

## Building Solid Foundations for Managing Solid-Earth Data with EPOS and EUDAT



*Luca Trani is a researcher and IT architect at the Royal Netherlands Meteorological Institute ([KNMI](#)), where he works for the R&D Seismology and Acoustics Department and for the ORFEUS Data Center (ODC), which is hosted by KNMI. ODC/KNMI is part of the ORFEUS (Observatories and Research Facilities for European Seismology) European Integrated Data Archive ([EIDA](#)) and is participating in the current phase of the EUDAT project together with the National Institute of Geophysics and Volcanology ([INGV](#)) in Italy and the Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences ([GFZ](#)). ORFEUS is deeply involved in the European Plate Observing System ([EPOS](#)).*

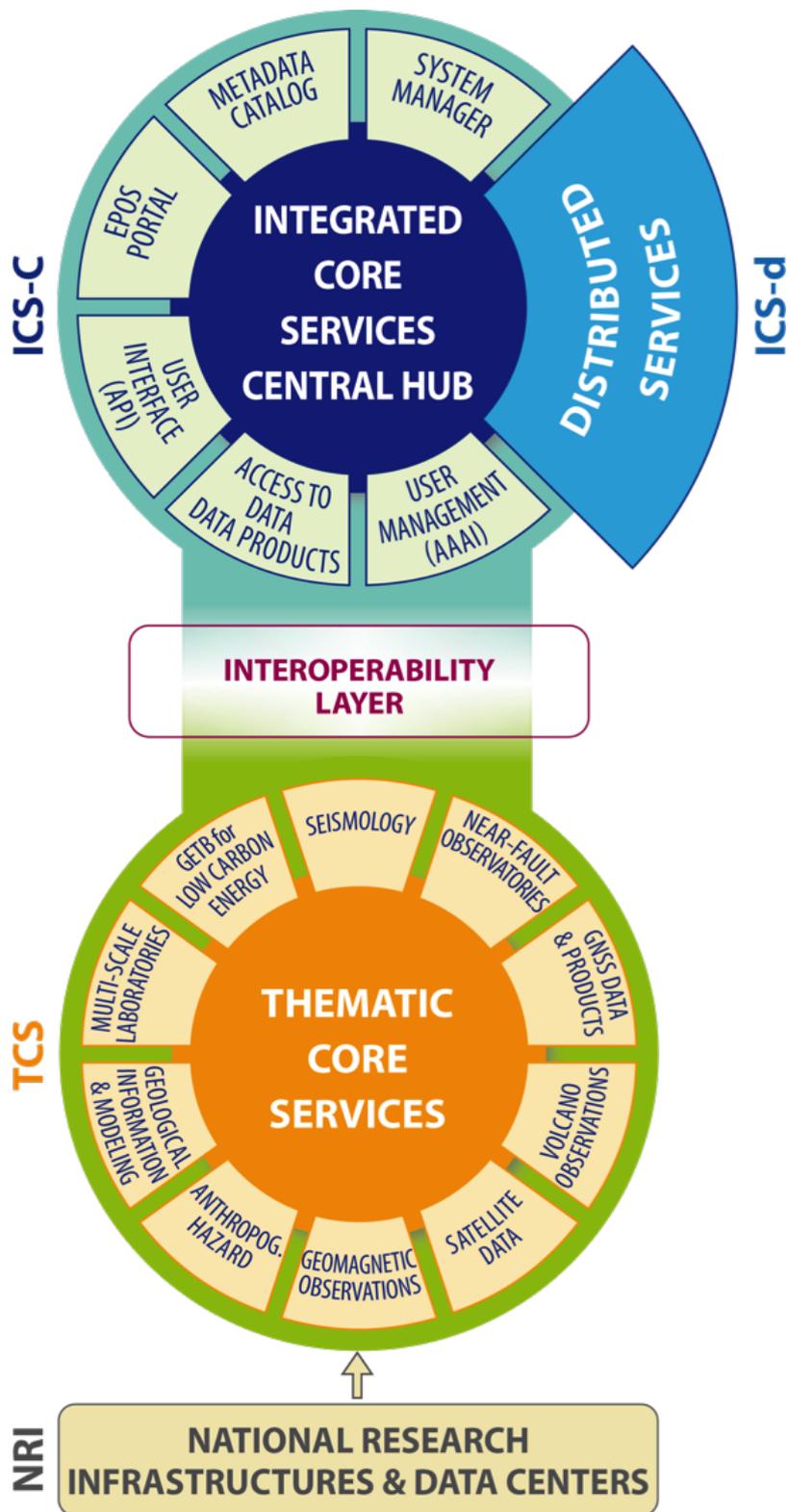
*Luca is responsible for the design and implementation of services and infrastructures for data management, and is currently the chair of the EIDA Technical Commission, which is in charge of IT development within EIDA. EIDA is a fundamental pillar in EPOS and, in particular, it is accountable for the EPOS seismological waveform data. Luca has been involved in the design of the EPOS architecture since its Preparatory Phase (EPOS PP). Some of his interests are interoperability, distributed systems, workflow engines and data intensive computing.*

