



## Towards a European Collaborative Data Infrastructure



Digital Agenda for Europe (A Europe 2020 initiative)

includes

Open Access to scientific knowledge strategy:

Objective is to optimise the impact of publicly-funded scientific research, both at European level (FP7, Horizon 2020) and at Member State level.

The strategy is to develop and implement open access to research results from projects funded by the EU Research Framework Programmes, namely FP7 and Horizon 2020.

Open access requirements are based on a balanced support to both 'Green open access' (immediate or delayed open access that is provided through self-archiving) and 'Gold open access' (immediate open access that is provided by a publisher).

In recent years, significant investments have been made by the European Commission and European member states to create a pan-European e-Infrastructure supporting multiple research communities.



## Towards a European Collaborative Data Infrastructure



### MOTIVATION:

In recent years, significant investments have been made by the European Commission and European member states to create a pan-European e-infrastructure supporting multiple research communities.

As a result, a European e-infrastructure ecosystem is currently taking shape, with communication networks, distributed grids and HPC facilities providing European researchers from all fields with state-of-the-art instruments and services that support the deployment of new research facilities on a pan-European level.

However, the accelerated proliferation of data – newly available from powerful new scientific instruments, simulations and digitization of library resources –, has created a new impetus for increasing efforts and investments in order to tackle the specific challenges of data management, and to ensure a coherent approach to research data access and preservation.

EUDAT aims to address these challenges and exploit new opportunities using its vision of a Collaborative Data Infrastructure.



## Towards a European Collaborative Data Infrastructure



EUDAT Consortium (FP7 Project)

Start date: 1st October 2011

- Duration: 36 Months
- Budget: 16.3 M€ (9.3M€ EC)
- EC Call: INFRA-2011-1.2.2
- Consortium: 25 partners from 13 countries
- Objectives:

National data centers, technology providers, research

Cost-efficient and high-quality CDI

Meetings users' needs in flexible and sustainable way

Across geographical and disciplinary boundaries

Leading organisation: Finnish Center for Scientific Computing (CSC)

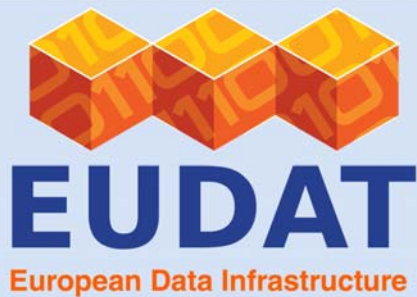
**Project Coordinator: Kimmo Koski**

**Project Manager: Damien Lecarpentier**

**Scientific Coordinator: Peter Wittenburg**

**Dissemination Manager: Nagham Salman**





## Towards a European Collaborative Data Infrastructure



EUDAT represents a unique partnership between research communities and data centers.

It brings together data service providers and users who are directly involved with the design of data services. The consortium includes key representatives from research communities in:

CLARIN Linguistics

EPOS Earth sciences

ENES Climate sciences

LIFEWATCH Environmental sciences

VPH Biological and medical sciences

INCF International Neuroinformatics Coordinating Facility

### EUROPEAN PLATE OBSERVING SYSTEM

The European Plate Observing System (EPOS) is the integrated solid Earth Sciences research infrastructure approved by the European Strategy Forum on Research Infrastructures (ESFRI) and included in the ESFRI Roadmap in December 2008.

The preparatory phase project (EPOS PP) started on 1 November 2010 and will last for four years.

The real challenge for EPOS is to successfully coordinate- and provide access to- the data infrastructures for solid Earth science in Europe. This requires strengthening the European capability to create high quality data, both observed and simulated, and to facilitate access to data products.

The vision for the EPOS Integrated Core Services is:

- to serve different stakeholders,
- guarantee access to data and data products (including software for data processing and modeling),
- provide tools for outreach, educational and training programs,
- establish conditions for long-term data preservation, with a focus both on IT solutions and governance.

EUDAT can provide IT solutions that would be difficult for the solid Earth sciences community to provide on its own.

## EPOS Data & Products:

Data will be available from the Solid Earth science disciplines which each community deals with, such as

- seismology,
- volcanology,
- geology and surface dynamics,
- geodesy,
- geomagnetism,
- analytical and experimental laboratory research,
- rock physics and petrology,
- satellite information.

## EPOS DATA AND PRODUCTS:

Available data from each community will be quality controlled according to the appropriate standards as defined by each of the disciplinary data providers;

In order to facilitate the integration among the communities, data had been categorized in the following levels:

Level 0: raw data, or basic data (example: seismograms, accelerograms, time series, ...)

Level 1: data products coming from nearly automated procedures (earthquake locations, magnitudes, focal mechanism, shakemaps, ....)

Level 2: data products resulting by scientists' investigations (crustal models, strain maps, earthquake source models, etc...)

