**D9.4 Report on Use Cases, Requirements, Metadata and Interoperability of WP09**

**Document information Summary**

|  |  |
| --- | --- |
| **Date** | 29.03.2018 |
| **Document title:** | Report on Use Cases, Requirements, Metadata and Interoperability of WP 9 |
| **Leader Partner** | Lauro Chiaraluce |
| **Main Author(s):** | Semih Ergintav, Ivano Carluccio, Gaetano Festa |
| **Contributing author(s):** | All Partners |
| **Reviewer(s):** |  |
| **Approved by:** |  |
| **Target audiences:** |  |
| **Keywords:** | DDSS, Use case, DMP |
| **Deliverable nature:** |  |
| **Dissemination level:** |  |
| **Delivery date:** |  |
| **Version:** | 3.0 |

**TABLE OF CONTENTS**

1. Introduction 3

2. Priority List of DDSS 5

3. TCS roadmap 8

4. Data Management Plan (DMP) 10

5. Use cases 11

Specific Use Case 13

Cross-disciplinary Use Case 15

Annex 1 – The Summarized Version of NFO DDSS List 17

# 1. Introduction

Near Fault Observatories (NFOs) are advanced multidisciplinary research infrastructures, based on multi-parametric networks, that monitor the chemical and physical processes in and around active faults, related with earthquake generation. These networks complement regional infrastructures with high-density distributions of seismic, geodetic, geochemical and geophysical sensors located in the near-fault distance (within 10-20 km) around the structures where large earthquakes are expected in the future. In conclusion, NFOs have related laboratories where learning best practices in sensor design and installation, in multi-disciplinary data analysis, and in advanced methods for real-time data processing.

Generally, three of the NFOs operate near plate boundary systems at South Iceland Seismic Zone, the Marmara Sea and the Corinth Rift. Two of them, Marmara Sea and Corinth, include offshore seismic sources that pose an additional tsunami hazard. While CRL and MarSite are international networks (Table 1), others are organized as national/local networks. Table 1 lists the names of the NFOs and their reference persons, to contact.

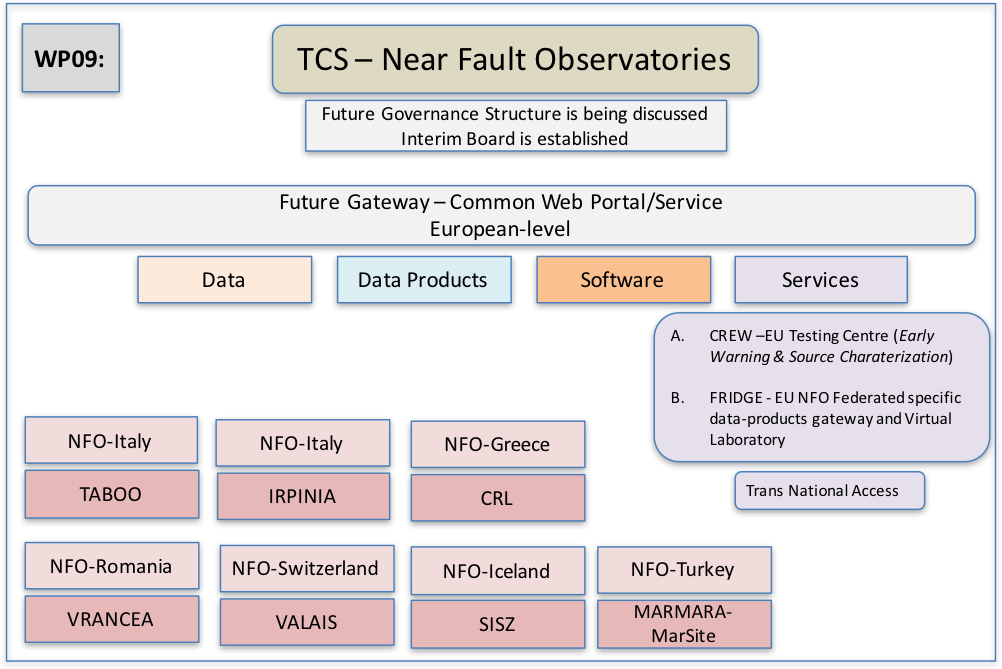
The organization chart (Figure 1) shows the data and service providers contributing to this TCS, providing a variety of DDSSs.

