D8.4 Report on Use Cases, Requirements, Metadata and Interoperability of WP 8

Document information Summary

Date	March 27 , 2017
Document title:	Report on Use Cases, Requirements, Metadata Catalogue Templates of WP8
Leader Partner	INGV
Main Author(s):	Alberto Michelini
Contributing author(s):	WP8 TCS-ICS liaisons
Reviewer(s):	
Approved by:	
Target audiences:	WP6 and WP7
Keywords:	Data, web services, use cases
Deliverable nature:	Report
Dissemination level:	public
Delivery date:	March 30, 2017
Version:	v. 0.1

Table of Contents

SUMMARY

This deliverable reports the status of the TCS-ICS integration of WP8 at M18. For this TCS, 46 Data, Data Product, Service, Software (hereafter DDSS) elements have been proposed. Out of these, there are 21 DDSS that are "implemented", 15 "under implementation", 9 in the "design phase" and 1 provided as "available".

We feel that all the services so far developed and operational (i.e., the 21 implemented) require some additional integration between each other that can be obtained only through the relevant *metadata* and the assignment of *persistent identifiers* (PID) capable to index uniquely the products. The final goal being of providing tools that, besides helping the researchers with additional products to obtain analysis of the data reliably and quickly, include provenance procedures that will allow researchers to reconstruct all the individual steps of the workflows that have been adopted. The overall machinery adopted by EPOS ICS & TCS should provide all those ingredients/procedures/tools necessary to perform "reproducible science".

The two proposed use cases both focus on the availability of data and follow similar data extraction procedures. The first use case consists of gathering "time-geographic" selected data & products. This use case, besides being simple, is effectively the starting point of any other more complex use case. The second use case is more sophisticated including the orchestration of data and products from several TCSs.

1. Introduction

This should include a brief description of the TCS and RIs within that TCS, or the individual RI if not within a TCS grouping. The description should include information on the TCS/RI location, contact person(s), major functions/offerings and (if an e-RI) a URL used to access it.

Main target of Task 8.3 of EPOS IP is the close interaction between the TCS-Seismology with WP6-7 to ensure the full integration and the interoperability of the technical developments of the EPOS-Seismology services with the IT infrastructures and general IT services of the EPOS ICS (e.g., AAI, PIDs, metadata structure & discovery services, connection to HPC resources, advanced visualization). In this task, the IT developers of the EPOS Seismology TCS coordinate closely with the activities organized by WP6-7.

The TCS Seismology in EPOS-IP has been conveniently structured into 4 main activities (pillars) – Task 8.4 *Access to seismological waveforms*, Task 8.5 *Access to seismological products*, Task 8.6 *Earthquake hazard and risk data, products and services* and Task 8.7 *Implementing Computational Seismology tools and services*.

The Task 8.4 deals primarily with the services for seismological waveforms and associated data, including the compilation of station metadata and fostering common data archival and sharing policies. All activities are coordinated under the umbrella of ORFEUS and its Service Management Committees. In turn, the activities include:

- Further developments of the software architecture of the European Integrated (seismological) Data Archive

- EIDA, making full use of current FDSN standard web services for station metadata (fdsnws-station) and for waveform data (fdsnws-dataselect).
- The development of a new class of data services for the strong motion community to be included into the European Strong Motion DB.
- EIDA integration of mobile networks and OBS.

The Task 8.5 deals with development, expansion and EPOS-integration of services to access seismological products provided by EMSC, its partner institutions, and the scientific community, building upon the existing infrastructures at EMSC and its partners. This includes:

- The platform that will collect and provide access to new products of scientific interest (authoritative locations and magnitudes, moment-tensor, shaking and damage maps, seismic source models, earthquake location methods, site response corrections).
- The European Archive of Historical Earthquake Data AHEAD (authoritative locations and magnitudes for historical earthquakes, macroseismic intensity data, studies on historical earthquakes).
- The improvement of services for earthquake parameter and earthquake observation information.

The task 8.6 focuses on earthquake hazard and risk data, products and services and will develop the services initiated by EFEHR under NERA and SHARE and achieve their integration within EPOS. Main target is the European earthquake hazard and to this purpose the activities involve the maintenance, revision, and enhancement of collections of seismogenic source models and of earthquake catalogues. To this end, the activities of this task concern the following services::

- European Database of Seismogenic Faults.
- European Ground Motion Prediction Equations database.
- European geotechnical engineering information database and web services.
- Services for the analysis of strong motion recordings in buildings.
- Earthquake risk.

Finally Task 8.7 targets the Implementation of Computational Seismology tools and services. This is motivated by the need for full waveform simulation in complex 3D media to make it possible wavefield simulations at frequencies in the "engineering range".

This all shows that the WP8 EPOS Seismology activities are quite diversified and embrace a very large spectrum of different activities stemming from raw data to sophisticated analysis deriving from several intermediate results as in the case of seismic hazard maps.

The overarching, flexible and easy-to-implement orchestration of the modular services provided by the activities listed above is pivotal to EPOS Seismology and to EPOS ICS.

2. Priority List of DDSS (Data, Data Product, Service, Software)

The first crude version of requirements were collected through a template (ppt file with tables) which were sent out by WP2 (by the WP2 leader Lilli Freda). These are used as a starting point for our interactions. During the last few months WP6 and WP7 teams have conducted a level-1 interview where information regarding this is mentioned. We need to collect now a more restricted list of DDSS elements

that are essential to be implemented during the first 24 months of the EPOS-IP project. Therefore a detailed priority list is requested from each TCS, in this section. Such list should describe each DDSS specifying its technical maturity (standards data format? standard metadata? Web services or APIs available? Is it under construction? Date of release of the full service?). For each DDSS which is going to be implemented within M24 we suggest to describe it by using the table in Annex 1.