### **Good morning, Luca. For people who are new to EPOS and EUDAT, please introduce us to EPOS...**

Right, briefly put, the European Plate Observing System ([EPOS](#)) is the integrated solid-Earth sciences research infrastructure approved by the European Strategy Forum on Research Infrastructures (ESFRI). It was included in the ESFRI roadmap in December 2008. Essentially EPOS is a long-term plan for integrating existing national and international research infrastructures within the Earth sciences.

EPOS aims to create a pan-European infrastructure for the solid-Earth science research that is needed to support a safe and sustainable society. In accordance with this scientific vision, the mission of EPOS is to integrate the diverse and advanced European research infrastructures for solid-Earth science research. This is important as research in the solid-Earth sciences is relying more and more on new e-science opportunities to monitor and unravel our dynamic and complex Earth system. EPOS will make it possible to perform innovative multidisciplinary research to give us a better understanding of the Earth's physical and chemical processes which control earthquakes, volcanic eruptions, ground instability and tsunamis, as well as the processes driving tectonics and the Earth's surface dynamics. EPOS will improve our ability to manage our usage of the subsurface of the Earth. Through the integration of relevant data, models and facilities, EPOS will enable the Earth sciences research community to make significant progress in developing new concepts and tools for finding key answers to scientific and socio-economic questions concerning geo-hazards and geo-resources, as well as the environment and human welfare.

To achieve the necessary levels of integration and interoperability between the participating research communities, EPOS has designed an architecture capable of organizing and managing the interactions between the different parties and assets involved in EPOS. The design of this architecture takes technical, governance, legal and financial issues into account, and thus EPOS envisages that the following four complementary elements will be included in its pan-European infrastructure.



**Figure 1: Overview of the EPOS architecture**

National Research Infrastructures (NRIs) will contribute to EPOS while being owned and managed at a national level. They will be the basic EPOS data and service providers. The NRIs have a significant economic value both in terms of their initial construction budgets and their ongoing yearly operational costs (which are typically covered

EUDAT receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 654065.



by national investments that need to continue during the implementation, construction and operation of EPOS).

The Thematic Core Services (TCS) will integrate the research communities within EPOS and their assets. In essence, the TCS will be a technical and governance framework where data and services are provided. It will also act as a forum where our research communities can discuss matters such as sustainability strategies and the implementation of their services, as well as legal and ethical issues.

The Integrated Core Services (ICS) are a novel e-infrastructure that we envisage consisting of services that will give researchers access to multidisciplinary data and data products, along with “synthetic” data from simulations, processing, and visualization tools. The ICS will consist of the ICS-Central Hub (ICS-C) plus various distributed computational resources (ICS-d). The ICS will be the place where the integration of the TCS occurs.

