**[Which e-infrastructures for tomorrow's environmental challenges?](https://ec.europa.eu/futurium/en/content/which-e-infrastructures-tomorrows-environmental-challenges)**

**Discussion and contributions received on Digital4Science**

The European Commission would like to explore synergies and identify gaps in the activities engaged by the e-infrastructures projects on **Earth sciences, energy and environment**, as well as to share knowledge and increase cooperation by seeking interoperable solutions, re-using promising applications and technologies, sharing computing resources, organising joint pilots and demos, and exploring common business and sustainability perspectives.

Questions to be addressed include:

* which e-infrastructure innovations and services need to be pursued in order to advance priority needs of the individual scientific domains and to maximise their societal impact?
* which e-infrastructures can be re-used by other domains?
* what needs to be done for their uptake and sustainable deployment?

Reflections concerning integrated e-infrastructure services (IT as a service to research) to enable data and compute intensive science in multiple research fields, the transformation of science (digital, data-driven, open), the evolution towards an open science cloud, etc. are welcomed.

[**Common requirements for future digital infrastructures**](https://ec.europa.eu/futurium/en/comment/5739#comment-5739) **– Donatella Castelli**

Below a set of common requirements for future digital infrastructures that emerged during the pre-RDA plenary workshop on "Research Data Infrastructures for Environmental-related challenges" held on 22 September 2015, in Paris. The workshop was coordinated by Donatella Castelli (CNR), Odile Hologne (INRA), and Michel Schouppe (European Commission - DG RTD). The presentations done in this workshop are accessible at: <https://rd-alliance.org/plenary-meetings/sixth-plenary/programme/research-data-infrastructures-environmental-related>.

• The three addressed challenges (i.e. climate change, marine resources conservation, food security) do face bigger volumes of data, higher diversity, and heterogeneity of formats, metadata structures and contents. They require multi-scale approaches with abilities to scale up and down from global to local scales.

• The three challenges point towards data infrastructures that should better integrate monitoring data, numerical model outputs, and socio-economic data. This is a pre-requisite in order to generate societal impact.

• All challenges are characterised by a growing interest in data provision, transparency and reproducibility. Permanent focus should be put on quality check, comparability, and validation of data and information. Data should be collected once and reused several times; proper licensing is not yet implemented systematically.

• Environmental-related challenges are trans-disciplinary; they require collaboration and co-design. Data and systems interoperability are essential if we want to create shared knowledge across different communities. Standardization of metadata and the use a generic ontology require stronger collaborations between scientists and data managers. Semantic web and big data technologies require expertise and manpower in informatics and data base management. Open data policies are encouraged at international (e.g. G8, G20, UN frameworks), European and national levels. Still the privacy issue matters a lot, for instance when geo-localised information on farms and crops may lead to identify the farmer. There is a lot to learn from anonymisation techniques used in other fields (e.g. in the health sector). Dealing properly with private issues will also require more of institutional commitment.

• Some researchers and organisations are still resistant to share their data openly or to document sufficiently data quality aspects of their data. More incentives are needed from research funding agencies in order to change behaviours. Open data sharing will be also facilitated by a wider deployment of better solutions for data traceability, attribution and citation. Initiatives would gain from being more user-centric than data-centric.

• Access to data for commercial use will require defining new economic models and partnership policies for data infrastructures. New agile players are emerging, notably in the US. Access to information for innovation will be greatly facilitated from new enriching data discovery capabilities on the web, for instance based on brokering technologies.

• Four pillars have been identified in order to build successful infrastructures:

* 1. (storage as service (e.g. geospatial data clouds); data curation and management (metadata generation, harmonization, exchange); data processing and analytics (e.g. enabled by cloud computing); and knowledge delivery (e.g. data publication);
	2. brokering approaches to cope with diversity, coupled to continued efforts to agree on internationally-recognised standards and methodologies,
	3. policies (such as on data usage, infrastructure governance, resource and service inclusion), capacity development and advocacy, and
	4. alignment of national initiatives with international ones.