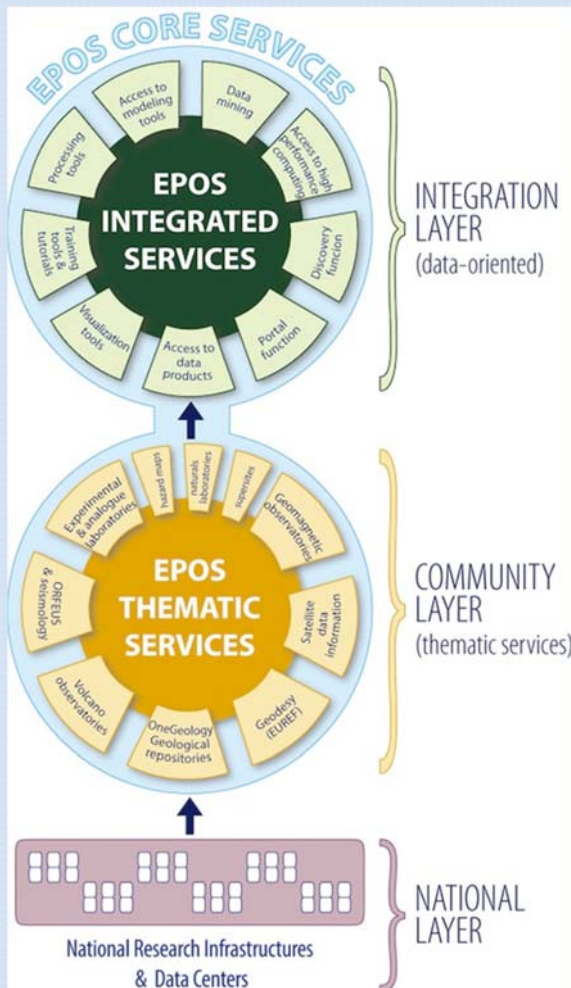
Level 3: integrated data products coming from complex analyses or community shared products (hazards maps, catalogue of active faults, etc....)



## e-ICT Architecture

To develop the EPOS architecture, a coherent ICT approach is needed. This is made up of four models, technically defined as:

- the resources model describes the available resources from national or local Research Infrastructures together with the resources needed to run the Integrated and Thematic Core services;
- the data model describes the metadata, the data taxonomy, the format and other technical details;
- the processing model controls the way processes are constructed and executed in the e-infrastructure;
- the user model controls the way in which the end-user interacts with the e-infrastructure.



## FIN-EPOS CONSORTIUM

Chairs: Annakaisa Korja (Institute of Seismology, University of Helsinki) and Markku Poutanen (Finnish Geodetic Institute)

Participating organization and principal investigators:

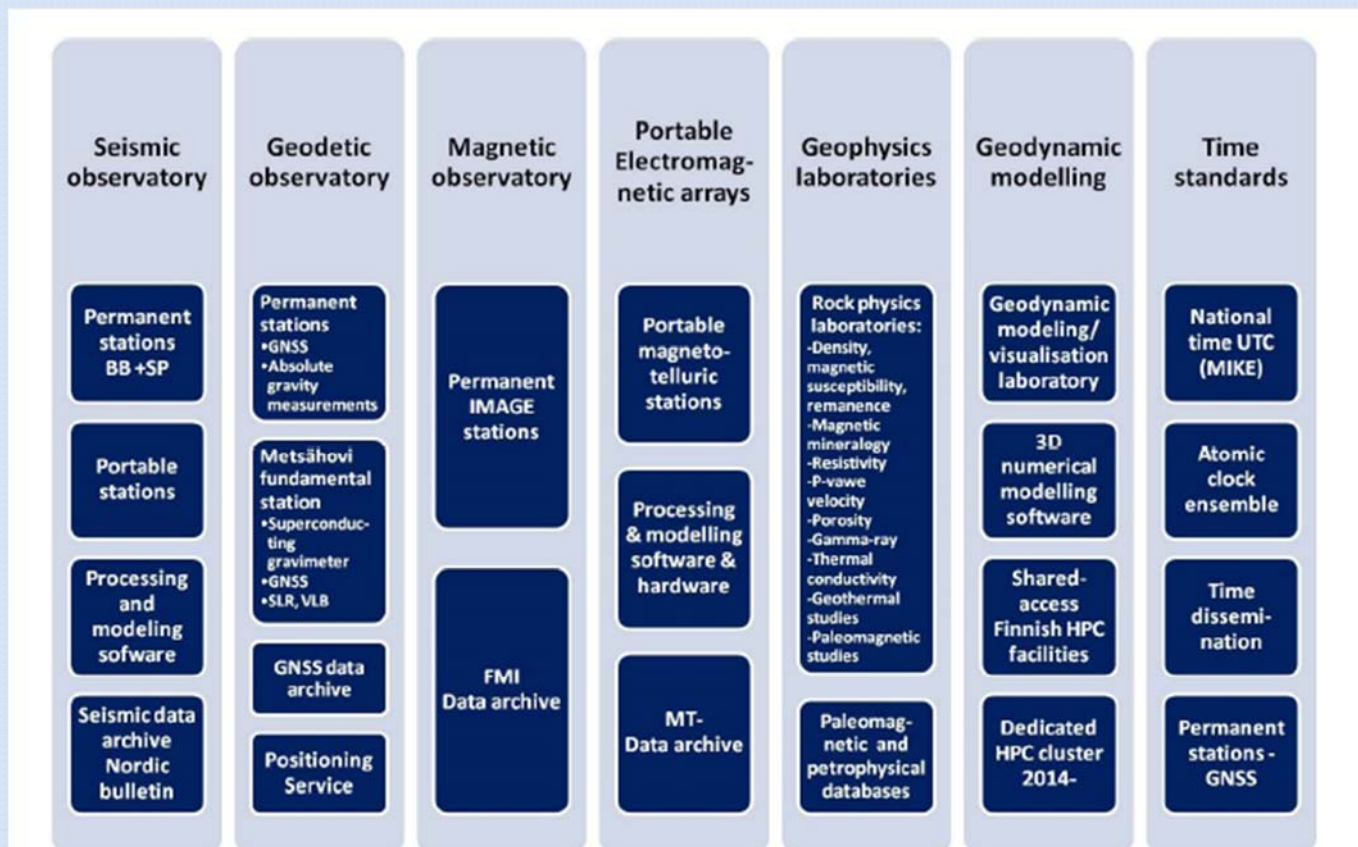
- RD Annakaisa Korja ISUH - University of Helsinki, Institute of Seismology
- Dir. Pekka Heikkinen ISUH - University of Helsinki, Institute of Seismology
- Asst. Prof. David Whipp ISUH - University of Helsinki, Institute of Seismology
- Prof. Ilmo Kukkonen UH - University of Helsinki, Department of Physics
- Prof. Karri Muinonen UH - University of Helsinki, Department of Physics
- Prof. Markku Poutanen FGI - Finnish Geodetic Institute
- Adj. Prof. Elena Kozlovskaya UO/SGO - University of Oulu, Sodankylä Geophysical Observatory
- Adj. Prof. Toivo Korja UO/GF - University of Oulu, Geophysics
- Dr. Kari Pajunpää FMI - Finnish Meteorological Institute
- Prof. Eija Tanskanen FMI - Finnish Meteorological Institute
- Adj. Prof. Satu Mertanen GTK - Geological Survey of Finland
- Dr. Damien Lecarpentier CSC – IT Center for Science
- Dr. Mikko Merimaa MIKES- Centre for metrology and accreditation