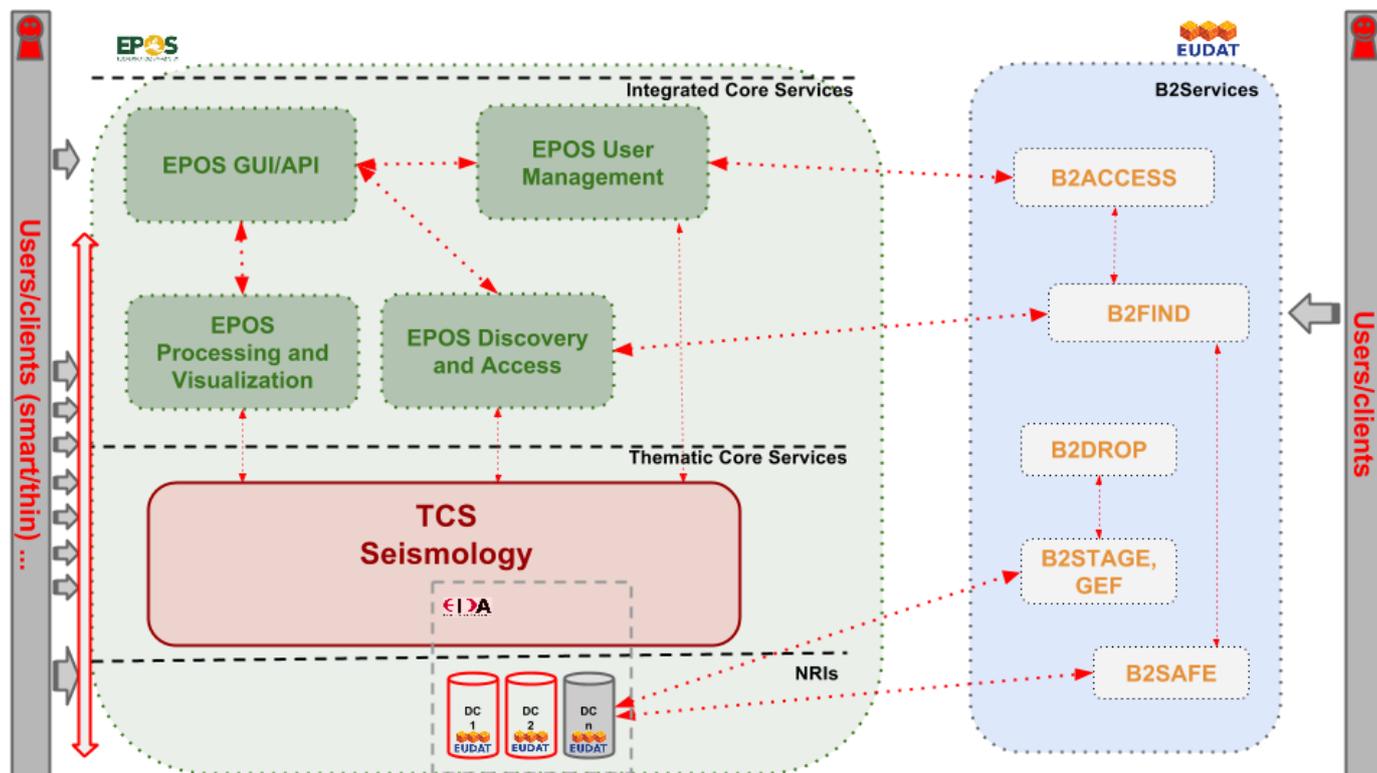
The Executive and Coordination Office (ECO) will be the EPOS headquarters and also the legal seat of the distributed infrastructure, and thus it will be responsible for governing the construction and operation of the ICS in addition to coordinating the implementation of the TCS.