This document reports the status of TCS-ICS integration of TCS WP9 and prepared by the contributions of all seven NFOs within WP9 (Table 1). 44 DDSS (Data, Data Product, Service and Software) elements have been proposed. All elements are related with continuous acquisition and archives of long time-series of multidisciplinary data and data products. To generate a distributed database of multidisciplinary and high resolution data and data products which are accessible through a common TCS gateway (FRIDGE), metadata sets of DDSS elements and required services are also defined in the new services of TCS WP9 or addressed to another well-defined thematic WP(s).

Finally, we described one internal and one cross-disciplinary use case, to demonstrate the availability of data and data extraction procedures, which will be the future basis for the common gateway FRIDGE. These use cases were analyzed in TCS-ICS meeting, Prague (2017).

|  |  |  |  |
| --- | --- | --- | --- |
| *RI Name* | *Location* | *Contact Person* | *DDSS provider(s)* |
|  |  |  |  |
| **TABOO** | Italy | L. Chiaraluce | INGV |
| **IRPINIA** | Italy | G. Festa | UNINA |
| **CRL** | Greece | P. Bernard | CNRS/ NOA |
| **MARSITE** | Turkey | S. Ergintav | KOERI |
| **VALAIS** | Switzerland | J. Clinton | ETH |
| **SISZ** | Iceland | K. Vogfjord | IMO |
| **VRANCEA** | Romania | A. Marmureanu | INFP |

Table 1. NFOs across Europe in WP9, contact people and associated DDSS main providers. For CRL, DDSS are provided both by French (CNRS) and Greek (NOA and Greek Universities) Institutions. DDSS of MarSite are provided by KOERI (Boğaziçi University, Turkey), CNRS (France), Kocaeli University (KU, Turkey), Istanbul University (IU, Turkey), TUBITAK (Turkey), INGV (Italy) and GFZ (Germany).



*Figure 1. NFO - TCS organization*

# 2. Priority List of DDSS

The DDSS Master Table of the WP9 has been reviewed with WP6-7 in Bucharest (Romania), June 2016 and in Naples (Italy), November 2016 meetings. Then it's been divided into two groups, based on their priorities. High priority group contains 24 DDSS elements (12 DDSS are Seismological Data or Data products, 5 DDSS are Geodetic Data or Data products, 3 DDSS are Geochemical Data, 1 DDSS is Strain Data, and 3 are Services) indicating the effective multidisciplinary within each NFO (Figure 2). Non-high priority group contains 20 DDSS elements from different disciplines (Figure 3). Almost all data and data products concern more than one NFO depending on the specific National or International RIs. Each DDSS linked to the corresponding harmonization group. Most of the elements from Seismology, Geodesy and Satellite are or will be exposed through the existing services from the authoritative WP (e.g. EIDA: WP08). Specific data, high-level data and data-products that will be produced by the NFO(s), will be exposed through the TCS data gateway.



*Figure 2. The multidisciplinary character of WP09 in high priority group.*



*Figure 3. The multidisciplinary character of WP09 in non-high priority group*

The harmonization of the available formats and metadata of data and data products was the first task in different WP(s). Then, following the multidisciplinary nature of NFO(s), WP9 tried to contribute to the 12 Harmonization Groups (HG). The following list shows the contact persons in each HG(s):

* Seismology (HG-01/03): Raffaele Di Stefano
* Geology and 3D/4D structural Models (HG-05/10): Semih Ergintav
* Geodetic and Satellite (HG-02/03): Semih Ergintav
* Geohazards (HG-07): John Clinton
* Geochemical data and Meteo parameters (HG-09/15): Antonio Caracausi
* Borehole data (HG-12): Pascal Bernard
* Episodes and Georesources (HG-06/08): Lauro Chiaraluce

In the IC-TCS meetings, many discussions have been realized between WP(s) and HG(s) and we got important feedbacks from HG-02, HG-09 and HG-015. Then, we reconstructed (renamed, concatenated moved to non-HIG list and deleted) some DDSS element, related with HG(s) discussions, in high priority list.

