SERA
Seismology and Earthquake Engineering
Research Infrastructure Alliance for Europe

DELIVERABLE

D5.3 Scientific roadmap for the active and passive seismological community

Work package	WP5, NA3: Networking Deep Seismic Sounding data and products
Lead	Uppsala University
Authors	Roland Roberts, Ramon Carbonell, Monika Ivandic, Kuvvet Atakan
Reviewers	-
Approval	Management Board
Status	Draft
Dissemination level	Public
Delivery deadline	30.04.2019
Submission date	30.04.2019
Intranet path	DOCUMENTS/DELIVERABLES/D5.3_Scientific_roadmap_for_the_active_and_passive_seis_community

This project has received funding from the European Union's
Horizon 2020 research and innovation programme under grant
agreement No 730900.

Table of Contents

Summary

This document reviews the general situation regarding possibilities for future access, within the EPOS framework, to controlled and passive source seismic data, and proposes a route towards developing an effective data access system for researchers working with European seismic data. The document also describes some of the issues related to this type of data that the system needs to be able to deal with, such as different forms of data and metadata, saving old “legacy” data, data ownership issues, and other relevant matters, but also provides suggestions and recommendations on how to build and organize the system for handling them.

This document describes the general EPOS architecture in relation to the characteristics of the relevant types of seismic data, including metadata definitions, and considers the type of administrative superstructure which appears necessary in order to implement EPOS services of the envisaged type. Both the volumes and technical characteristics, and various complications related to intellectual property rights, suggest that a pan-European standing committee or group appears necessary in order to optimally deal with the seismic data. A model for how this might be achieved is described. In addition, we point to additional areas where EPOS should consider the development of services, resources permitting.

This report has been produced by workers within the working group, primarily from the University of Uppsala and CSIC in Barcelona. Contributors include Roland Roberts, Ramon Carbonell, Monika Ivandic, and Kuvvet Atakan.

1 Overall structure of EPOS data access concept, SERA, and seismic data

The following text presents in summary the basic features of the EPOS project and the related database. This text is written based on original EPOS documentation (e.g. Atakan and Michalek, 2017) and SERA deliverable D6.4 (Atakan et al., 2018), that are updated with current developments.

The European Plate Observing System (EPOS) is a single, pan-European Research Infrastructure plan for sharing solid Earth Science data, observations and research results. Its mission is to integrate the existing and future advanced European Research Infrastructures for solid Earth Science into a single, distributed and sustainable infrastructure taking full advantage of the new e-science opportunities and warranting increased accessibility and usability of multidisciplinary data. Through its IT platform ICS (Integrated Core Services), EPOS will enable sharing and integration of science data collected by the research infrastructures and facilitate common access to services from a single online environment.

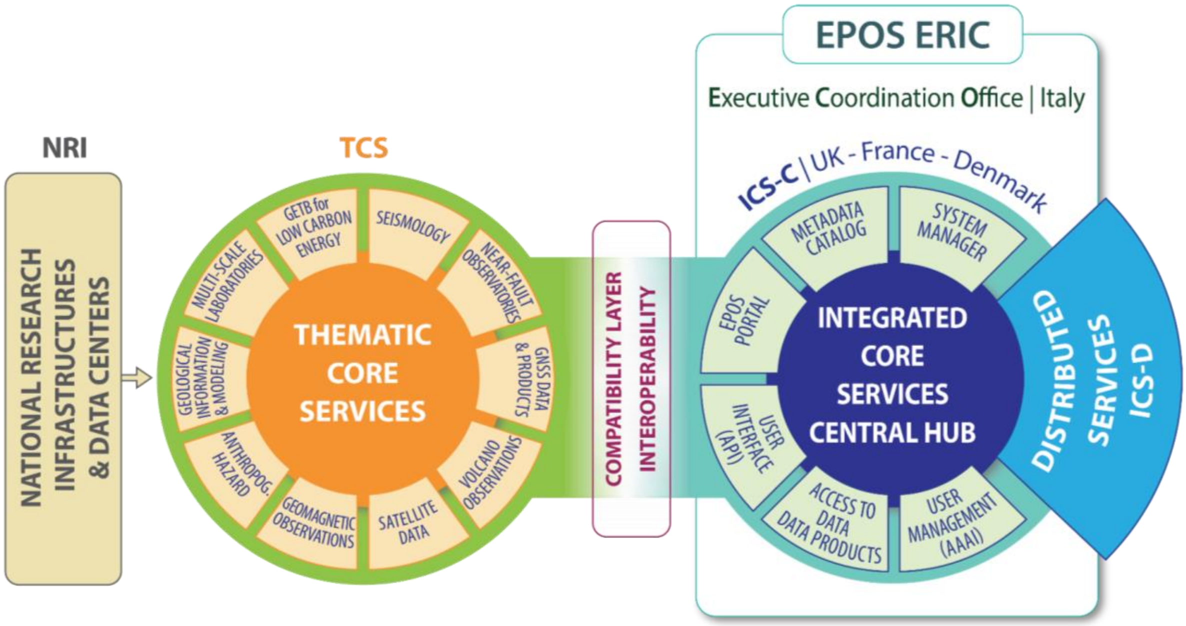

Figure 1. EPOS functional architecture.