The List of all the WP8 DDSS so far proposed is provided in the "WP8_DDSS_list" table that is to be considered the master (dynamic) table available at
https://docs.google.com/spreadsheets/d/1S-0RvAX35F63aJirfAZBwpS6wZ3j78aX-XkxDMm8on8/edit#gid=0 .
In the table there are listed all the 46 DDSS and for each one there are provided information on:

- WP8 Pillar name
- Service name
- DDSS element name
- API
- Type (Data, Data Product, Service, Software)
- Status (implemented, under implementation, design phase)
- Data / Product format (standard)
- Documentation resource (URL)
- Hosting Institution(s)
- Responsible contact person
- Comments
- Priority (high/medium/low)
- Maturity (high/medium/low)
- Date for delivery to ICS (Ready or Year/Month)

Information on the status and developments of the DDSS is progressively updated as it becomes available. Based on this information, the following table summarizes the status of the DDSS and is also adjourned dynamically (March 2017).

Summary of DDSS	
number	**DDSS Status (column F)**
21	Implemented
15	Under implementation
9	Design phase
1	Other
46	Total

In the following, we have followed the same activity scheme provided in the project and summarized in the introduction to group the DDSS. The DDSS *priorities*, the levels of *maturity* and the *readiness* for provision to the ICS (i.e., either Ready or <Year>/<Month> for delivery) are listed in the last three columns of each table. With regard to the priorities, most of the DDSS listed have high priority but not all are implemented or are likely expected to be implemented within the first two years of the implementation phase. We also note that many of the services have been implemented already but they undergo adjournments during EPOS-IP.

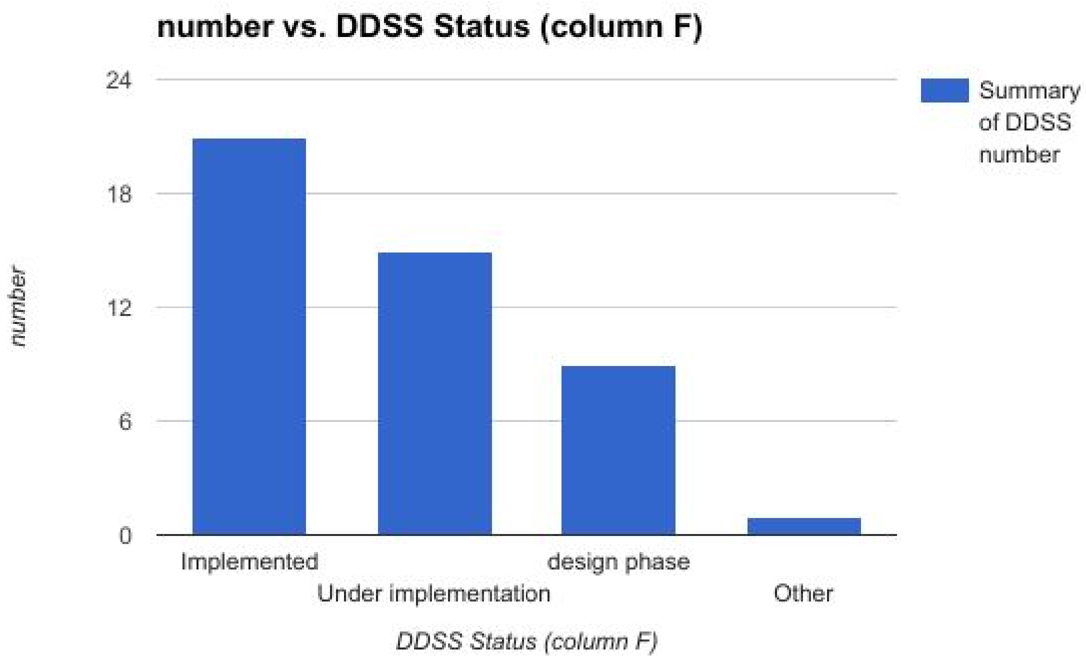


Figure 1. The histogram summarizes the status of development of all the DDSS of EPOS-S.

Access to seismological waveforms

The DDSSs are composed of two main groups – EIDA and RRSM/ESM – of rather mature data and services. The EIDA services incorporate and adhere to standards agreed within the FDSN (International Federation of Digital Seismograph Networks; https://www.fdsn.org) and therefore globally compliant in seismology. Similarly, an effort is made for the time-window waveform event data of the ESM DB. In this case, however, the global strong motion community does not seem to have reached consensus on the waveform services to be developed and for ESM it has been chosen to follow similar developments as those of the FDSN and, by so doing, set some global standards for waveform event DBs such as ESM. Noteworthy, the waveform event data webservices are consistent with those developed at UC Berkeley, Seismological Laboratory. In all cases, we are referring to RESTful web service interfaces for accessing common FDSN data types.

The EIDA next generation services include, in addition to the now standard *fdsnws-dataselect* and *fdsnws-station* to retrieve data and station information, respectively, new data services and products that will allow for an overall modernization of the EIDA suite of services: *EIDA Routing Service (ERS)*, *EIDA Mediator Service (EMS)*, *EIDA Authentication Service (EAS)*; and a new interactive *EIDA web interface*. Specifically, the routing-service will provide routes to data and services distributed across the federation, the AAI system compliant or possibly the same adopted by EPOS will be able to cope with the distributed services whereas the quality metrics service (WFcatalog) and the mediator service will provide the capacity to filter the data according to a defined set of metrics and of metadata.
Finally, the *station book* will facilitate the access to the most complete and adjourned station metadata.

Near real time access to unrevised strong motion data is provided through the RRSM DB which is already in place and operative. The *European Strong-Motion* (ESM) DB (interactive) is operating since 2015 and it will be enriched with the standard *fdsnws-event* for event parameters information and *fdsnws-station* for station information web services. In addition, the *waveform-event* - a newly developed service to access and download windowed waveform event data – is under development and while concluded its prototyping it will become operational by M24.