• When considering budget allocation, many institutions have the tendency to privilege their core missions sometimes at the expense of investing properly into metadata generation and infrastructures for data sharing. Data management is not cheap. Data sharing has a cost. There is a strong demand for best practices and solutions allowing reducing these costs while maximising data impact.

• Sustainability of research data infrastructures is a recurrent problem. In this period of public austerity, there is need for new economic models sustaining long term operation of these infrastructures. There is also a difficulty related to the transitioning of research to permanent data infrastructures. It is hard to attract more users to contribute to infrastructure development while not providing a fully operational service yet to those users.

#### [**Continued concertation and engagement with related projects**](https://ec.europa.eu/futurium/en/comment/5749#comment-5749) **– Wouter Los**

In our meeting of Monday 9 November 2015 in Brussels, I suggested to continue with interacting with a focus on (a) how we can benefit from each other for selected short term problems to be addressed, and (b) on clarifying common challenges to be explained in a common white paper.
(With respect to the latter I think that sharing and re-use of data, models, web-services and workflows would be an excellent topic for studying complexity in (interacting) environmental systems).

There was support to have also other relevant projects, such as the ENVRIPLUS project, involved ([http://www.envriplus.eu](http://www.envriplus.eu/)). Next week will be a project meeting of ENVRIPLUS in Prague and I will consult the key individuals there for organising a joint meeting.
Let us start discussing how we can boil down the magnitude of possible meeting topics to a few crucial ones, such as mentioned above under a and b.

[**Summary of discussions - Group on Earth, Energy, Environment**](https://ec.europa.eu/futurium/en/comment/5750#comment-5750) **– Angelo Rossi**

**e-concertation 9 Nov. 2015: Summary of discussions of Project Group on Earth Sciences, Energy and Environment**

The 11th e-concertation meeting held on November 9th 2016 in Brussels included a breakout meeting of the project group on Earth Sciences, Energy and Environment, including representatives from several H2020 projects (see Annex 1). Original materials are included (see Annex 2).

The project group brings together e-infrastructure projects that stem from calls on Virtual Research Environments and on HPC Centres of Excellence, which address a variety of topics including: an energy-related centre of excellence (EoCoE); weather and climate simulations in high performance computing (ESiWACE); analytics on big earth data cubes (EarthServer-2); earth science data monitoring (EVER-EST); a data e-infrastructure for marine and fisheries (BlueBRIDGE); use cases on climatology (VI-SEEM); and support to multi-disciplinary data-driven sciences (VRE4EIC). This thematic project portfolio is complemented by support from transversal e-infrastructure projects including: data knowledge and digital services for open science (EGI Engage); the European partnership for high performance computing services (PRACE); a collaborative data e-infrastructure (Eudat2020); the European research and education networking commons (GeANT); digital identifiers for scientific artefacts and researchers (Thor); and text and data mining including applications on use cases in life sciences, food and agriculture (OpenMinTed). In the meeting participated also representatives from other projects such as EDISON, MAGIC, TANDEM, SESAME-NET and the FP7 project ODIP.

The mandate of the group is to explore synergies between the projects as well as to share knowledge and increase cooperation by seeking interoperable solutions, re-using promising applications and technologies, sharing computing resources, organising joint pilots and demos, and exploring common business and sustainability perspectives.

The discussion was triggered by 3 questions, which were posted in the Digital4Science platform. The present report compiles the main issues addressed, although not necessarily in the order of discussion.

* + Question 1) Which e-infrastructures (innovations and services) are needed to cover priority needs of the individual thematic domains?
	+ Question 2) What possibilities exist for re-use of these e-infrastructures by other domains/disciplines?
	+ Question 3) What needs to be done for the uptake of existing services and their sustainability and maintenance?