EPOS functional architecture is based on a three-layer structure where the bottom layer, consisting of the national Research Infrastructures (RIs), represents the backbone of the EPOS (Figure 1). The second layer represents the community developments at the European level, where Thematic Core Services (TCS) are organised. Currently 10 different TCSs are operational in EPOS with varying degree of maturity. Some the TCS communities are quite mature and were established more than a century ago, whereas others were recently formed. The 10 TCSs are: TCS Seismology – WP08, TCS Near Fault Observatories (NFO) – WP09, TCS GNSS data and products – WP10, TCS Volcano observations – WP11, TCS Satellite data and products – WP12, TCS Geomagnetic observations – WP13, TCS Anthropogenic Hazards – WP14, TCS Geological data and modelling – WP15, TCS Multiscale Laboratories – WP16 and TCS Geo-Energy Test-Beds – WP17.

The third and the uppermost layer is the pan-European level where Integrated Core Services (ICS) are offered as part of the EPOS architecture. ICS Central-hub (ICS-C) will be hosted by the Geological Surveys in UK (BGS), France (BRGM) and Denmark (GEUS) jointly. The distributed resources (ICS-D) are designed such that external resources, such as High-Performance Computing (HPC) and High Throughput Computing (HTC), visualisation processing and analysis represent services that are decentralised and offered by third parties such as European Open Science Cloud (EOSC).

The National Research Infrastructures (NRIs) are research infrastructures and data centres that provide data to the Thematic Core Services. Thus, the NRIs represent the EPOS data providers that will guarantee access to quality-checked data and products and they provide services at national level.

The Thematic Core Services (TCS) are the pan-European e-infrastructures that provide and disseminate data and services to specific communities and international organisations (e.g. ORFEUS for seismology). The TCS are community-specific integration and they represent transnational governance framework with the scope to integrate the data, metadata and services arriving from various national and international infrastructures and data centres.

The Integrated Core Services (ICS) provide a new interface for users by adopting data access policies aligned to open science principles. The ICS make data, services and products accessible to users in a useable form that allow innovative, disciplinary and cross-disciplinary research. This e-infrastructure will allow access to multidisciplinary data, products (including synthetic data from simulations, processing and visualization tools), and services to different stakeholders, including but not limited to the scientific community (i.e. the main EPOS users). The key element of the ICS in EPOS will be a Central Hub (ISC-C) where users can discover and access data and data products available in the TCS and NRIs, as well as access a set of services for integrating and analysing multidisciplinary data.

The establishment of EPOS has a general time line which consist of four phases: The Conception Phase (2002-2008), The Preparatory Phase (2010-2014), The Implementation Phase (2014-2019) and The Operational Phase (2020 and after). In October 2018, EPOS became a European Research Infrastructure Consortium (ERIC) and thus a legal entity (EPOS-ERIC), with 12 countries signing as founding members. The ERIC legal framework provides EPOS with a legal personality and extensive legal capacity recognised in all EU Member States and the flexibility to adapt to the specific requirements of each infrastructure. Its legal seat is hosted in Italy (Rome) at the National Institute of Geophysics and Volcanology (INGV) headquarter.

Initially, in the EPOS preparation phase and EU-financed implementation phase project, focus was largely on specific types of data whose scientific and practical importance is clear and where a feasible and relatively rapid route towards implementation of EPOS access to such data could be identified. This includes e.g. seismological data, primarily in the sense of signals produced by natural or anthropogenically stimulated earthquakes. Controlled source seismic data, i.e. where a controlled man-made signal source is used, was discussed during the EPOS preparatory phase, but it was decided not to have any major focus on such data in the initial phase. Reasons for this include:

- the assessed need to focus initially on attaining proper functionality for data from a limited number of areas
- that a major concept motivating EPOS was natural hazards
- that there are very large volumes of controlled source seismic data
- that much of this data has been collected by commercial companies

This implies various complications regarding research access to the data. The issue of controlled source seismic data was revisited in the planning of the SERA project, leading to the inclusion of this work package in the SERA project. It is important to stress that this work package is a very minor part of SERA, and that resources for this work have been rather limited. The intention has thus been to take some steps on the route towards full practical integration of seismic data into the EPOS framework, but not to develop fully functional software solutions.

Different types of seismic data exist. In the SERA work package, it was decided to focus on “Deep seismic sounding” (DSS) data, in the sense of long-range seismic refraction and wide-angle reflection data. Long-range in this context means recording signals from sources at distances up to tens, hundreds and even thousands of kilometres. The recorded signals include “diving waves” which have followed curved wave-paths deep into the earth and back up to the surface, and wide-angle reflections from impedance contrasts within the Earth, such as that between the Earth’s crust and the underlying mantle. The effective penetration depth of such signals depends upon various factors including the length of the profile and the size of the source. Many DSS projects provide information though the whole crust and the upper parts of the underlying mantle. Focusing first on DSS data was regarded as sensible for various reasons: Much of this data has been collected by academic institutions and is thus fairly readily accessible and total data volumes are limited. The scale of the elucidated part of the Earth is, often clearly relevant with respect to the treatment of the earthquake seismic data, already actively included in EPOS. In addition, due to increasing environmental restrictions, it is difficult to use the large explosions necessary for some DSS surveys. Therefore, many projects are in practice not repeatable. So, it is important to try to fully secure data for the future, even from DSS projects which were performed several decades ago.