Finally, there is an ongoing discussion on the services to be provided and how these will interface with EIDA for the *Mobile station pool* and the *OBS coordination teams*.

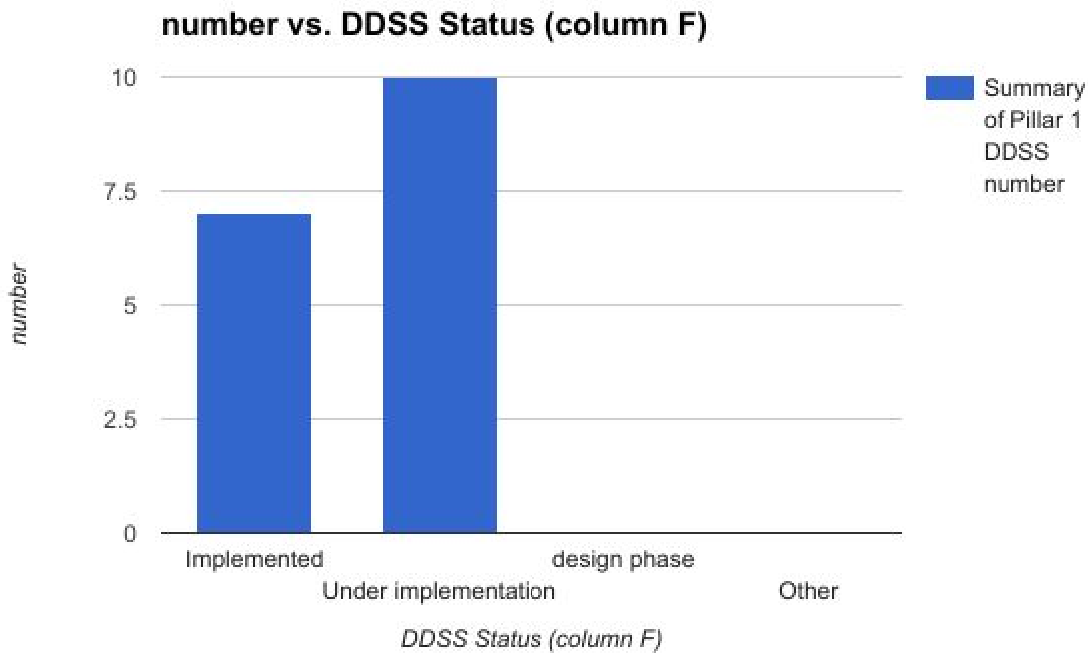


Figure 2. The histogram summarizes the status of development of all the DDSS of Pillar 1 (waveforms).

Seismological Products

These are standard products deriving from routine, often automatic, analysis like earthquake catalogues provided by EMSC and other seismological centers: source parameters, moment tensors and similar. DDSS relevant to *Seismological Products will be* integrated to the seismic portal (http://www.seismicportal.eu/) which is the final host for the services (web portal and its associated web-services). The interactive platform resides on backend databases for hosting and web-services to facilitate the access to the existing datasets and products. Both the databases and web-services are built upon open-source standards.

Two main data products for earthquake selection include the *fdsnws-event* (http://www.seismicportal.eu/fdsn-wsevent.html) for the selection of earthquakes according to many different parameters like geographic location, depth, magnitude,... and the *FE-region lookup* (http://www.seismicportal.eu/feregions.html, Flinn-Engdahl Regions are a seismic and geographical regionalization scheme proposed in 1965, defined in 1974 and revised in 1995) have been provided since long by EMSC and they are currently implemented as web-services and are running on the seismic portal.

In addition, "*Moment Tensors*" (http://vigogne.emsc-csem.org/mtws/) and "*Testimonies*" web-services (http://vigogne.emsc-csem.org/mtws/) are implemented and are currently in testing phase. These are standard products deriving from routine, often automatic, analysis like earthquake catalogues provided by EMSC and other seismological centers.

The "*SRCMOD*" web-service will be useful to retrieve the finite fault rupture models of large earthquakes and the "*UNID*" web-service is designed to identify the origin parameters of the EMSC catalogue from a unified identifier and vice versa. Both services are currently under implementation.

Overall, EMSC is developing a platform upon which to orchestrate these different services with the target of creating a federated environment similar to that developed for the waveform data through EIDA.

Another important DDSS part of this pillar is AHEAD (Archive of Historical Earthquake Data) that provides parameters (e.g. event identifiers, locations and magnitudes) of past earthquakes, and macroseismic intensity data. Parameters are provided via the web portal, using an fdsnws-event service (preferred and alternative solutions; formats: QuakeML 1.2, CSV, GeoJSON), or using an OGC WFS (only preferred solutions; format: GML) service. Macroseismic intensity data can be accessed - with some limitations due to still-to-be-solved legal issues - via the web portal or via a dedicated web service (preferred and alternative solutions; formats: QuakeML 2.0, CSV, GeoJSON) still in an early experimental phase. As the underlying data comes from a variety of sources such as published papers, earthquake catalogues, or more generally, grey literature, bibliographical metadata has a key role in AHEAD, and a dedicated experimental web service is available (formats: DC elements OAI XML, DC terms RDF, BibTeX).

Finally, the AVIOS service (http://avios.geopsy.org/) is an interactive, web site procedure for automatic site characterization processing. The GUI features and core modules of the Geopsy.org software - which AVIOS is built on - has been completed. The working prototype of the proposed AVIOS web service will be opened to WP8 community selected users for testing and validation by may 2017.

