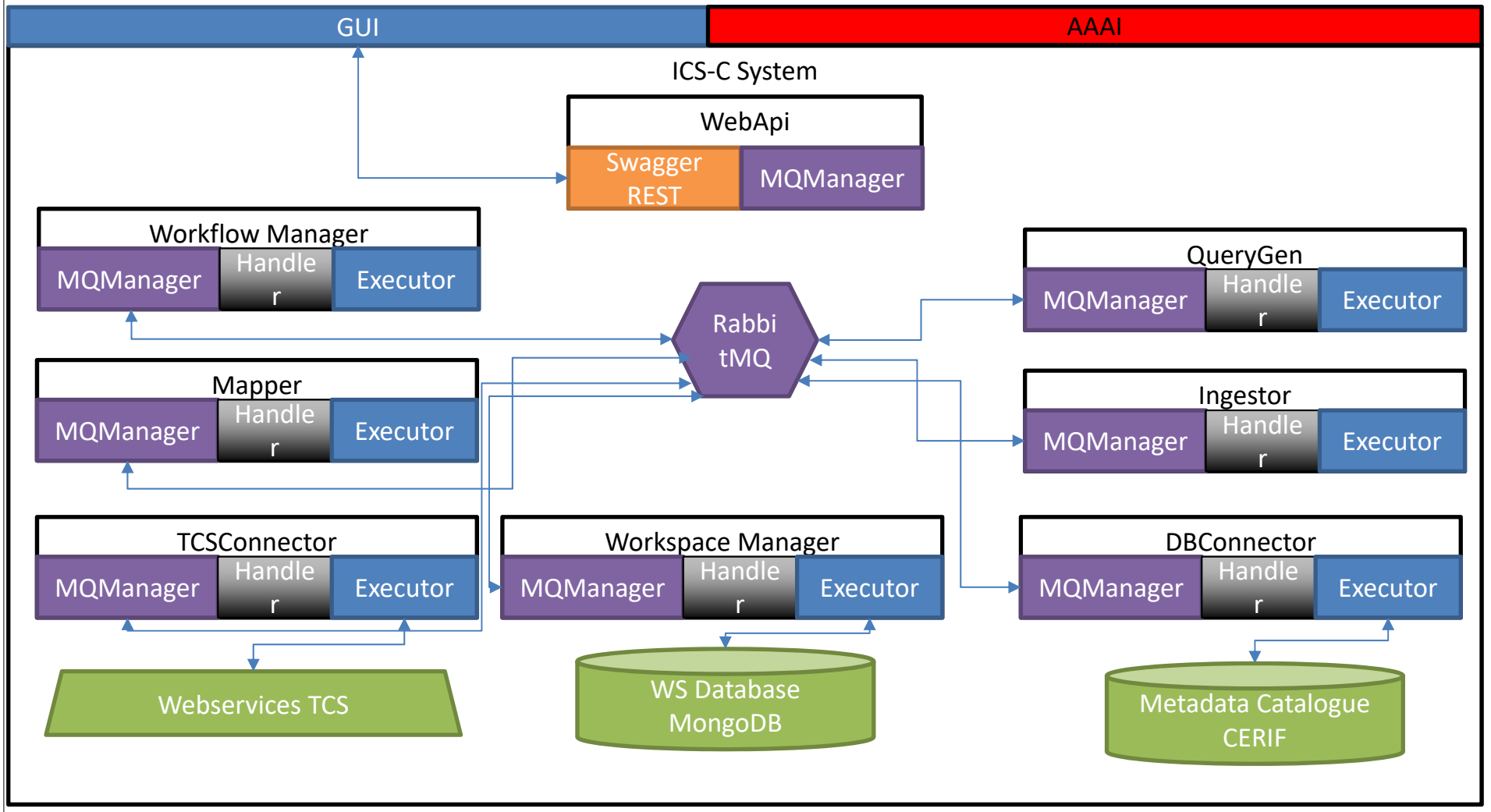
1. User story
2. GUI ←→ Web APIs connection
    1. Operations and Actions: who defines actions?
3. Harvesting vs. brokering
4. Metadata:
    1. Ingestor / harvester / mapper
    2. OGC

# ~~Tomorrow~~ today

- Harvesting vs. brokering
- Manage ocg services (layer name)
- WP14
- WP10
- JSON webapi
- System bugfix
- GUI

# Backup slides


AAAI
GUI
EPOS ICS-C System
WebApi
Orchestrator
Workflow Manager
QueryGen
Mapper
RabbitMQ
Ingestor
TCSConnector
DBConnector
Metadata Catalogue
TCS webservices
WSConnector
Workspace Catalogue
MQ Manager
Handler
Executor


GUI
AAAI
ICS-C System
WebApi
Swagger REST
MQManager
Workflow Manager
MQManager
Handler
Executor
QueryGen
MQManager
Handler
Executor
RabbitMQ
Mapper
MQManager
Handler
Executor
Ingestor
MQManager
Handler
Executor
TCSConnector
MQManager
Handler
Executor
Workspace Manager
MQManager
Handler
Executor
DBConnector
MQManager
Handler
Executor
Webservices TCS
WS Database MongoDB
Metadata Catalogue CERIF