It is common to distinguish between the character of data described above and “near vertical incidence” or “reflection seismic” data, where the sensors registering ground vibrations are relatively close to the source. Much such data is commercial, and volumes of data can be vast. This SERA work package does to some extent consider near vertical incidence data, but primarily only in terms of some considerations of how such data may be made available in the future. There is no clear demarcation between DSS and near vertical incidence data. Nevertheless, the distinction between DSS and other data is meaningful and practically important in our context.

Other categories of seismic data exist. One such, which has become popular recently, is the use of methods which analyse seismic noise, i.e. simultaneous signals from many sources, as opposed to signals from individual earthquakes or controlled sources. One popular method for such passive studies is based on an elegant theoretical relationship between the cross-correlation function of data recorded at two stations and the seismic Green’s function between these two stations. Such studies require the cross correlation of long, continuous, series of data recorded simultaneously at the two stations. In some senses, such data is most naturally categorized together with earthquake seismological data, because both types of data involve long term recordings to analyse signal sources which we do not control. However, such passive studies also have much in common with some types of controlled source studies, e.g. that they may elucidate the same volume of the Earth. Therefore, passive data will be considered in this work package.

In the EPOS design architecture, observed data will in general not be transferred to a common, central EPOS database, but will remain distributed. To make data discoverable, metadata must be available in the central portal function, i.e. in the EPOS “ICS” (integrated core services). The architecture also includes functions known as TCSs (thematic core services) which are intended to facilitate the inclusion of data of specific types into the EPOS system. The inclusion of controlled source seismic data into the EPOS portfolio will require coordination between the involved instances, common and standardized data models, protocols and encodings, various types of software for data management and access, and perhaps some types of analysis. For the common system to work effectively and to be fully in line with standards in a broad sense (c.f. EUs “INSPIRE”) common standards and technologies must be agreed and implemented, in order to ensure full interoperability when working with different types of data simultaneously. Thus, it appears that some new TCS function must be designed and implemented, if DSS and other seismic data are to be fully integrated. This could mean the expansion of the remit of an existing TCS, or the design of a new unit.

When designing systems for data curation and access, it is often important to have considered what it is expected that the data will be used for, specifically in our case, for research. The envisaged research uses may significantly affect which functionalities the system should offer and are likely to strongly influence priorities regarding which types of data to first focus upon, and which forms of application software it is most important to develop or implement in the system.

2 Controlled source seismic data: Previous, current and likely future use and significance

Almost all of the Earth is physically inaccessible to us. Geological processes sometimes propagate material from depth to at or near the surface where it may be physically sampled, but this only indirectly elucidates structures and processes at depth. There are a number of different geophysical methods which can provide information about structures at depth. It is probably fair to say that, with current technology, seismic methods are the most powerful and significant methods for probing the deep Earth. Seismic signals can penetrate to all depths within the Earth, providing useful information on structures, including seismic P- and S-wave bulk velocities, the presence of boundaries causing an impedance contrast, and the presence and properties of material which is seismically anisotropic. Much of our current understanding of the inner structure of the Earth is based on such data. This includes e.g. that there is a crust overlying a (chemically) distinct mantle, all overlying a core of different material, the outer part of which is molten. Seismic methods can also reveal structures on much smaller scales.

The basic concept of controlled source seismic methods for probing the Earth is in general very similar now to what it has been for several decades: A suitable signal is generated and recorded by a number of sensors, deployed at or close to the surface of the Earth. Analysis of the travel times of different phases (wave propagation paths from source to receiver) provides information about velocity structures, impedance contrasts etc., and more detailed investigation of the details of the recorded waveforms can provide further information.

While the basic concept is not new, there have been very major technical improvements in the application of seismic methods. Sensors have improved, both in technical specifications as well as reduced costs, mean that many more sensors can be deployed as compared to earlier. Recording three components of ground motion (as opposed to only the vertical component) has become much more common, allowing enhanced analyses, especially regarding S-waves. Communication between sensors, e.g. using telemetry, has improved dramatically, reducing logistical problems, reducing costs for some types of survey, and allowing sensor configurations which were previously difficult to use. New technical sensor configurations, such as ocean floor sensors, optical fibre sensors, and snow- and landstreamers, offer new possibilities for measurement. Three dimensional arrays of sensors (and sources) have become much more common. The controlled sources themselves have been further developed, compared to the simple sources (often, explosions) used earlier. Offshore, arrays of airguns "firing" simultaneously are used to provide signal strength and to focus the source signal as desired. On land, vibrators transmitting signals at different frequencies may be used individually or together, offering the possibility of enhanced signal to noise ratio by prolonged measurement periods.

Seismic data has been routinely collected in digital form for many years. Improved analogue to digital converters (with more bits) allow better recordings, enhancing the possibilities for analysing weak signals buried in the waveform data, especially when multichannel data from many sensors is available. Improved surveying and geodetic methods for simple and precise estimation of the location (including altitude) of sensor positions can significantly improve results. New, increasingly sophisticated, data analysis software is continuously developed, making the ambition of modelling the full waveform data gradually more tenable. It is relatively common to reprocess older "legacy" seismic data, using tools which were not available at the time of initial processing. Sometimes this can significantly improve the analyses.

Thus, equipment (sources and receivers, communication, and ancillary equipment) has improved significantly, as have data analysis and modelling tools. Very many sensors are now often used, sometimes in an array as opposed to simply along a profile, opening new possibilities for noise reduction and signal enhancement when processing.