According to the WP9 meeting in Bucharest (2016), the WP will offer open access to the data and data products. Embargoed data will be opened only after the embargo period and they will be distributed as non–embargoed data. Hence, we do not plan different rights, conditions and licenses for access to EPOS legal entities, non-commercial entities and others. However, authorization is required, mainly, to assure traceability. We generally agree with the AAAI services and Open Access Data Policy for user’s registry and secure data access. In this framework, we support the policy of identified users.

Another effort was to identify the metadata standards of the NFO specific data and data products and to create a service for them. For this purpose, FRIDGE (EU NFO Federated specific data and products gateway and Virtual Laboratory), as a NFO service, is under development. This is the common gateway to all NFO(s) in order to discover and download the NFO specific data and high level data products. It will also host simple visualization tools for multidisciplinary data. Specific data and data products stored in local DBs, available for data mining, interaction between the codes and disciplines. Figure 4 shows a demonstration, using multiparameter/multidisciplinary data sets.



*Figure 4. Interaction of the multidisciplinary data, to create high-level data sets.*

IT team of WP9, now, creates a common DB(s) standards for NFO specific data and data products, merging the available structures in NFO(s). For example, TABOO has useful format for geochemical data and data products. Also, all NFO uses QuakeMl. Using the advantage of this approach, IT team demonstrated two new web services related with Vp/Vs and radon time series. Details are given in Use Case section.

CREW (EU Testing Center for Early Warning & Source Characterization) is also defined as another EU level service. It is a testing facility, built on real-time and offline high-resolution data, aimed at fostering the development of next generation methodologies and software for real-time monitoring of faulting processes. The first version of CREW is operational in real-time mode for Early Warning and uses the data from a seismic network, provide alerts and evaluate the performances (Figure 5). In CREW 2.0 (under implementation) we expect that the system will also work on offline (certified) datasets in playback mode, whose earthquake parameters can be downloaded by the user. The user can perform an offline analysis and use the testing centre uploading results and comparing performances of his software against standard software.



*Figure 5. Snapshots from CREW web interface*

Finally, the current DDSS priority list for WP9 was organized. It contains the available information about metadata standard, data format, the service names, DDSS categories in ICS, name of related Harmonization group. The summarized version of the DDSS Master Table of the WP9 is given in Annex-1.

# 3. TCS roadmap

To organize a more restricted list of DDSS elements, that needed to be implemented during the first 24 months of the EPOS-IP project, we decided to work on the standardization and provision of DDSS elements, signed as having HIGH degree of priority (Figure 1). After that, we will focus on the implementation of DDSS elements having Medium and Low priority to expand the number of the products (Figure 2). In February (2017), our list merged with Master DDSS list under the control of ICS (Annex-1).

As can be seen in Figure 3, non-high list mainly consists of data products, since they are complex elements and need integration with other WP(s) in HG(s). Most of the elements of Seismology, Geodesy and Satellite are well defined using the existing formats and services in the authoritative WP(s), with the inputs of HG(s). We push the HG(s) to estimate the roadmaps and timelines in our DDSS list. Until now, using the feedbacks, we simplified our roadmap and removed risky elements. Unfortunately, this process is very slow. To eliminate any irrecoverable mistakes, we decided to define some formats independently.

The TCS strategy for **USER ACCOUNTABILITY and REGISTRY** is currently **under discussion**. We are discussing the possibility of using EUDAT B2ACCESS platform for FRIDGE as an alternative of inheriting completely ICS strategy.

Table 2. Present status of NFO DDSS list,

|  |  |  |
| --- | --- | --- |
| Seismological data | Vel. Seismic waveforms Continuous (EIDA - Virtual Network) | high - done |
| Seismological data | Seismic Station Information (EIDA - Virtual Network) | high - done |
| Seismological data | Acc. Seismic waveforms Continuous  (EIDA - Virtual Network) | high - done |
| Seismological data | VP/VS (Time series) | high - done |
| Geochemical data | Radon Decay Counts (with Local temperature in radon sites as correction parameter) | high - done |
| Geochemical data | Meteo parameters | high |
| Geochemical data | CO2 soil flux measurements | high - done |
| Seismological data | Historical EQ Catalogue | high |
| Data Services | CREW - EU Testing Center Early Warning & Source Characterization | high - done |
| Data Services | FRIDGE - EU NFO Federated specific data-products gateway and Virtual Laboratory | medium |
| Seismological data | OBS waveforms | medium |
| Seismological data | Infrasound waveforms | medium |
| Seismological data | Strainmeter (water level, pressure) | medium |
| Seismological data | Earthquakes Catalog | medium |
| Seismological data | Earthquakes parameters (location, mag, phases, moment tensor) | medium |
| Seismological data | Focal mechanisms | medium |
| Seismological data | Seismic Velocity and Attenuation Models (1D, 2D, 3D, 4D) | medium |
| Seismological data | Site Characterization | medium |
| TNA | Transnational Access | Decision of EPOS community is waiting |