***Thanks for that overview, Luca. When did the EPOS project start and what is the current status of the four components that you mentioned?***

The EPOS-PP project was launched in 2010 and lasted four years. Afterwards there was a follow-up with the EPOS Implementation Phase (EPOS-IP) project, which is currently ongoing and which will finalise the implementation and lead to the envisaged operational phase (from 2019). EPOS builds on top of existing NRIs and fosters and promotes the consolidation of the research communities into the TCS. Resources and guidelines will be provided to achieve the harmonisation and integration of Data, Data products, and Services and Software (DDSS) between the TCS and the ICS, and also within the TCS. The TCS services typically have existing infrastructures as their backbones, however they will need dedicated developments to meet the EPOS requirements. The TCS services have different levels of maturity ranging from well-established to novel, and a major challenge will be to bring all the services to the same level of maturity within the EPOS-IP project. The ICS constitutes a novel infrastructure which will be constructed from scratch but which will benefit from the preliminary studies and experience gained during the EPOS-PP project where, in particular, an architecture was designed and tested in an incremental process. After a call and a selection procedure, the Executive Coordination Office (ECO) country has been selected. The ECO will be established during the third year of the EPOS-IP project, that is to say when the EPOS European Research Infrastructure Consortium (ERIC) starts. The timeline of the technical developments fits nicely with the incremental uptake of the EUDAT services.

***Which EUDAT services is EPOS using or going to be using, Luca?***

In actual fact, potentially every EUDAT service could be of interest to EPOS. Given our broad scope and the fact that we embrace several heterogeneous research communities with a wide variety of requirements and use cases, EPOS constitutes the perfect ground for deploying and harnessing the full power of the EUDAT tools and services.



**Figure 2: Possible integration of EUDAT services**

Pragmatically speaking though, it is best if the uptake of EUDAT services is progressive and incremental. EPOS has a strong component that is currently involved in the current phase of the EUDAT project, namely ORFEUS/EIDA. The EPOS representatives within EUDAT (namely KNMI, GFZ and INGV) act as precursors in that they are the first to be adopting EUDAT services by embedding them in their facilities and daily data management routines. This integration process actually started during the initial phase of EUDAT. These three organisations – GFZ, INGV and KNMI – which are supported by national infrastructure providers – respectively the Karlsruhe Institute of Technology (KIT), the Italian consortium CINECA, and the Dutch e-Infrastructure provider SURFara – currently utilise (or are planning to use) services like B2SAFE, B2DROP, B2STAGE, B2ACCESS, and B2HANDLE. And this list is going to grow as soon as new EUDAT services become mature enough to be embraced by our research communities. Moreover these three current EUDAT partners within EPOS have an important role in outreach which is bidirectional – on the one hand, they are targeting other EPOS communities and raising awareness about the EUDAT services, and on the other hand they are helping in the EUDAT development process by bringing in new requirements that come from the broad EPOS ecosystem in order to improve and enhance EUDAT’s services.

***It is great to see how enthusiastically and comprehensively EPOS is embracing the EUDAT services. Luca, you mentioned that it is best if the uptake of these services is progressive and incremental. Would you share with us how is EPOS managing the uptake of the EUDAT services in this way?***

Certainly. In this initial phase we have targeted a subset of the EUDAT services as candidates for our uptake plans. In our distributed data infrastructure (namely EIDA), we would like to perform the deployment of the following B2services: [B2SAFE](#), GEF (the Generic Execution Framework), [B2STAGE](#), B2ACCESS, [B2DROP](#), and [B2FIND](#). Furthermore we would like to generate a metadata catalogue from [B2SAFE](#) and extend the functionalities of the afore-mentioned EUDAT services, in particular [B2SAFE](#) and [B2FIND](#), in order to match the requirements of our research communities.

Some of the typical things that our researchers need to do, that is, our initial use cases, include:



- management, replication and preservation of community data archives,
- data discovery through the metadata catalogue(s), including more in-depth searches using community-specific parameters and community-level services, for example, through [FDSN](#) services web services or corresponding services, and data download, and
- data discovery and staging into and out of high performance computing (HPC) resources where sophisticated analyses, such as data-intensive computations, are performed.

***Some of these services have already been made available to EPOS researchers, haven't they, Luca? How are the EUDAT services that are already in use helping the EPOS community and researchers in their work and research?***

Yes, one example that I can mention is [B2SAFE](#). This service is currently being used to facilitate long-term preservation of seismological datasets that are enriched with persistent identifiers (PIDs) and replicated onto external data facilities. We plan to extend this use case in order to improve access to the data and increase the likelihood of new discoveries using the data.