# INTRODUCTION: FIN-EPOS CONSORTIUM

## The scientific focus of FIN-EPOS

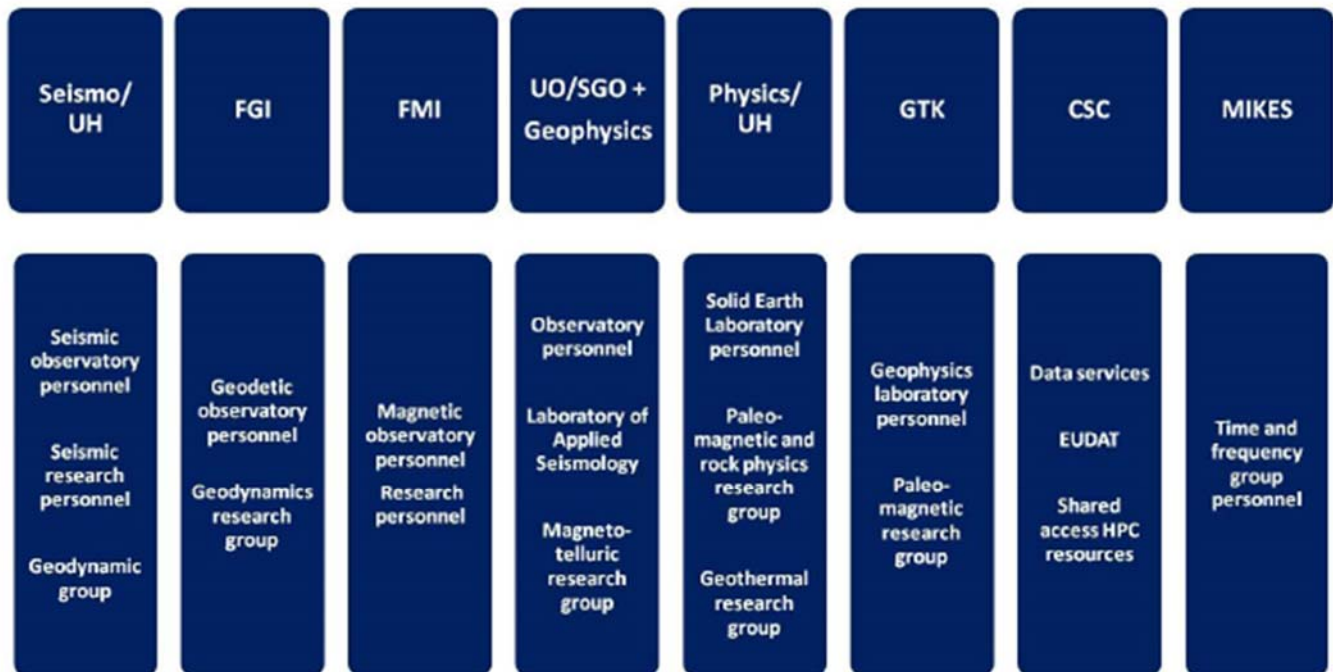


## FIN-EPOS RESEARCH INFRASTRUCTURE AT THE MOMENT:





# Finnish national EPOS council



## FIN-EPOS CONSORTIUM

A timetable of FIN-EPOS

2014 EPOS preparatory phase

- Finnish scientists join all pertinent working groups
- Finland signs LOI for EPOS ERIC
- Data management plan

2015 -2020 Construction phase

- Finland becomes full-member
- Finland joins General assembly
- Scientist are nominated to advisory boards
- Data flows to Thematic core services
- Finland joins in EPOS ERIC

## CLARIN:

### Common Language Resources and Technology Infrastructure

The CLARIN project is a large-scale pan-European collaborative effort aimed at making language resources and technology readily available for the whole European Humanities (and Social Sciences) community. This includes coordinating the development of appropriate resources. Amongst other things, CLARIN will offer scholars tools for computer-aided language processing.

The European Resources Infrastructure that CLARIN will create is based on an open European Federation of strong service centres and repositories that jointly provide:

- (i) knowledge of existing language resources
- (ii) coordinated creation of, archiving of, and access to such resources,
- (iii) access to services and tools that would allow scholars to utilise such resources, and
- (iv) bundling of and access to expertise related to specific language processing problems.

## What the CLARIN initiative offers:

- Comprehensive service to the humanities disciplines with respect to language resources and technology.
- Technology for overcoming the many barriers created by institutional, structural and semantic interoperability problems and fragmenting the resources and tools landscape.
- Tools and resources that will be interoperable across languages and domains, thus addressing the issue of preserving and supporting the multilingual and multicultural European heritage.
- Comprehensive training and education programs that include university education in the different member states.
- Improvement and extension of web-based collaborations, i.e. creating virtual working groups breaking the discipline boundaries.
- Development or improvement of standards for language resource maintenance.
- A persistent and stable infrastructure that researchers can rely on decades to come.



## CLARIN Key Technologies

To achieve these challenging goals CLARIN will be built on and contribute to a number of key technologies coming from the major initiatives advancing the eScience paradigm:

- CLARIN will use Data Grid technology to connect data repositories as implemented in the DAM-LR pilot project, as well as web services provided by the various centres;
- CLARIN builds on ideas launched by the Digital Library community to create Live Archives, and will further such initiatives;
- CLARIN incorporates, and contributes to, Semantic Web technology to overcome structural and semantic encoding problems;
- CLARIN incorporates advanced multi-lingual language processing technology that supports cultural and linguistic integration.

## European Network for Earth System Modelling (ENES):

### MOTIVATION:

Because of the direct consequences that climate changes can have on national economies and on our lifestyles, accurate scientific information is required by government and industry to make appropriate decisions regarding our global environment.

It is therefore the responsibility of the scientific community to accelerate progress towards a better understanding of the processes governing the Earth system, and to improve our predictive capabilities.

We need to build an advanced software and hardware environment in Europe where we can develop, improve and integrate our most advanced high resolution climate models.

## What is the role of ENES in EUDAT?

The ENES community will add to EUDAT's existing services and long-term archived data for interdisciplinary applications.

It is also working on scalability aspects of the federation, on workflow engines and web services, data curation and preservation, authentication and policy rules as well as interfaces to data archives in a federated environment.

ENES is making its expertise with its own community federation of data servers available to EUDAT: This federation will be integrated into EUDAT.

## Why is EUDAT important for ENES?

ENES is already being hit by the 'data tsunami', and this volume of data will just continue to grow.

By collaborating with other scientific disciplines also experiencing these data challenges, ENES can adapt the architecture of its own federation of data servers to meet this new reality.

The climate research community will also benefit from easier access to data from other disciplines, because climate researchers, and especially those working on evaluating the impact of climate change, require data from multiple scientific fields to perform their research effectively.