# 4. Data Management Plan (DMP)

Generally, at the TCS level, we will propose a DMP to demonstrate our long-term sustainability during the operational phase. Now, individual DDSS elements discuss their own DMP, focusing on the following topics:

* Data storage and maintenance responsibility
* Data curation responsibility
* Data management and governance structure
* Financial commitment securing the operational costs

As a first step, we plan to make data findable, including provisions for metadata. For EIDA data, metadata can be explored using fdsnws/station. DOI are assigned, at the level of the network operating the stations. We grouped all seismic stations within EIDA under Virtual networks related to single NFOs, which are now operational. Currently there are no plans to include searchable keywords, and version numbers would need to be introduced for data, metadata and products. Data produced by NFOs are open and they are intended to be interoperable, that is allowing data exchange and re-use between researchers, institutions, organizations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available open software applications, and in particular facilitating re-combinations with different datasets from different origins). This is already realized via EIDA, and FRIDGE and Geodetic data are heading towards this goal. Data and metadata vocabularies, standards or methodologies are followed in EIDA (Seed, StationXML, QuakeML) to make the data interoperable. These are being developed for FRIDGE.

Licenses continue to be explored for NFO, and NFO will adopt appropriate creative commons licenses for each dataset and products. EIDA, a mature service, is developing data quality assurance processes. FRIDGE needs to develop these in future.

Currently, NFOs are left to their own planning to manage data security / recovery / transfer, which is not acceptable in the long run. This matches current rules for EIDA nodes. EIDA needs to explore certified repositories in order to ensure long term preservation and curation through standard methods. FRIDGE will need to do the same.

# 5. Use cases

Using the advantages of multi-disciplinary data sets, we can utilize many simple use cases. Table 3 includes general categories of these examples.

Some of them have been already implemented/tested during IT meeting (the first one in Rome, 2017). We first focused on VP/VS and Radon time series because they may represent a sort of prototype respectively for complex and standard time series generation. In addition to this, for the geochemical data (e.g. Radon), we had no international standard for the unit measurement, sample rate and format, thus it was the opportunity to work also within the Harmonization Group (mainly with the Volcano Observatories community), to define a complete metadata for such raw data. Once defined all these parameters, the WP09 IT group prepared the API/web services that together with the keywords and metadata, were moved to the ICS-C dictionary and portal.

All the two web services we implemented are also going to be part of FRIDGE.

Generally, to create these use cases, we solved complex problems that are related with

* the scientific concept
* the query building
* the common DB structure
* the query performance
* the input parameters standard definitions and defaults
* the output (data and metadata) format standard

After this exercise, IT group of WP9 produced the following products:

1. a text document (API description) organized into sections describing
   * the concept
   * the input parameters and related standards priorities and defaults
   * the DB query itself
   * the payload (output) with related standards (this last part is in progress and almost finished)
2. a "portable" package including
   * the working server
   * the needed DBs (with example data)
   * the API (in python)

The minimum purpose of the “portable” package (in particular a “docker”, working on Docker Platform) is to demonstrate the two-working web-services at any place. As a first example, our IT group used it in TCS section of ICS-TCS meeting (Prague, 2017).