There are limitations to what may be achieved using seismic data. The resolution of Earth structure which is possible to achieve is constrained by various factors. One of these is the number and spatial distribution of receivers and sources. To some extent, this can be controlled, but cost and practicalities limit the number of sensors which may be used. On land, sensors must be at the surface of the Earth (unless they are in a borehole or otherwise underground), and offshore in the water or on the seafloor. The limitations where sensors may in practice be deployed limits possible resolution, and the limited number of sensors also limits the possibilities for signal enhancement and noise reduction by processing. In addition, the finite frequency of seismic signals limits resolution. Seismic signals are damped on propagation through the Earth, with higher frequencies in general being more heavily damped. This means a practical upper limit on frequency, and thereby a limit on resolution. This resolution limit is normally discussed in terms of Fresnel theory, with a rule of thumb that resolution of better than about a quarter of a wavelength is not achievable.

A reasonable question is therefore if the current limits on what can be achieved using seismic data are related to the fundamental physics of the situation (wavelength of the signal etc.), or if more and better data can (continue to) improve resolution. In general, despite the greatly increased number of sensor (and, often, source) locations used in surveys, there remains a significant potential for improved results by further improvements in the sensor arrays, specifically the number of sensors used. Thus, we are likely to continue to see significant increases in the number of sensors deployed in many surveys. This implies that seismic methods will continue to improve, probably significantly in the future, and also that the total volumes of data generated are likely to (continue to) increase considerably.

On all scales, from global (e.g. structures on the core-mantle boundary) to large scale (crust and upper mantle) to small scale (upper crust and near surface structures) in general available data means that we are far from the limit of resolution which is probably theoretically and practically attainable. In fact, with some exceptions, such as very intensively investigated oil- and gas-bearing sedimentary areas, our knowledge of structures within the Earth is very limited relative to what can be achieved with current technology.

This relates to the question of possible “diminishing returns” with seismic methods, in the sense that while a new, modern and extensive survey in an area previously investigated should be able to produce significantly better results, it is not necessarily the case that the scientific advances will motivate the cost and effort of the project, given that there are some earlier results. However, the Earth is very large, and seismic profiles sample a limited volume of the Earth, geographically close to the profile. It follows that the existence of previous data from a limited geographical area certainly does not preclude the possibility of significant new results if the area is revisited. In general, we are far from the point where new seismic surveys lack potential for significant new insights into geological structures and processes.

New data collection and analysis tools can mean that different experimental configurations become relevant, with possible implications for how data storage and access systems should be designed. One example of this is that the passive cross-correlation methods demand long continuous data sequences, rather than the short time segments relevant for classical controlled source studies.

Possible changes which may be relevant include:

- Further major increases in the number of sensors used
- More advanced sensors, including improved telemetric communication
- New types of sensor e.g. land-streamers, ocean bottom sensors, rotational seismology.

- A continued increase in the use of three-component sensors

Large international projects, gathering sensors from many different institutions, have long been common in some types of seismic projects, including long-range refraction (DSS) profiles. If very large numbers of sensors are to be routinely used, it seems likely that rather intimate collaboration between institutions may be required. The EPOS system should probably support this. At what level is less clear. EPOS has already established a database regarding the existence of some types of seismological equipment. Extending this in some appropriate manner might be sensible, with the aim of facilitating international co-planning of new projects.

2.1 DSS studies including near vertical incidence, distant earthquake and passive seismic data

There are a limited number of “long-range refraction” profiles. Earlier SERA reports contain extensive, but not complete, descriptions where such experiments have been carried out.

Despite the age of some of this data, it is potentially of continued scientific significance, for various reasons. Apart from DSS studies, we have few tools for investigating the Earth on a crustal and upper mantle scale. The total volume of data is limited, and it is sensible to try to secure this data long-term, especially given that many of the projects are unlikely to be repeated in any form, because of the very large sources (explosions) necessary will not be allowed. Near vertical incidence seismic can provide useful data from such depths, but these projects are very expensive, and in general can only cover rather limited geographical areas. Analysis of data from teleseismic and regional earthquakes can provide information about seismic velocities and boundaries in the crust and upper mantle. However, for both these types of data (near vertical incidence and distant earthquake data) the body-wave (P- and S-wave) ray-paths involved are close to the vertical. For DSS data, the ray paths are in general much more horizontal. Thus, the data types are in this sense clearly complementary. Seismic velocity structure on a crustal scale may be well-revealed by using local source tomographic techniques, but this works well only where there are many earthquakes geographically distributed over an area, and if these occur at a range of different depths. Passive studies of ambient noise using cross correlation can provide information about structures at crustal and upper mantle depth. However, in most cases on this scale the method works well only for surface wave noise components. The effective penetration depth of Rayleigh waves increases with increasing period and thus wavelength, so lateral resolution at depth is strictly limited.

2.1.1. Future developments and use of DSS

As previously mentioned, methods for the analysis and modelling of controlled source seismic data have improved significantly over time. Re-analysing legacy data using improved, modern, methods can sometimes reveal significant new features in the data and in the Earth.

An area of increasing relevance is joint inversion of different types of data sampling the same volume of the Earth. This may be seismic data together with some other type of information (e.g. gravity or electromagnetic data) but may also be different categories of seismic data. In this context, the legacy DSS data may be very useful even though the information content in the entire data set may be small compared to other more modern data, because the special character of the DSS data (e.g. many dominantly sub-horizontal wave-paths) can mean that it supplies important complementary information, possibly significantly constraining models primarily based on other data.