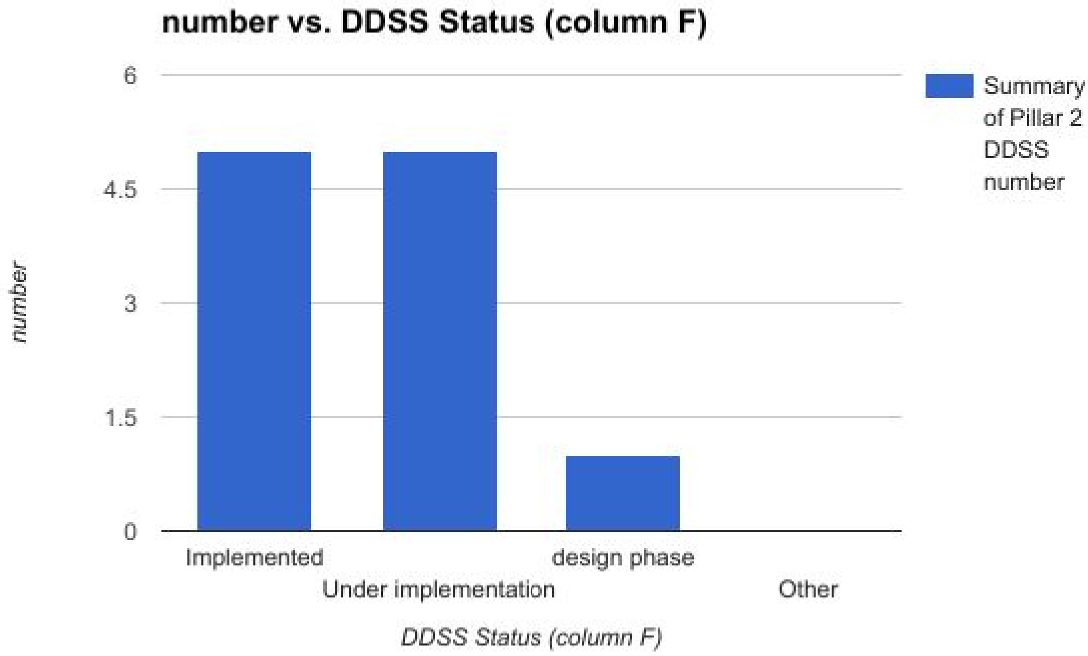


Figure 3. The histogram summarizes the status of development of all the DDSS of Pillar 2 (Seismological products).

Hazard and Risk

Services

DDSS relevant to seismic hazard and risk are available and integrated at EFEHR (www.efehr.org). The main EFEHR service is the web portal and its associated web-services. The interactive platform resides on backend databases for hosting and web-services to facilitate the access to the existing datasets and products. Both the databases and web-services are built upon open-source software and open standards. The databases adopts the postgreSQL RDBMS and its tools, whereas the web-services are developed based on in-house technologies such as Java Servlet and Java WebSocket inherited from Apache Tomcat® packages. The front-end is a web-interface implementing Jetspeed technologies fully conformant to the Java Portlet standards.

The access of data and products is currently open without any authentication service. EFEHR follows, as far as the data providing institutions and projects allow, an open data policy. The underlying data have Creative Commons® licenses to ensure that the content of the distributed datasets can be reused, abstracted, modified material to produce new information. The geo-spatial content, either hazard results (maps, curves, and uniform hazard spectra) or raw input models (active faults, seismic source zones, and seismic catalog) is implemented by the OpenGIS® Web Map Service Interface Standard (WMS). For more advanced users, a documented APIs is also provided to support the building of their own applications directly on top of EFEHR data services. For technical documentation and (automatic) generation of service clients, we provide service documentation in WADL (web application description language) format, i.e. WADL xml, machine readable and WADL html extraction, human readable.

An immediate priority of EFEHR is to populate the databases and to disseminate the exposure data (European building stock inventory) and vulnerability information. Of interest is to integrate the outcomes of NERA project, the European Building Stock Database (NA7 deliverable). Distribution of the exposure data will open even more the audience of EFEHR to engineers and industry. Earthquake hazard datasets and products combined with the exposure datasets will constitute the necessary input for hazard and risk assessment.

Additional activities foreseen within EFEHR are related to updates of different components of the regional hazard model, development of tools to develop and validate input models to support the update of the European seismic hazard model. These tools are to be developed on open-standards and openly distributed through EFEHR. Note that currently EFEHR does not provide access to any risk services or results. However, the risk services and products might be extended from the existing hazard services when they become available.

Products

Hazard products as distributed by EFEHR are:

- Input models and data: Fully harmonized datasets: fault sources, earthquake catalogues, area source model, strong motion data and earthquake rate forecasts.
- Output: Extensive seismic hazard products, namely hazard curves, uniform hazard spectra and maps for Europe and national assessments; As of today, EFEHR displays the output of:
    - 2013 Reference European Seismic Hazard Model (ESHM13) as developed within SHARE project.
    - 2001 Global Seismic Hazard Assessment Program (GSHAP) outcomes.
    - 2015 updates of the Swiss Seismic Hazard.
- Other products:
    - Cross-links to software platform that facilitate on-demand computation of regional and site specific seismic hazard, risk and economical impact.
    - Standardized input files based on the “Natural hazards' Risk Markup Language” (NRML) that links to the OpenQuake implementation of the ESHM13.
    - European Geotechnical Database.
    - Relevant documentation of input datasets, hazard models and web-services.