EPOS aims to achieve what is known as federated data management and discovery. (In short, that means that all our different databases will seem as though they are working together as one very large database, so, for example, people will not need to worry about searching through all the individual databases one at a time.) Data management is a pivotal issue to any distributed data archiving system: it should always be possible to identify where data is stored, to find out whether the same data is available at data centres aside from the one where the master copy of the data is stored, and to know what quality checking has been performed and when. These are all characteristics that ensure that a distributed system, such as that of EPOS, will be both robust and reliable. At the same time, data acquired by EPOS (and replicated onto EUDAT resources) should be discoverable and promptly made accessible. Although EPOS is developing its own solution for the purpose of achieving maximum interoperability, EPOS is also seeking to exploit the services provided by EUDAT as much as possible. A roadmap should be defined to gradually permit EPOS users to easily manage, discover and access data, whether it is stored on our own EPOS resources or on storage systems proper to EUDAT.

In combination with [B2STAGE](#) we might soon simplify the staging of datasets onto/from computational resources, thereby minimising data transfers.

Federated identity management is also a major requirement in EPOS - this would enable all our members to use the same credentials to access all our databases and networks. We would like to simplify and harmonise the authentication and authorisation processes used at our different sites and [B2ACCESS](#) could provide us with a viable solution for doing that.

Some of our users are interested in using [B2SHARE](#) to publish the results of experiments and make them available to their colleagues. We also have a use case focused on [B2DROP](#) that aims to minimise data flows to/from users' machines by creating a personal user space in the "cloud" where temporary requests can be stored.

We have identified a number of possible applications of EUDAT services to date, and doubtless many others will arise. In general, the EUDAT services contribute significantly by automating many technical chores, simplifying the daily work of our researchers and improving overall efficiency.

***Thanks, Luca. Was there a particular reason why EUDAT's services were of interest to EPOS? Where there other alternatives that EPOS could have used?***

As EPOS is integrating different research communities with different levels of maturity, it is fundamental to our requirements to have a large degree of freedom in order to make interoperability possible. EUDAT services are particularly suitable for facilitating this process as they can be deployed across our research communities to complement or augment the existing services belonging to the communities with more mature data management,

---

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.



---

as well as being used by the communities with less mature systems as a gateway towards their integration within EPOS.

An additional consideration in the choice of the EUDAT services is that EUDAT constitutes a solid backbone that could ensure sustainability in the long term. Of course, given the broad scope of EPOS and our range of requirements, our whole set of issues cannot be addressed solely by EUDAT. For instance, the provision of the computing resources needed for simulation and analysis will need to be handled by a pool of potential resources that result from integrating different e-infrastructures. However EUDAT can constitute the glue and underlying middleware that facilitates the exploitation of those resources in a consistent way.

***So it sounds as though the collaboration with EUDAT has made it easier for EPOS than trying to solve all the data-service problems on its own...***

Yes, the EPOS Implementation Phase has just started with the EPOS-IP project being launched on the 5th of October this year. EUDAT will for sure simplify the task of implementing an infrastructure as complex as that envisaged by EPOS. EPOS will leverage existing services and infrastructures by integrating and building on top of them. EUDAT can be a complement to EPOS by sharing responsibilities and amortising the development efforts, thus making it possible for us to focus on specific requirements rather than building everything from scratch. One of our key principles at EPOS is “Do not reinvent the wheel” therefore we will rely on EUDAT where possible for specific components which can provide fundamental “bricks” to build the EPOS “house”.

We can and do achieve much more by working together, and I would like to acknowledge my colleagues at INGV and GFZ for the contribution their expertise has made to this interview. I would also like to share a final remark... I believe that the collaboration between EPOS and EUDAT has the potential to become a clear and long-lasting example of successful synergy between e-infrastructure providers and research communities. Collaboration implies much more than just technological exchange; it is mainly about people and building trust - EUDAT and EPOS have all the ingredients to establish and maintain a successful framework for cooperation for many years to come.

[Read more](#)