It follows that, at least to some extent, it is important to secure DSS data, even older data, for the future.

Appropriate actions may include:

- Continued work (cf deliverables D5.1 and D5.2) to identify where relevant data exists, in what form, how accessible it is, and who may be contacted about the data.
- Definition of which forms of data EPOS should focus on (recorded waveforms, processed sections, derived models, geological interpretations etc.)
- A function for access to software appropriate for analysis of DSS data and (especially) DSS data together with other forms of data.
- A coordination function for coming projects, where relevant to help design and perform DSS projects, but also to allow “piggyback” projects to e.g. near-vertical incidence offshore projects, where complementary wide-angle data may be collected by deploying stations on land close to the airgun profiles. Equipment appropriate for wide angle studies may differ slightly from that for near-vertical incidence studies, in that target frequencies may be lower, and sensor spacing higher, but some types of instrumentation may be appropriate for both types of study. Earthquake seismological sensors may be well-adapted to gathering DSS data, depending on frequency sensitivity, ease of deployment etc. Therefore, a coordination function regarding instrumentation for DSS studies may be helpful.

2.1.2 Inclusion of the near-vertical incidence data into EPOS

“Reflection seismic” surveys are often major, expensive field campaigns, collecting very large quantities of data. Spatial scales can vary, with some surveys focusing on the upper crust, some on the entire crust and part of the upper mantle, and others studying small scale structure very close to the surface. There is a link between the depth of interest and the appropriate spacing between recording profiles, or for 2D arrays, the extent and density of the sensor array. Much near-vertical incidence data is collected commercially; some by academic institutions and government authorities. Data ownership issues must be dealt with appropriately in a future system for facilitation of access to such data. Some countries insist that such data be deposited with a government authority, to make it secure for the future, but there are often strict confidentiality rules regarding access to and use of such data. Total data volumes may be very large indeed.

There are well-established and generally accepted data formats for the storage of industrial reflection seismic data, and these are commonly adopted also by academic institutions collecting such data.

Offshore, much high-frequency acoustic reflection data is collected, using e.g. side-scan sonar. Conceptually, this is very similar data to “seismic reflection”. Depending on instrumentation and how this is used, such surveys may reveal only seafloor topography, or may also show structures in the underlying shallow sediments. There is ultimately no strict demarcation between these different data types (sonar and seismic reflection). Sea-floor topography data is certainly relevant in an EPOS connection, but accessibility to such information need not necessarily be via access to the acoustic data itself, the derived data (seafloor topography) may be used. Much of the high frequency (acoustic) data reflects only the properties of the water, seafloor, and perhaps the very shallow underlying structure. In general, this data is not of central importance in an EPOS context, because of the limited possibilities of joint analysis with other types of data. Some high frequency studies do, however, reveal subsurface structures to some depth, and it is possible that such data is EPOS-relevant. Where such data is available in a form consistent with that used for reflection seismic data, it can in principle be included as such in the EPOS system. It seems doubtful that any priority should be given to other forms of high-frequency data, at this stage.

2.1.3 Passive seismic data and the benefits of integrated access to passive and DSS data

The analysis of continuous streams of data, recorded simultaneously by different stations, can be used for large scale studies of the Earth, to depths of tens of kilometres and beyond. Semantically, it is common to distinguish between observational seismology, which records naturally occurring (and anthropogenic) individual events, and methods which are based on the recording and analysis of a seismic “noise” wavefield, consisting of temporally overlapping signals from many different sources. While observational seismology is passive in the sense that we do not control the source, the term “passive seismology” is generally used to refer to methods for analysis of a noise field, as opposed to signals from individually identifiable events. Many passive analyses use a methodology based on two-station cross correlation, which is related to the point-to-point seismic Green’s function between the two stations. However, other analysis methods may also be used to analyze the noise field, and facilities for passive analysis should not be limited to a specific analysis methodology. At some frequencies, there are strong noise signals generated by the interaction between waves in water (generally, the Sea) and the underlying material, and often this noise field has a morphology (directionality characteristics) well-suited to passive seismic analysis. Passive analyses of such ocean-generated “microseismic” noise in general reveal surface wave signals which are sensitive to Earth structure do depths of some tens of kilometers. This is the depth range of relevance for most DSS projects. It follows that a system with integrated access to such data and to DSS data is sensible. In addition, passive methods are to a rapidly increasing extent being used for smaller scale projects, comparable in spatial scale with reflection seismic projects. The higher-frequency sources of relevance here are in general not oceanic in origin, but are generated by many different natural and anthropogenic phenomena relatively close to the recording site. In such higher frequency data, clear body wave signals may be found, i.e. signals of the same types as those using in reflection seismic surveys. There appears to be major potential for the further application and development of such high frequency passive methods, and this is currently a field undergoing rapid development.

There are also other types of seismic noise, and associated analysis methods. This includes e.g. volcanic tremor data, which is of relevance for the EPOS activities related to volcano monitoring. It is therefore important that the seismic data facilities which EPOS offers are appropriately generic in character, in order to be well adapted to the various different types of seismic data.