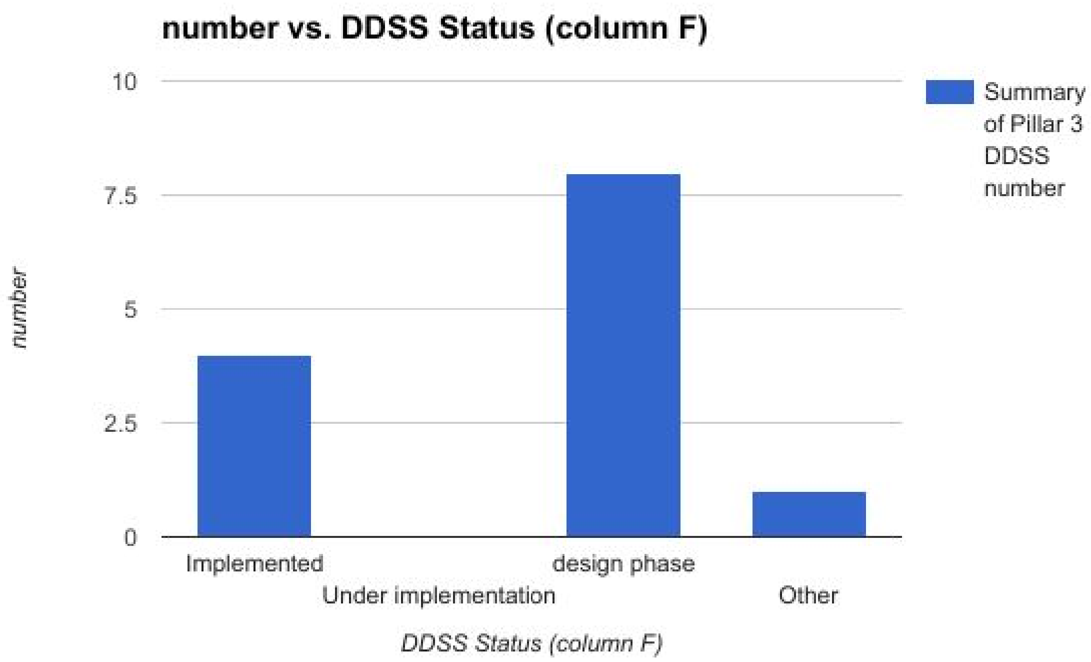


Figure 4. The histogram summarizes the status of development of all the DDSS of Pillar 3 (Hazard & Risk products).

In addition to the fault sources specifically adopted in the ESHM13, EFEHR provides access to the European Database of Seismogenic Faults EDSF via link to the relevant web site developed during the project SHARE (diss.rm.ingv.it/share-edsf/) where the database is still hosted. Recent developments of EDSF concern the preliminary realization of a Spatial Data Infrastructure (SDI) currently available for testing purposes at http://namazu.rm.ingv.it. This SDI provides access to EDSF data and metadata using the following Open Geospatial Consortium (OGC) standard protocols: WMS and WFS for data, and CSW for metadata (http://www.opengeospatial.org/standards). Metadata were compiled according to the recommendations of the INSPIRE directive.

Computational seismology

The computational seismology DDSSs are quite mature relying on the VERCE platform. The current suite of DDSSs targets the forward modeling of the seismic wave field in three-dimensional, laterally heterogeneous structures at different scale-lengths (i.e., from local to global). The products include:

- Calculation of synthetic waveforms.
- Visualization of waveform propagation.
- Production of synthetic shake maps.
- Gathering and processing data and synthetic waveform.

- Calculation of waveform misfit.

These products (DDSSs) are ALL available and implemented upon the VERCE platform (https://portal.verce.eu/). They have been developed adopting top IT solutions and have addressed from the start issues such as the AAAI and the selection of the most appropriate workflows. In addition, care has been made to the drafting of a user manual available at http://www.verce.eu/Training/UseVERCE/VERCEPortal-UserGuide-vers1.1.pdf.

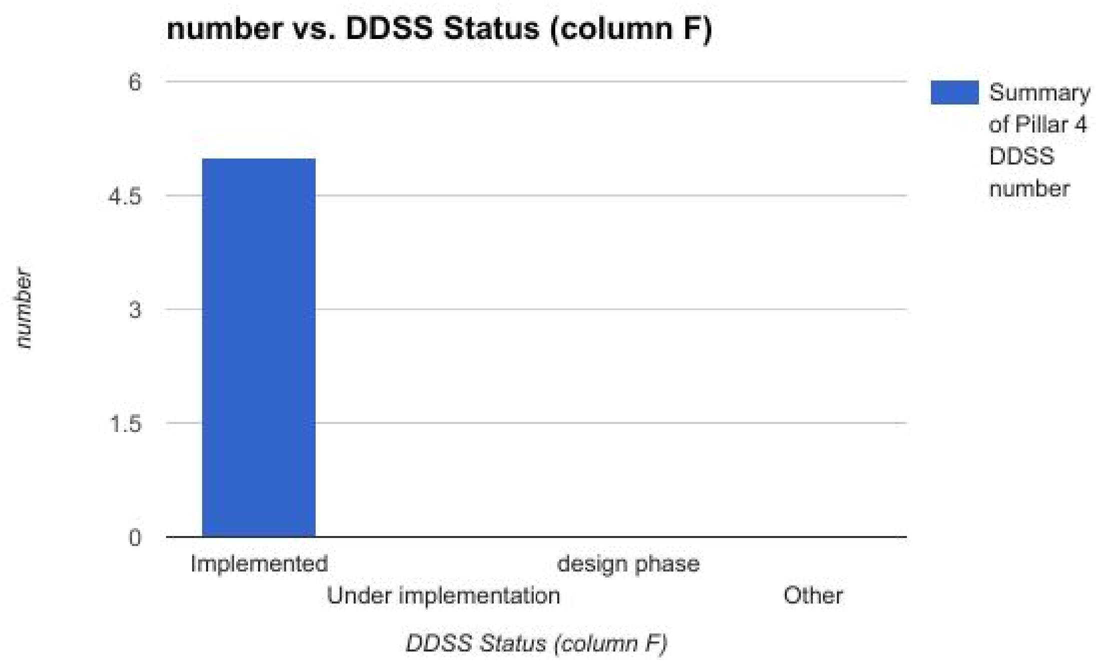


Figure 5. The histogram summarizes the status of development of all the DDSS of Pillar 4 (Computational Seismology products).

3. TCS roadmap

For those DDSS elements that are not in the priority list (see point above) TCS communities need to prepare a realistic time-table with the implementation of these additional DDSS elements. There should be a clear and realistic plan for the implementation of the additional DDSS elements including not only the technical implementation, but also their sustainability. This would require a Data Management Plan (DMP) including data storage, data curation, governance and financial issues.