**Answers to Question 1 -** Different thematic domains are characterised by different needs, some more technology-driven, some primarily science-driven. Different layers or levels exist, represented by e-infrastructures, research infrastructures and virtual research environments (VRE). A key need is to bring together both technology and domain/research actors. Some sub-communities have different definition or perception of data (e.g. the larger Geo community is used to standards in mapping, while other communities could benefit from higher-level abstraction to access relevant data). This is particularly relevant for multi/cross/trans-disciplinary projects and approach. One could bring data to users, but on the other hand, users can be also brought to (access) data, including availability of processing, tools and overall awareness of available data services (catalogue of services, discussed later in plenary sessions). A user-centered vision is considered important, and some infrastructure or services could be very specific, other more generic. In this respect, the importance of data quality and potential editorial processes on data themselves has been stressed. Some common grounds, across domains and service types, exist: in this respect, the publishing workflow/framework is common to the overall research/technology communities.

**Answers to Question 2** - Co-design was stressed to be important in upcoming WP 2016/2017. Group discussion stressed the necessity of co-understanding before co-design. Achieving interoperability across systems, as well as brokering and integration with existing ones, is considered important. The possibility of harvesting across services need to be based on standards (e.g. Open Geospatial Consortium - OGC W\*S - [http://www.opengeospatial.org](http://www.opengeospatial.org/) ), as well as workflows. In order to re-use or jointly use services and data, sharing and communication across platforms, stakeholders and countries/geographic regions is needed. Some existing or upcoming projects could offer services to actors active in other directorates.

**Answers to Question 3** - The difficulty for several players to achieve sustainability of services was expressed. Domains do need sustainability and expect high-availability of services, particularly when it comes to data archives and data access services. The basic question of the business case and model (including who is the paying customer) in science and model was raised (discussed also in EUDAT). Basic infrastructures need to obey to community (and formal) standards, or at least, to start with, to be aware of them, and corresponding policies should be developed. One could learn from existing isolated services and efforts that were developed until now: they and their assessment will be beneficial for developing new platforms and services. Somewhat similarly to the generic publishing workflow (see Q1 discussion), basic resources can be cross-cutting, while specific tailor-made services, possibly building on them are more specific. Other issues: An aspect that was not discussed but participants felt to be relevant and important is the use of Open Government Data platforms: data served from them can be valuable for researchers, and possibly used or linked to services of various kinds, such as those related to geo-hazards.

**Actions**: An updated mailing list, including the additional project representatives that participated to the meeting, will be posted in the D4S platform. The platform (D4S) should be used more systematically for identifying specific issues for closer cooperation between the projects. In this respect, a matrix of use case (science) per technology would be set up to facilitate the exploration of additional potential points of contact, interaction, and cross-fertilisation. In addition, specific joint meetings (face2face) could be considered to stimulate interest in exploring synergies and enhance cooperation. It was first proposed to establish a link with the project ENVRIplus of the Research Infrastructure programme and to explore the possibility of a joint meeting.

Reporting by Angelo Pio Rossi and Reinhard Budich

**Annex 1 - Participants** - CASTELLI Donatella (CNR-ISTI) ROSSI Angelo Pio (Jacobs University Bremen) LOS Wouter (University of Amsterdam) BUDICH Reinhard (MPI für Meteorologie) LECARPENTIER Damien (CSC - IT Center for Science) ALBANI Mirko (ESA) STEIJAERT Andres (SURFnet) CHEVERS John (GÉANT) SALOMONI Davide (INFN CNAF) UTRERAS Florencio (Red CLARA) GLAVES Helen (British Geological Survey) DESPLAT Jc (Irish Centre for High-End Computing - ICHEC) LAZAROU Constantinos (The Cyprus Institute) JEFFERY Keith (Keith G Jeffery Consultants) MATYSKA Ludek (Masaryk University) DIEPENBROEK Michael (University Bremen) ALLINE Damien (IRD) PINEDA Oriol (PRACE AISBL) Chaired by Antonios Barbas (EC) Jean-Luc Dorel (EC)

**Annex 2 - Original notes -** Original notes are pasted below. They are as well present on the etherpad used during the breakout meeting: <https://etherpad.gwdg.de/p/ESEE-Meeting2015-11-09> 1)