2.2 Data access through EPOS: Data characteristics and Metadata concepts

Seismic data is stored in various different formats. This is partly a consequence of the practicalities around different types of data being different (e.g. single station contra sensor arrays; single source can array of sources), partly due to historical developments. It does not seem sensible or feasible, at least in the short to intermediate term, that a common data format should be used for the many diverse types of seismic data. However, for the data to be integrated into a homogenized EPOS system, the data should be delivered to the user in a standardized form, or possibly a limited number of standardized forms. This will require an interface between each data archive and the EPOS system, which will convert data to standard form.

A necessary step in coordinating data access is clearly defined homogeneous metadata standards.

Data ownership, and responsibility for long term curation of data, should not be transferred to EPOS.

When addressing questions related to metadata, a first step is to consider the general character of the data. For seismic registrations, considerable effort has already been expended within EPOS for earthquake and induced seismicity data (**Appendix I**; provided by Henning Lorenz, UU, TCS-GIM, WP15; see 5 Appendicies (page 19)).

Some types of DSS data (list not exhaustive) are presented in the Table 1.

Table 1. Examples of DSS data.

Data Category	Data Types	Data descriptions and comments
Raw DSS data (deep seismic sounding, long range refraction, wide-angle data)	Time series data for an individual sensor point Associated location information, including altitude Sensor type Original recording equipment type and current form of storage Station comments	The time series ground motion data may be one component (usually the vertical), but is more commonly three component. It should be clear which coordinate system the locations are expressed in. For older data, station locations are from map-reading, and may be inaccurate. Where possible, how the position was measured should be included. Sensor type should include the sensor manufacturer and identification number, if possible, with a link to information regarding sensor performance. Older data may have been collected on e.g. f.m. tape. It may be important to know this e.g. because the timing accuracy may depend on the recording mode. Sensors may be deployed on different material (e.g. bedrock or sand) and the operator may have noted noise sources nearby.
DSS data sections	Commonly, reduced time sections, after filtering. These may be digital time series, or may be in the form of images	Information about station position has been converted to source-receiver position. The reduction velocity is important. Filter parameters are important For images, presentation parameters, such as the form of scaling of the traces, should be noted. Some sections using more advanced processing e.g. including three component information may be relevant
Data extracted from the DSS sections	Arrival times of identified phases in the section	The information should, if possible, include what the phase is identified to be, who picked the phase, and from which section (filtering etc.).

Models derived from DSS data	This is primarily the results of modelling of the arrival time of identified phases in the data, producing a two-dimensional velocity model, of P or S velocities	Entities will be represented by the occurrences or mines, which could either be spatially be represented by points or polygons. The data are already handled by Minerals4EU and it should be examined whether these services can simply be integrated "as is".
Orientation data	Location of the entire profile	This includes the locations of all stations and source points, possibly for several profiles within the same larger project. Descriptive information, such as project name and date, are important.

A very similar table can be presented for near-vertical incidence data. However, this is much more likely to be available in properly documented digital time series form, focus is on arrays (of receivers and sources) rather than single station, and much ancillary information (e.g. locations) will be more reliable. Processing sequences are often more complex for reflection sections than for refraction (DSS) sections.

For passive raw data, an important factor is the relative timing accuracy between stations to be compared by cross correlation.

2.3 Links to other organs within and outside EPOS; Organisational considerations

While all seismic data records ground motion, the character of the controlled source seismic data is in some ways different from that of the earthquake registrations already included into the EPOS system. However, the system for handling the controlled source data should, as far as sensible, build upon the existing concepts for the earthquake data. This will mean some common features in the metadata structures, but also considerable differences. The data formats used for the earthquake data internally in the EPOS system (mini-SEED) and for deliveries to users are unlikely to be fully appropriate for the controlled source data, especially the seismic reflection data, where SEG-Y is an industry standard. Therefore, quite different data storage formats can and should be used. DSS data may be more similar to earthquake data in character, implying that a common data format is more feasible. This is because, while the DSS data set is in a very real sense array data, the DSS data may be recorded by individual stations, each having its own specified location and instrument and site characteristics, recording sources one by one. Therefore, a structure seeing each single station recording as an individual "event" may be sensible. The near-vertical incidence data is in character more fundamentally array data, and it is most natural to think of the data as such. Whether or not it is best to consider common data structures for near-vertical incidence data and DSS data, or earthquake data and DSS data, depends on various practicalities. Probably, it is best if both possibilities are easily feasible in the database.

Passive data, in its various forms, should be essentially similar to data recorded by earthquake observation systems (permanent stations in seismological networks).

Irrespective of this, close coordination with the existing earthquake seismology component in EPOS is clearly necessary. In the first instance this means ORFEUS (Observatories & Research Facilities for European Seismology), which has the role of a TCS for the seismological waveform data. Administratively, one possibility could be to request that the role of the ORFEUS TCS is expanded in order to also include controlled source seismic data. This would presumably immediately take the form, at least in an initial period, of a distinct working group under ORFEUS. The apparent alternative of a quite separate TCS for e.g. the seismic reflection data is probably undesirable.

The data access functions via the ORFEUS TCS are based on a number of (EIDA) nodes distributed around Europe. It is possible that the role of these nodes could be expanded such that in a technical sense they also dealt with the controlled source data. Administratively, such a node system could be a problem. Much of the controlled source data is owned by commercial interests. In general, it seems unlikely that such interests will be prepared to directly interact with EPOS, in terms of delivering data. They may, however, be prepared (or obliged) to deposit data in some national repository, under state control. For robust and convincing control over access to data, some of which may have restrictions imposed by the owners, it seems likely that an operative function at national level may be necessary. It is a small step from this to the concept of data delivery nodes in each participating country. In fact, the number of EIDA nodes has been increasing, so the difference between regional and national nodes is clearly both conceptually and practically smaller than previously.