The Seismology TCS is well on the way and we think that the fundamental DDSSs for access and download of *waveform*, *station* and *event* data are in place. These are the fundamental and minimal services for the researchers accessing the EPOS infrastructure. There are, however, several DDSSs still in the design or in the software prototyping stage and are not mature and will unlikely meet the M24 deadline for being operational. As of M18 (March 2017), the EPOS Seismology master table "WP8_DDSS_list" of the status of the DDSS table offers a realistic snapshot of the current level of the implementation of the DDSS.

4. Data Management Plan (DMP)

Each TCS should have a DMP for their own sustainability during the operational phase of EPOS (EPOS-IP). It would be essential that each DDSS element defines its own DMP discussing the following topics: i) Data Access Policy, ii) Data storage and maintenance responsibility, iii) Data curation responsibility, iv) Data management and governance structure, v) Financial commitment securing the operational costs.

A Data Management Plan is very relevant to insure preservation of all the data acquired. Although this issue applies to all the types of data of the TCS, it is especially relevant to the raw data acquired in the field that are at the origin of all the other "derived" data. For seismic waveforms, this is a well known issue and it is customary that e.g., each primary node of the EIDA federation applies the opportune preservation and curation policies to the data acquires and/or is responsible of. In addition, there has been discussed in the integration meeting held in Prague at the beginning of March 2017 that adoption of data curation standards as those provided by the Digital Curation Centre (DCC) https://dmponline.dcc.ac.uk/ is to be explored better..

5. Use cases

Please describe at least one envisaged use case for your TCS using the form below. Ideally you would be able to describe one internal and one cross disciplinary use case. Please keep in mind that when modelling the Use Case it is of crucial importance to focus on what the Use Case does, not on how it does it. Example use case is provided in Annex 2.

I Use case

Use case name / topic - **EQINFO** Very basic use case to gather all seismological data and data products after a strong earthquake in a given area and plot or list them.
Use case domain - This use case is: - Discipline-oriented, namely focusing on the discipline of: primarily seismology but it can be useful for other disciplines - Multidisciplinary, namely focusing on the disciplines of: seismology, geology, earthquake engineering
Use case description - *In this section the use cases will be outlined. This section may require iterative refinements.* - As a researcher in seismology (and/or as a duty officer at an operational center that needs to provide scientific information to people and media), I want to know all the relevant information available for a given area where a strong earthquake occurred in the past (or just struck). This will allow for the identification of the main tectonic, geological, seismological features and information of the area. This is important to gain more scientific information for proper first assessment of the existing information, download the data of interest and to better target new research to study further the area and the earthquake specifically (or for the information to be divulged to the media for example).
Actors involved in the use case - *A list of the actors who communicate with this use case.* - <IT System leader, Researcher, (duty officer of seismological operational center)> - <Seismologist>
Priority - *How important is this Use Case to the TCS?* - Very High.

Pre-conditions - A strong earthquake has occurred in a given area.
Flow of events – user view - *Determine the start and the end of the typical Use Case scenario. If a use case scenario grows too complex, it can be split up into subsequent scenarios.* - Basic sequences and needed steps (user view) - The following steps are needed to answer the question: 1. Either a <researcher> or <operational center> provides origin time, earthquake location and magnitude of a M6+ earthquake in the Euro-Mediterranean area. 2. < Researcher > wants to know what is known and available of the epicentral area and in particular: a. Maps of seismicity (e.g., 24H, 3d, 3m, 1y, 10y from the event origin time, OT) b. Maps of the available focal mechanisms (or moment tensors) c. Plots of the recorded strong and weak motions ordered by distance from the epicenter d. Shake maps of the earthquake e. Maps of historical earthquakes f. List of the historical earthquakes g. Seismic hazard map of the region h. Map and list of the known faults in the area
System workflow - system view - 1. In order to execute step 1, the existence of an operational center is not required since the user could be a scientist interested to trigger the whole procedure for research purposes (in this case the earthquake might be an old one). 2. The system based on the earthquake location identifies the: a. Rectangular area where to make the following queries to the different DBs and/or the country and associated institutions that have competence for the area and where the products are available (e.g., shakemaps, earthquake parameters). b. Shared storage area where the products will be made available. (The storage area is meant in a very broad sense since the material could be made available inside e.g. a predefined document as a report). 3. In order to execute step 2a, the system must perform the following actions: a. Query the DB(s) and extract the past events according to time (i.e., 24H, 3d, 3m, 1y, 10y) and to magnitude (e.g., M>3.0). b. Make as many maps with the seismicity as requested (one or more standard formats with predefined background geographic information (terrain, national and administrative boundaries, municipalities, roads and railways, etc.). c. Push the plots to the pre-defined storage area above. 4. In order to execute step 2b, the procedure is similar to 3 with: a. System queries the moment tensor DB(s). Note that there could be more than that available at EMSC.

b. Plots on the map the available moment tensors and, if available and reliable according to some quality index, the automatic solution.
c. Push the plots to the pre-defined storage area above.

5. In order to execute step 2c, the procedure involves the following:
   a. Submit the query using the FDSN web services to EIDA to download the relevant weak and strong motion data.
   b. Plot the waveforms according to distance using a time vs distance plot service.
   c. Push the plots to the pre-defined storage area above.

6. In order to execute step 2d, the procedure involves the following:
   a. Use the web services to query the shakemap DB (it is available at INGV, NIEP, NOA, ETHZ, ...) and download the relevant maps (MMI intensity, PGA, PGV, PSA03, PSA10 and PSA30).
   b. Push the plots to the pre-defined storage area above.