**Table 3** Example Use Cases,

|  |  |
| --- | --- |
|  | **Use Cases** |
| **Time Series** | X Chemical Component (Rd, CO2, T) in time |
| Geodetic Displacement (all comp) |
| Number of earthquakes per day |
| *b*-value in time |
| VP/VS, VP, VS, Attenuation in time at site |
| Heat Flux in time |
|  |  |
| **Maps with surfaces** | X Chemical component distribution |
| Map Layers from 3D models (VP/VS, VP, VS, Poisson ecc) |
| Faults surfaces |
| Geo Layers top and bottom surfaces |
| Seismological Discontinuity surfaces |
| Heat Flux |
| Vertical Sections through 3D models |
|  |  |
| **Positional maps** | Interactive Maps of selected sites position |
| Interactive Maps of selected earthquakes position |
| Space-Time earthquakes distribution |
|  |  |
| **Histograms and Dispersion** | Selected earthquakes parameters errors distribution |
| Selected earthquakes' depth distribution |
| depth errors vs Origin-time errors |
| Dromocronas (Distance vs P and S Traveltime) |
|  |  |
| **VLAB** | VP/VS, VP, VS, NumEqks/Day, CO2 flux, Radon Flux, Geodetic Displacements ALL at site |

In the following part, we specified one internal and one cross-disciplinary use case in detail.

## Specific Use Case

|  |
| --- |
| **Use case name/topic**: *Selecting and viewing earthquakes’ distribution in maps and vertical sections* |
| **Use case domain** This use case is *related to one single discipline: seismology. The goal is to allow the user to select earthquakes locations from the specific NFO-DB based on some spatial-temporal-quality criteria and to plot the distribution in a geographic map and along a vertical section which extremes are interactively selected on the map. The aim is to help understanding the fault system geometry and spatial-temporal trend of the seismicity pattern.* |
| **Use case description**: *As a seismologist, I want to observe the spatial and temporal distribution of earthquakes sources in the fault system volume in different 2D views. I want to be able to chose:*   * *the type of location to use for plots (obtained with different location codes and velocity models)* * *the quality of such locations (rms, gap, h\_err, v\_err, nP, nS and so on)* * *the time window* * *the volume (geographic area + depth range)* * *the magnitude range* * *the magnitude threshold to use different symbol and show focal mechanisms if available*   *I also want to use a color palette to identify the time range of plotted locations*  *As soon as the map is ready, I also want to be able to draw an oriented rectangle on the interactive map to produce a vertical section with locations, as wide as the rectangle short side and as long the rectangle long side and as deep as the deepest earthquake.* |
| **Actors involved in the use case**   * *User Seismologist* * Seismology researcher |
| **Priority:** *High* |
| **Pre-conditions**: *User must have logged in* |
| **Flow of events – user view**  The following steps are need to answer the question:   1. *seismologist user chooses*    1. *the specific NFO [the use case is focused on just one NFO]*    2. *a lapse of time*    3. *a geographic region inside the NFO area (or the whole)* 2. *seismologist user chooses*     1. *criteria to select earthquakes’ locations by*        1. *type of location*       2. *quality of location*       3. *magnitude range*    2. *magnitude threshold to show strongest earthquakes with different symbol and eventually attribute a focal mechanism if available*    3. *weather to use a color palette for time*    4. *symbols to use for earthquakes beyond magnitude threshold* 3. *seismologist user produces a geographical map with gray shaded topography, seismic stations distribution, earthquakes’ locations plotted as dots colored as a function of time and strongest earthquakes plotted with the same color code but with different symbol and focal mechanism associated if available* 4. *as soon as the map is ready the seismologist user*    1. *draws an oriented rectangle on the map*    2. *obtains a new Distance/Depth (XY Cartesian) plot with all the locations included within the rectangle from depth 0km to the deepest earthquake (scaled plot) with the same color and symbol code as the map and focal mechanisms if available.* 5. *The locations, magnitudes, focal mechanisms, related locations’ quality parameters are restituted to the user in a quakeML catalog with all metadata included.* |
| **System workflow - system view**   1. *The user interface receives the input parameters for query\_locations* 2. *It activates one tasks (query\_locations) connecting to the specific NFO database* 3. *The task connects to the database and performs a query typically combining information from locations, quality, magnitudes, and focal mechanism tables* 4. *The task a temporary VIEW (see mySQL) in the DB storing the results from the combined query (one line per location)* 5. *The data contained in the VIEW are also stored in a downloadable quakeML catalog file* 6. *A listening service takes data from the view and produces the interactive map* 7. *The interactive map allows to draw the rectangle based on which extremes a sub\_task is activated selecting earthquakes from the view based on polygonal in/out function and locations are plotted in a vertical section* 8. *A button to reset the vertical section plot and redraw the rectangle is made available* |
| **Post-conditions** *The View is kept in the DB until the user session is active or until the session timeout is passed* |
| **Extension Points** If the use case has extension points, list them here.  *No extension points.* |
| **« Used » Use Cases** If the Use Case uses other Use Cases, list them here.  *No other use cases.* |
| **Other Requirements** This can include non-functional requirements related to the Use Case.  *Privacy legislation, response time of the system* |
| **(to be filled in by WP7) Class diagram and sequence diagram** |