3 Recommendations

The EPOS community should consider the establishment of a function, possibly a new group, with national representatives who will together develop concepts for the full inclusion of controlled source seismic data into the EPOS framework. This group could be ORFEUS, TCS-GIM or it could be part of a new separate TCS (e.g. Thematic Core Services in Earthquake Engineering).

The technical details regarding practicalities of the inclusion of DSS data into the EPOS framework should be investigated and documented (deliverable 5.4 of this work package), and some initial practical investigations should be carried out and assessed (deliverable 5.5). Appropriate metadata and data formats should be explicitly defined.

Node software capable of transforming relevant data formats (data at national and local level) into a standardized EPOS metadata and data format should be developed. Such software will allow the stepwise establishment of data supplying nodes at national level. For the near vertical incidence data, the data structures should be as closely in line with industry standards as is feasible. For other data (e.g. much of the DSS data) the formats already used by ORFEUS may be appropriate. These nodes should also be capable of facilitating access to derived data, including data sections and derived models.

Legacy data, where the risk of data loss may be significant, should be explicitly identified, and priority should be given to securing such data for the future. The most obvious mechanism for achieving this appears to be via the national representatives in the group mentioned above. In some cases, it may be appropriate to transfer data from its original owner to a new institution, in order to facilitate long term storage and access. In general, this should be dealt with at national level, but if this is not possible, then a common function, based on agreement, should be considered.

When and if resources allow, EPOS should offer functions which can aid in the organization of new large-scale seismic projects. This includes offering channels for information regarding possibly available equipment.

EPOS should also offer software repository functions, for various types of data, including controlled source seismic. However, as much advanced software for reflections seismic processing is commercial, an ambition that EPOS should offer access to all or most relevant software is not appropriate.

In a subsequent step, a link between literature and data would be advisable. It is of special interest for old data, which generally lack of a Digital Object (DO) and from which several publications have arisen. Scholix (Framework for Scholarly Link eXchange, Burton et al., 2017) aims to enhance an open information system that link scholarly literature and data. A Scholix Link Information Package will contain information about the original data and literature, as well as information about the link package itself (metadata). Therefore, access to all seismic information (both data and publications) would be straightforward.

In general, there seems to be only relatively weak motivations for EPOS considering offering services related to controlled source seismic data beyond rather basic (but effective) data access functions. However, some simple tools for examining subsections of data may be useful for helping users to assess data relevance. Naturally, the possibility of showing metadata information graphically (primarily in the form of maps), and effective tools for examining in detail selected geographical areas, selecting data sets based on such images etc., is part of the basic EPOS ICS concept, and should be offered.

An internal coordination among the SERA Work Packages where integration to EPOS is relevant, is necessary. There is already some parallel work done in WP6 (see Deliverable D6.4), which might be useful to consider in the coordination effort. There seems to be a growing interest from the Earthquake Engineering community to become a dedicated new TCS in EPOS. This may provide an opportunity to develop more coherent set of services and their subsequent integration in EPOS.

4 References

Adrian Burton, Martin Fenner, Wouter Haak, Paolo Manghi, 2017. Scholix metadata schema for exchange of scholarly communication links. https://doi.org/10.5281/zenodo.1120265

Atakan K., Michalek J., (2017) EPOS-IP Deliverable 6.1 (M24). Description of TCS Requirements and Use Cases. European Plate Observing System (EPOS) Implementation Phase Project No. 676564, InfraDev programme - Horizon2020, 24p.

Atakan, K., Athanasopoulou, A., Avouris, N., Bousias, S., Lamata Martinez, I., Molina, J.F., Pegon, P., Sintoris, C., Tsionis, G. (2018). Review of current SERIES and EPOS databases. SERA Project Deliverable D6.4, 49p.

Mikael Pedersen, Sylvain Grellet, Henning Lorenz, Martin Hammitzsch, Sébastien Hameau, 2017. D15.5: Specification and documentation of web service interface for geological thematic services, EPOS.

5 Appendices

5.1 Appendix I: Active Seismic Index

The Active Seismic Index is a register that will contain discovery data for datasets that origin from any kind of seismic field survey. The primary purpose is to integrate discovery functionality for such data sets in EPOS via a service provided by the TCS Geological data and models.

The approach will follow the TCS design of discovery indices and associated services for each data domain (of which data from active seismic experiments will be one). The intention is to implement the functionality described in the use cases that have been added to deliverable D15.4 in an internal revised version: D15.4_M18_int-rev1, March 2018.

Seismic surveys are traditionally subdivided by the primary design of the experiment into seismic reflection surveys (mainly for imaging the upper to middle part of the Earth's crust) and seismic refraction surveys (Deep Seismic Sounding, DSS). However, this subdivision has no impact on how the survey is described by discovery attributes. Thus, it is taken care of by one attribute that will allow a clear distinction in a search process in case this is desired by the user.