7. In order to execute step 2e and 2f, the procedure involves the following:
   a. Use the web services to query the AHEAD DB and download the available event catalogue for the area.
   b. Push the list in CSV format for example to the storage area above.
   c. Plot the retrieved historical earthquake catalogue.
   d. Push the plots to the pre-defined storage area above.

8. In order to execute step 2g, the procedure involves the following:
   a. Submit a query to retrieve the seismic hazard map for the target rectangular area.
   b. Plot the hazard map with the earthquake epicenter represented.
   c. Push the plots to the pre-defined storage area above.

9. In order to execute step 2h, the procedure involves the following:
   a. Submit a query to retrieve the known faults for the target rectangular area.
   b. Plot the known faults with the different symbols according to the faulting mechanism.
   c. Push the list of faults and their relevant parameters in CSV format to the storage area above
   d. Push the plots to the pre-defined storage area above.
10. Notification of completed job is sent to the seismologist, to the duty officer of the seismological operational center.

Post-conditions

- -

The job is finished when the products are present onto the created storage area (workplace storage).

Extension Points

- -

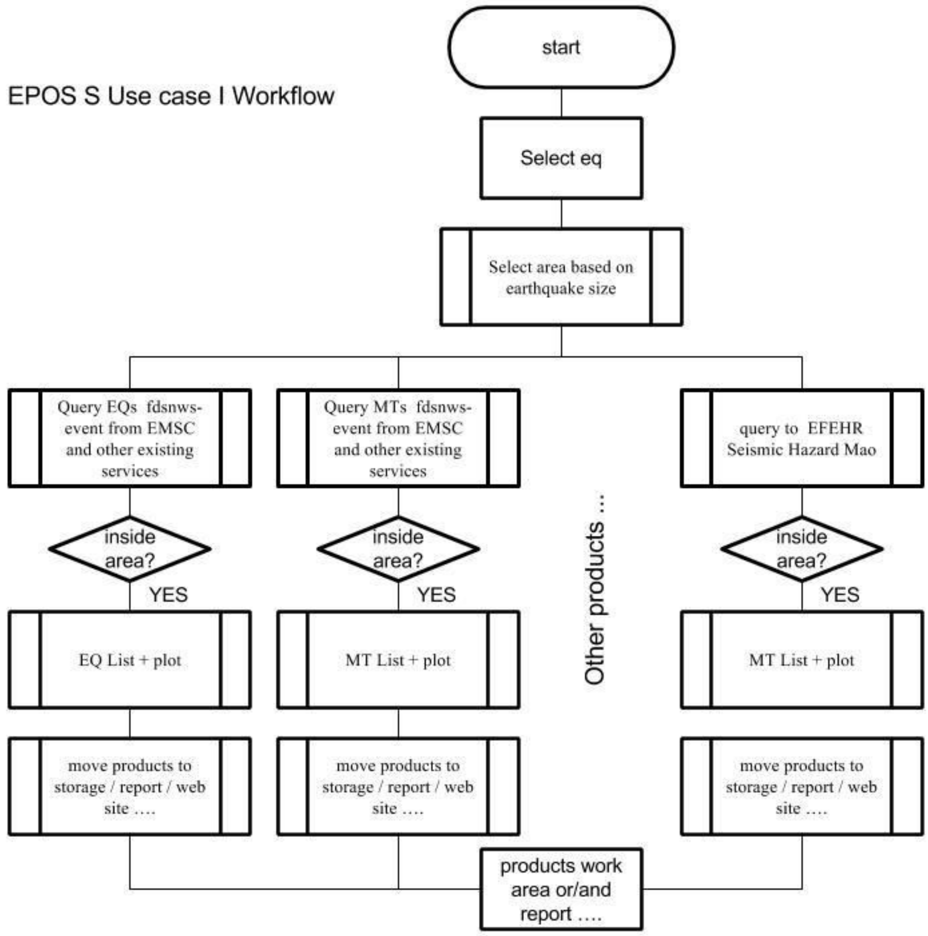
The use case can be easily extended to add more products: for example, the 3D seismic velocity structure is available. Similarly, the resulting products can be assembled into a report which can be edited and commented by the researcher.

« Used » Use Cases

- -

No, the use case is very basic and it avails of individual services provided for the most in the TCS-Seismology.

Other Requirements - Prompt availability of the data in a near real-time environment can become a prerequisite in case the user requests the information shortly after the earthquake of interest has occurred.
After the interview (to be filled in by WP7) - *Create class and sequence diagram for each use case.* - Class diagram and sequence diagram: see Annex 3 – Workflow diagram common symbols.



II Use case

Use case name/topic - A landslide event in the study region and he/she wants to investigate whether it had been triggered by an earthquake.
Use case domain -

This use case is: - Discipline-oriented, namely focusing on the discipline of: primarily geologists interested in landslides. - Multidisciplinary, namely focusing on the disciplines of: seismology, geology, geotechnical engineers.
Use case description - *In this section the use cases will be outlined. This section may require iterative refinements.* - The interest of the user has been triggered by a landslide event in the study region and he/she wants to investigate whether it had been triggered by e.g. a local earthquake. As the user has just observed the landslide but has limited information on the cause of the event, the user will try to collect data and try to identify when the event happened and what triggered the event.
Actors involved in the use case - *A list of the actors who communicate with this use case.* - <IT System leader, Researcher>   - <Landslide geologist>
Priority - *How important is this Use Case to the TCS?* - Moderate.
Pre-conditions - A landslide has occurred in a certain area.
Flow of events – user view - *Determine the start and the end of the typical Use Case scenario. If a use case scenario grows too complex, it can be split up into subsequent scenarios.* - Basic sequences and needed steps (user view) - The following steps are needed to answer the question: 1. <Researcher> chooses the geographical position and the presumed time of a landslide to defines the target area. Retrieves all the products available in a time interval (t0, t1) in an area with radius R< 100 Km from the location of the event (cf. Use Case I for detail). Restrictions based on the dataset and/or product type, name, etc... can also be applied. e.g. seismic waveform data (sensors type = broad-band station, accelerometric station), seismic event data (earthquake bulletin/parameters), GNSS data (sensor type = cGPS), satellite data (multispectral satellite images),... 2. The user will query the TCS-Seismology services (ORFEUS/EIDA) to retrieve continuous seismic waveform data from broad-band and accelerometric stations within the defined time window for the given region (circular area centered on the landslide event). The user may filter the seismic waveforms for a number of pre-calculated quality parameters and request domain specific metadata describing the sensors. This first spatial-temporal query returns a package containing the given waveforms in the domain specific format (mseed) filtered by a number of quality parameters and the domain specific metadata format (station XML). 3. The user will query the TCS-Seismology services (EMSC) to discover whether any seismic event within the defined time window for the given region (circular area centered on the landslide event) may have triggered the landslide. From a quick analysis the user identified that there was a seismic event in the target area that may have triggered the observed landslide. 4. The user will now query the TCS-GNSS data on a narrowed time interval and region to check whether a cGPS station is available in the area and eventually recorded some deformation at the surface.