## Cross-disciplinary Use Case

|  |
| --- |
| **Use case name/topic**: *Viewing and comparing Vp/Vs ratio to Radon concentration in time* |
| **Use case domain** This use case is *Cross-disciplinary. The goal is the comparison of  two “entities” related to seismological (Vp/Vs time series) and geochemical (Rn concentration time series) disciplines. The scientific reason for this is looking for changes (in space and time) related to the deformation process (e.g. fracturing and/or fluid migration processes) occurring during the pre- co- or post-seismic phase.* |
| **Use case description**: *As a seismologist, I want to observe and compare the spatial and temporal behavior of P- and S-wave velocity ratio (referable to the rock volume elastic parameters) with the temporal and spatial pattern in Rn concentration in a defined rock volume (referable to on-going deformation processes), looking for statistically coherent change points, thus possibly ascribable to the same undergoing physical process such as for example the earthquake preparatory phase.* |
| **Actors involved in the use case**   * *User Seismologist* * *Geochemistry researcher* * Seismology researcher |
| **Priority:** *Medium* |
| **Pre-conditions**: *User must have logged in* |
| **Flow of events – user view**  The following steps are needed to answer the question:   1. *seismologist user chooses*    1. *the specific NFO [the use case is focused on just one NFO]*    2. *a lapse of time* 2. *seismologist user chooses*     1. *one specific seismic station*    2. *criteria to select seismic rays by earthquakes quality and takeoff angle*    3. *criteria to generate a moving mean (boundary and default criteria are suggested by the Seismology researcher*s group)    4. *a magnitude threshold to show strongest earthquakes on top of the two time-series* 3. *seismologist user chooses*    1. *one specific site from suggested neighboring Radon sites active in the same period or part of it, based on the step 2*    2. *criteria to correct for meteorological data and to produce a moving mean (boundary and default criteria are suggested by the Geochemistry researcher*s group) 4. *seismologist user produces a combined X (time) Y (Vp/Vs ratio) and Y’ (Rn concentration) dispersion plot with symbols and strokes*   Alternative sequences and needed steps (user view)   1. *seismologist user may go for step 3 (Geochemistry) before step 2 (Seismology). Step 1 is forced to be the beginning.* |
| **System workflow - system view**   1. *The user interface receives the input parameters for query\_VpVs and query Radon* 2. *It activates two tasks, one per CPU, connecting to the specific NFO database* 3. *Each task connects to the database and searches, for the chosen recording site, records that match to the required criteria: SQL queries might be complex or simple queries, depending on the DB structure, operating on the basic tables containing*    1. *Vp/Vs: P arrival times (and related quality parameters), S arrival times (and related quality parameters), earthquakes locations, quality of earthquakes locations, takeoff angles, back-azimuth angles*    2. *Radon concentration: Rn counts, site correction parameters, meteorological site parameters* 4. *Each task restitutes a file on disk in a standardized format for time series (to be defined)* 5. *A listening service takes the files as soon as they are ready and produces an interactive plot where only scales, symbols and colors can be changed* 6. *A button to “Change criteria” sending to the selection page from step 2 is made available (see post-conditions)* |
| **Post-conditions** *The request is kept in memory to be reloaded as default in step 6 of the System workflow* |
| **Extension Points** If the use case has extension points, list them here.  *No extension points.* |
| **« Used » Use Cases** If the Use Case uses other Use Cases, list them here.  *No other use cases.* |
| **Other Requirements** This can include non-functional requirements related to the Use Case.  *Privacy legislation, response time of the system* |
| **(to be filled in by WP7) Class diagram and sequence diagram** |

# 

# Annex 1 – The Summarized Version of NFO DDSS List