* + Which e-infrastructures are needed to cover priority needs o the thematic domains? different layers: e-infra + e-research infra + vre heterogeneity of end users needs (technology vs. domain science expertise) bring technology and domain/research together (some) end users need map data, non-file (facilitate of use data without getting dirty hands on files) --> level of services to deal with data abstraction vs. perception on information in data need for quality of data (editorial process on data) multi/cross-disciplinary approach(es) growing own (project/partners) computing resources vs. additiona/external resources, e.g. grid user-view --> user-centered vision Bring data to user vs. bring user to data (and processing / tools) discipline-specific vs. generic infrastructures how generic should e-infra be? model from passing from local to e-infra resources analytics: interface for easy run analytics on data --> standards + data (not necessarily in the same providers) common grounds: publishing workflow/framework (generic, for any domain)
	+ what possibilities for re-use by other disciplines? co-understanding needs before cooperating / co-design Interoperability across systems / brokering / integration with existing systems harvesting metadata across services? standards W\*S services (WPS, WCS, WCPS) workflows (workflows across services) Sharing & communicating across platforms and countries (linked to generic from point #1) data and computing Relationship with projects from other DG ?? What can we offer? (see actions)
	+ Status of uptake of existing services and sustainability? how can they be maintained? community expects high-availability services --> data archives domains need sustainability - difficult to fund business case in science? why? developing services for what? e.g. to do computing? how paid? cost issue: It's a matter of the business model, also discussed in EUDAT Basic infrastructures need to obey to community standards, or at least be aware of them develop corresponding policies learn from existing isolated services/efforts up to now --> new platforms/services polarisation of funding between science & infrastructure --> basic resources are cross-cutting + community tailor-made services.
	+ Additional questions to be eventually addressed for future discussion? Use of Open government data platforms --> useful for researchers --> how to link it? e.g. geohazards Actions DONE make a synthesis of today's discussion on the platform (based on email iteration)
	+ OPEN: PLEASE ADD YOUR REMARKS -->editing on <https://etherpad.gwdg.de/p/ESEE-Meeting2015-11-09>) / Who: ALL When: Monday 16th November 2015 noon CET OPEN: identify and amend to the present cluster of co-funded h2020 project (open, more can join) Who: ALL When: Continuously OPEN; d4s: identify issues for longer time perspective, communicate through the platform + eventual synchronous media Who: ALL When: Continuously OPEN: plan a real f2f meeting (enlarged?) Who: ALL by email list + d4s When: end Q1 or Q2 2016? (doodle to come) DONE: produce a joint statement (see here) OPEN: act as connection/ambassador across groups/projects to increase awareness Who: ALL When: Continuously OPEN: ENVRI+ establish a link between this group (cluster), and explore possibilities for specific actions Who: Wouter Los + Helen Glaves When: Prague meeting week starting Nov. 16th 2015 OPEN: look at existing use cases vs. proposals of techology per sector Who: ALL, report to Reinhard Budich (reinhard.budich@mpimet.mpi.de) When: Continuously.

**[Mailing List established](https://ec.europa.eu/futurium/en/comment/5752%22%20%5Cl%20%22comment-5752)**

Mailing list ESEE-SIG@GWDG.DE, and invited the following people:

donatella.castelli@isti.cnr.it
p.baumann@jacobs-university.de
an.rossi@jacobs-university.de
tiziana.ferrari@egi.eu
edouard.audit@cea.fr
fieg@dkrz.de
biercamp@dkrz.de
reinhard.budich@mpimet.mpg.de
kimmo.niittuaho@csc.fi
Damien.Lecarpentier@csc.fi
mirko.albani@esa.int
john.chevers@geant.org
dale.robertson@geant.org
yannis@di.uoa.gr
adam.farquhar@bl.uk
mdiepenbroek@pangaea.de
oprnjat@admin.grnet.gr
philippe.rohou@ercim.eu
keith.jeffery@keithgjefferyconsultants.co.uk

Please spread the word to those I forgot/did not know about. Thanks.
Mit freundlichen Grüßen R. Budich