Continuous GPS data will be retrieved together with the domain specific metadata. Results show no deformation visible at the nearest station.

5. The user will now retrieve satellite images (i.e. TCS Satellite data) and try to identify if the seismic event may have triggered the landslide. From the satellite images it is clear that the seismic event and the landslide event are unrelated, thus no chance yet to address at least the origin time of the landslide event. But now the user has a restricted time frame for the origin time of the landslide event identified from the satellite images.

6. The user goes back to the previously downloaded continuous waveforms and tries to detect the landslide event using the seismic data in a restricted time window based on the evidence in the satellite images. The user is successful to identify the landslide event origin time using the seismic waveforms.

7. Having identified the origin time of the landslide event and having proved that this was not triggered by a seismic event the user would like now to estimate the mass movement of the landslide. Unfortunately, it turns out that the seismic waveforms are in a format which is not easily manageable by our user (i.e. raw data) who is used to handle seismic waveforms in physical units.

8. Consequently our user decides to stage-in his assembled data set to the computational facilities (i.e. HPC or HTC, depending on the needs) which match the requirements of the specific computational problem and where the actual computation takes place after the retrieval of the data (transparent to the user). The user will then have a new assembled data set with instrumental correction applied according to the domain-specific metadata.

9. The user can now finally estimate the mass movement involved in the landslide event using a simple relation between the amplitudes in physical units and the mass involved.

10. The user has now an extensive data set composed by: instrumentally corrected waveforms, satellite images proving that the landslide event was not triggered by a seismic event, landslide (possibly with an identifier) with origin time and mass estimation and his/her own simple software written to detect the landslide on the seismic waveforms.

11. As a final step the user might want to share the results of his/her analysis together with the identifiers of the data products generated by him/her and make data and results available to his/her colleagues. The system provides an easy way to bundle the whole package, assigning identifiers and proper metadata, and make it available for discovery.

Once published the results of the analysis could be retrieved by scientists studying the same event from a different perspective. But, in order to facilitate their understanding and usability, the results need to be equipped with metadata which are rich enough and describe the type of product and the way to use it. Cross disciplinary links should also be made possible for example by means of annotations.

System workflow - system view

- -

1. Enable discovery based on multi-layered metadata: from high level discovery metadata to domain specific metadata (*Req 1*).

2. Identify, integrate and link data and products by means of PID (*Req 2*). 3. Enable data processing and visualization on distributed resources (*Req 3*). 4. Collect and manage provenance information (*Req 4*). 5. Bundle, publish and share results (*Req 5*). 6. Enable citation and reproducibility of science (*Req 6*). 7. Provide accountability and usage statistics *(Req 7*).
Post-conditions - The job is finished when the products are present and it has been set up accountability and usage statistics.
Extension Points - None.
« Used » Use Cases - Yes, the use case although different from the previous one adopts a very similar structure.
Other Requirements - Availability of metadata and PIDs.
After the interview (to be filled in by WP7) - *Create class and sequence diagram for each use case.* - Class diagram and sequence diagram: see Annex 3 – Workflow diagram common symbols.

6. CONCLUSION

The Seismology TCS is well on the way with a number of standard services offered through RESTful web services (i.e., REST architecture based web services). There are tentatively 46 DDSS that can be part of this TCS. Out of these, there are 21 DDSS that are implemented, 15 under implementation, 9 in the design phase and 1 provided as "available".

There are two important issues that need to be addressed. The first regards all those services that are being developed but are not yet ready. Some of them have been already prototyped and must undergo the final verification to be made operational whereas others are still at an earlier stage. Overall, our master table evidences that the "Waveforms", "seismological products" and "computational seismology" DDSS are all either "implemented" or "under implementation" whereas the development and the availability of the DDSS for the Hazard and Risk products and services is somewhat lagging behind with still 8 DDSS in design phase.

Secondly, we feel that all the services so far developed and operational (i.e., implemented; 21) require some additional link between each other. Indeed, we feel it is still missing that "glue" necessary to link the DDSS to one another. This glue must be in the form of relevant *metadata* and of *persistent identifiers* (PID) capable to index uniquely and access the products.

The final goal consists of providing tools that, besides helping the researchers with additional products to obtain analysis of the data reliably and quicker, include *provenance* procedures that will allow researchers to reconstruct all the individual steps of the workflows that have been adopted to calculate the results. In simple terms, the overall machinery adopted by EPOS ICS & TCS should provide all those ingredients/procedures/tools necessary to perform “reproducible science”.

Regarding the use case, the two proposed are both focussed on the availability of a data and follow similar data extraction procedures. The first use case is really simple and it consists of just the gathering of selected data & products. This use case besides being simple it can be very useful also in the context of operational centers and civil protection purposes. The second use case is somewhat more evolved and it entails data and products from several TCSs.