

# Copenhagen e·Infrastructures Observatory White Paper

---

## *Towards a European strategy for an e·Infrastructures Observatory*

### The state of play

Europe has built a world-class, high-capacity, high-performance communication and computing infrastructure, complemented with scientific data infrastructures, and e-Science services to facilitate research. Those electronic infrastructures (or e·Infrastructures) empower scientific communities with ubiquitous, trusted and seamless access to facilities, resources and collaboration tools, bringing to them the power of technology for communication, computation, storage and instrumentation.

Scientists have now light-speed access to distributed scientific facilities and databases, in-silico experimentation, as well as new services for collaborative work across geographical, disciplinary and organisational boundaries, strengthening scientific excellence and promoting European innovation and industrial competitiveness.

EU policy and financial support to e·Infrastructures has been -and continues to be- at the global forefront, reaffirming the priority that Member States place on research as the foundation of European competitiveness. Yet, designing and implementing improved policies and strategies requires better knowledge of the status of e·Infrastructures development at global, regional and country level, including evolution over time, as well as, scientific, social and economic impact due to their use by individuals, communities and people.

Several initiatives and projects in different contexts are currently concerned with addressing the above need<sup>[1][2][3]</sup>, using different methodological approaches and are currently at different maturity levels. Such efforts include the development and maintenance of international data bases/repositories for e·Infrastructures-related components (e.g. GEANT/TERENA compendium, e-IRG knowledge base, CEENGINE knowledge base, EuroRIs-Net+ knowledge repository, TOP500 supercomputers, etc), carrying out of surveys and assessments (e.g. RI-IMPACT study, ERINA+ assessment, e-FISCAL study, etc) and other initiatives.

Among these initiatives, a novel approach that responded to the e·Infrastructures stakeholders' need, is the e·nventory project<sup>[4]</sup> that established the foundations for the roll-out of the first deployment phase of the **European e·Infrastructures Observatory**; an on-line platform, offering interactive, user-driven visualisation tools and an extensive set of benchmarking indicators to:

- facilitate multidimensional and polymorphic monitoring and analysis,
- support fact-based policy and learning and
- disseminate achievements of e·Infrastructures in Europe and beyond.

The European e·Infrastructures Observatory has already reached a fully functional state and is freely available to the public at large ([www.e-Observatory.eu](http://www.e-Observatory.eu), [www.enventory.eu](http://www.enventory.eu)); the e·Observatory:

- offers **7** intuitive, interactive and user-friendly visualisation tools,
- features a core set of more than **45** benchmarking indicators for monitoring e·Infrastructures development progress, usage, impact and investments,
- provides access to more than **10.000** individual figures and other related data sets, and
- consolidates more than **18** months of stakeholders' feedback and consensus on its structure and functionality.

## The challenges

Through consultation with key e-Infrastructures stakeholders, a series of recommendations have been collected as feedback towards addressing current gaps and hurdles and evolving the state of affairs towards a strategic, sustainable and holistic framework for keeping track of e-Infrastructures development over time and global regions of interest.

The challenges currently faced include the following noteworthy items that are currently missing:

- A **periodic, project-based assessment** that can be aggregated at overall EU Programme level.
- A holistic approach for consistently applying **data validation/curation** in order to ensure that the collected data accurately represent the e-Infrastructures status quo.
- A set of data points that would enable to cover **regions beyond EU** with the same monitoring framework thus allowing for cross-comparisons of EU e-Infrastructures to international ones.
- A set of indicators and benchmarks to monitor **emerging e-Infrastructures** such as data infrastructures, cloud, as well as other areas of interest.
- A universal framework for monitoring e-Infrastructures impacts and trends at EU/MS (and beyond) level and revealing interrelations to **macroscopic socio-economic indicators** by utilising “factual”, unbiased metrics and data without subjective interpretations.
- A **collaboration framework** that would be able to accommodate the views of owners, users, and sponsors of e-Infrastructures at both EU and National level.