Another important distinction is the layout of the survey:

1. "2D survey": the survey moves in a line across the Earth's surface and images the structure below as a profile.
2. "3D survey": the survey images an area either by a moving or a fixed array of seismic receivers that are distributed over an area on the Earth's surface and which record simultaneously.
3. Borehole survey: the survey follows the three-dimensional path of a borehole. Receiver arrays are moved along a borehole between recordings, the source is either placed on the surface or at (fixed or variable) depth in a second borehole (cross-hole survey). Multiple receiver-boreholes are possible. The seismic source and receivers can also be integrated into a single logging probe that is used to characterise the near surroundings of the investigated borehole.
4. Complex geometries, e.g. addition of far-offset or off-line receivers to the above.

It is not possible to describe the exact geometry of an entire survey in the discovery data, but the type of geometry will be reflected by an attribute with attached controlled vocabulary.

5.1.1 Discovery data for the Active Seismic Index

The discovery data for datasets from active seismic surveys attempt to reflect the information that a potential (re-)user of the data requires to discover data sets of interest and decide which of those actually is relevant for the intended purpose. Visualisation of the survey design and results (images, link to models) might be an option for the future. The attributes of the discovery data are based on:

- attributes in the databases of the geological surveys
- information that is regarded as important in scientific publications (i.e. presented to the reader in a prominent place and not hidden in the methods or annexes)
- initiatives/projects to make science data openly available

- discussions in the scientific community
- an appropriate research infrastructure that allow for exchange of scientific data.

Table 2. The discovery data for datasets from active seismic surveys.

Attribute	Explanation	Remark
	General information	
name	Name of the survey	
nameCampaign	Name of the acquisition campaign or project	
identifier	Unique identifier for the survey	DOI URL for data publication; URL to database record; or similar
identifierCampaign	Unique identifier for the campaign	if available; e.g. the URL to a landing page
surveyDesign	Original design of the survey	code list (reflection, refraction)
surveyGeometry	Geometry of the seismic survey	code list (e.g. 2D, 3D, borehole 'moreActiveSeismicsIndex-basics_v0.1 specific?', ...)
surveyDepth	Imaging depth of the survey - depth below surface for 2D, 3D - radial for seismic downhole probe - difficult to define for VSP & cross-hole	How can we handle the different measures that can be given here, i.e. time (TWT) and distance?
nominalFold	Nominal fold of the seismic survey	What to do with this attribute for level 0 data? Fill it with the value known from the processed data?
description	Text with additional information on the survey purpose and/or summary of results, e.g. abstract	free text
specification	Text with additional information on the survey design & acquisition (including environmental conditions and other factors that could have an influence on the data)	free text
processing	List of processing steps that were performed to create the data set from raw data	free text

dataFormat	Format of the data set that is linked via attribute ‘dataset’	code list
dataQuality	Quality assessment of the data set	c.f. discussions about borehole
creator	Persons responsible for the data and intellectual content (processed data)	
custodian	Person or organisation that is the custodian of the data	
contact	Contact person or organisation that can provide more information about the data set	
license	URL to license text	
legalConstraints	Explanation in case no license is given or certain legal restrictions apply	ISO
Funding reference	Contributor funder and funder ID	
detailedDocumentation	Reference to publication or similar with detailed information on the survey	Need to decide whether to have (multiple) link(s) here
Related work	Publication, thesis or similar which refers to the dataset	Citation, DOI, URI…
preview	URL to preview image of the data set	
dateAcquired	period of data acquisition alt. year of acquisition	Format according to ISO 8601
Geometry – Spatial coverage	Geometry that represents the survey, e.g. - CDP line as polyline - polygon that represents the area of the 3D survey - 3D borehole trace or borehole start point - GeoNames permanent link of nearby villages	
dataset	URL to data set alt. download page	
dependencies	Physical objects (e.g. borehole) and data sets (e.g. raw data) that the present data set depends on/was derived from	List(?) of unique identifiers, e.g. DOIs, IGSNs
Technical parameters		
receiverType	Type of receiver	code list

receiverFrequency	Design frequency of the receivers used in the survey	
receiverModel	Model name or designation of the receivers used in the survey	It’s more important for the seismic source, but I added it for consistency/HL
spreadType	Type of receiver spread	code list
numberChannels	Number of receiver channels	
receiverSpacingX	Distance between receiver in the direction of movement (for stationary 3D survey, shortest distance between receivers)	
receiverSpacingY	Receiver spacing perpendicuilar to ‘ReceiverSpacingY’ (for 3D surveys)	Is this enough to describe the survey? Special set-ups could be described in ‘specifications’
numberReceiverPoints	Total number of receiver points	
sourceType	The type of seismic source used during the survey	code list
sourceModel	Model name or designation of the seismic source	If known. Good information for the expert
sourceSpacing	Distance between the source points	Not sure how to handle more complex set-ups, maybe not necessary here
nearOffset	Distance between source point and first receiver	
maximumOffset	Distance between the source point the most distal receiver	
recordingSystem	Model name or designation of the seismic recording system	
sampleRate	Sample rate applied during recording	
recordLength	Recording time at each source point	Mainly useful for level 0 data
FieldLowCut	Lowest frequency recorded	
FieldHighCut	Highest frequency recorded	

Contact

Project lead	ETH Zürich
Project coordinator	Prof. Dr. Domenico Giardini
Project manager	Dr. Kauzar Saleh
Project office	ETH Department of Earth Sciences Sonneggstrasse 5, NO H-floor, CH-8092 Zürich sera_office@erdw.ethz.ch +41 44 632 9690
Project website	www.sera-eu.org