Furthermore, as far as specific classes of indicators are concerned for monitoring e-Infrastructures:

- on the **networking** side, matters are quite well-addressed owing to the maturity that has been reached all over Europe for NRENs after a sufficient period of time and a consecutive series of flagship EU projects (currently in its GN3 implementation), coupled with the presence of a dedicated data collection mechanism which is carried out through, well-established and acknowledged related organisations (DANTE & TERENA). The GEANT/TERENA Compendium is responsible for a substantial amount of available indicators that exhibit a high degree of reliability in terms of time-span and geographical coverage.
- on the **grid** computing side, things seem to be on a transitional period, with several sources of data available containing a significant range of indicators, data and statistics, but sometimes lacking consistency in terms of time-span and geographical coverage (when compared to what the TERENA compendium has achieved). A one-stop-shop convergence can be expected to take place soon through EGI.eu and EGI-InSPIRE. This should lead towards a single, quantitative and qualitative point of reference for Grid progress monitoring (and potentially for clouds) that will enable a unified and seamless access to the Grid state-of-play to be made available.
- On the **supercomputing** side, the respective data aggregation framework is still at its early stages as far as an EU-hosted initiative is concerned. Besides the top500.org initiative, there has been little progress from the European side in terms of setting up a data repository mechanism. It is anticipated that the following phases of the flagship PRACE project could lead such an initiative and could help develop a single data collecting mechanism for supercomputing infrastructures.
- On other emerging e-Infrastructures (e.g. data infrastructures, cloud computing) both data aggregation as well as appropriate indicators, are still at their infancy, if not inexistent.

Lastly, the aspects of **time-span** and **geographic coverage**, even for “mature” indicators, are quite challenging to address in an all-inclusive way: in several cases the time period for which data is available may be limited, with partial data being available for some (usually earlier) years; likewise, geographic coverage of collected data, even when it addresses in principle all regions and countries, several incomplete cases may be included in the whole list. Combined, time-span and geographic coverage gaps, make it challenging to obtain complete and consistent data for all observed indicators.

## The roadmap

The results achieved to this date have generated awareness among e-Infrastructures stakeholders with regard to using as well as developing further the e-Observatory. Stakeholders that have declared their interest in e-Observatory developments include flagship EU initiatives such as GEANT, PRACE, EGI, EUDAT, OpenAIRE, as well as global e-science frontrunners from the USA, Russia, China, Australia, South Africa, Taiwan and Latin America. In order to take things further and conceive a pan-European strategy for a sustainable e-Infrastructures monitoring framework,

### *it is recommended that*

1. **The Commission and e-Infrastructure policy bodies** shall explore possibilities how to support a holistic framework for monitoring e-Infrastructures development, capitalising on and expanding the results of the European e-Infrastructures Observatory, to:
  - **Cover global regions** of interest where significant e-Infrastructures actions are taking place and/or international cooperation initiatives have been supported by the EU;
  - **Address emerging e-Infrastructures** and related initiatives that will be key enablers of the European Digital Agenda 2020, such as data infrastructures, cloud computing and broadband access initiatives;
  - **Standardise and automate data harvesting** and repository processes, in a manner interoperable with existing data repositories, in order to support updating as new data becomes available at the data source;
  - **Deploy novel visualisation tools and techniques** for representing the development of new e-Infrastructures themes, revealing Pan-European as well as Trans-Continental collaborations and impacts of e-Infrastructures that are key enabling pillars of international e-Science.
2. **The Commission** shall facilitate consultation and consensus processes at a global level towards the establishment of a global e-Infrastructures Observatory via the engagement of international e-Infrastructures experts, representing both regions of interest as well as themes of interest.
3. **Running flagship e-Infrastructure initiatives/projects** shall strengthen the collection and aggregation of data values for selected indicators, through a systematic and well-defined framework, for each of the e-Infrastructure fields (networking, DCI, HPC, etc.) by the organisations/structures that coordinate the respective initiatives (e.g. DANTE/TERENA, EGI.eu, PRACE AISBL, etc); wherever possible this process should build up a time series including past events.
4. **The Commission and National Governments** shall pursue a holistic approach for consistently applying data validation/curation to ensure that the collected data accurately represent the e-Infrastructures status quo.
5. **The Commission** shall ensure that future flagship e-Infrastructure initiatives/projects incorporate appropriate monitoring and assessment procedures according to relevant indicators suitable for the long-term e-Infrastructures assessment and benchmarking, which if not in the project proposal should be incorporated into projects at the project negotiation phase.
6. **The Commission** shall pursue, in close cooperation with **National Governments**, its efforts to support the maintenance and update of a European e-Infrastructures Observatory to keep up-to-date with the on-going developments in this field at international level.
7. **Major e-Infrastructure stakeholders**<sup>[5]</sup> shall facilitate, in close cooperation with **pan-European coordinating organizations and national bodies**, the collection of indicators in two emerging, important aspects of the development of European e-Infrastructures: sustainability<sup>[6]</sup> and human capacity<sup>[7]</sup> of e-Infrastructures. In neither case is there any information gathered consistently at present, if at all.

---

[1] Community Support for Research Infrastructures: evaluation of pertinence and impact, study report, commissioned by DG Research and DG Information Society and Media, September 2009.

[2] Role of Research Infrastructures for a Competitive Knowledge Economy, DG Research - Directorate B - Research Infrastructures and ESFRI – European Strategy Forum on Research Infrastructures, Workshop proceedings, 29-30 June 2009

[3] Research Infrastructures Work Programme 2010, European Commission C(2009)5905 of 29 July 2009, page 21.

[4] Seventh Framework Programme, Research Infrastructures, Grant Agreement no RI-261554 ([www.enventory.eu](http://www.enventory.eu)).

[5] **European and National state institutions** (e.g. European Commission, European Parliament, National governments, etc.) that are sponsoring electronic infrastructure initiatives and place a high value in their policy agenda on the impact assessment, post-mortem analysis and ex-ante evaluation, to help design better and more successful future programmes; **Policy bodies** (e.g. e-IRG, e-IPF, etc.) that have a clear mandate of supporting the development and the sustainability of electronic infrastructures and need an impact assessment tool to monitor the progress of achieving their objectives; **Flagship e-Infrastructures Projects** (e.g. GEANT, PRACE, EGI, EUDAT, OpenAIRE) that carry out specific electronic infrastructure activities and need to monitor their impact during deployment as well as retrospectively.

[6] As far as sustainability is concerned, it refers to the whole-of-life carbon footprint, totalling the cost of creating, running and decommissioning the e-Infrastructure components. The opportunities for energy savings across e-Infrastructures are huge and of considerable interest to the EU, as we strive for more environmentally sustainable approaches to all activities. Indicators that could be useful are ones such as the “total energy consumption” of national e-infrastructure components (possibly broken down by network, data centre and HPC centre). This figure would be valuable in its own right, to demonstrate trends in more efficient use of infrastructure, but would also be extremely helpful when combined with some of the other indicators, so that the cost per unit of installed disk space or the cost per process cycle of HPC systems in different countries could be calculated and displayed. Overall, each country should be looking to reduce these numbers and ratios in all instances.

[7] As far as **human capacity** is concerned, respective indicators are, perhaps, more speculative and difficult to measure, however, given that EU is interested in increasing the capacity of Europe to utilise new technologies and innovation, a set of metrics that indicates where and how much of this activity is taking place across the different countries would be valuable. Such types of measures could include, e.g. the number of university graduate or post-graduate courses in environmental sustainability for IT operations and in programming for energy efficiency, or new professions who understand in great detail e-Infrastructure systems and their exploitation and who can undertake the programming and scheduling activities required to carry out research most effectively, leaving the researchers themselves to concentrate on the research issues. Such indicators, even if simply numeric at the moment, could be extremely valuable in longitudinal studies, showing how human capital is being enabled and expanded across the EU.

The European e-Infrastructures Observatory has been developed by the **e-nventory** project, co-financed by EU's 7th Framework programme for R&D under Grant Agreement no RI-261554.



The **e-nventory** project acknowledges the invaluable support and contribution of its [Advisory Board](#), and the following European flagship e-Infrastructures initiatives and coordinating organisations.

