

E-Infrastructures & Data Management Collaborative Research Action (E-Infra CRA)

A Belmont Forum Initiative

Jean-Pierre Vilotte¹, Jean-Yves Berthou², Patrick Monfray³, Mathieu Girerd⁴

Aim and Scope of E-Infra CRA

The Belmont Forum⁵, a coalition of national science funding agencies from 13 countries⁶, is supporting an 18-month Collaborative Research Action (CRA) on *E-Infrastructures and Data Management*⁷ to implement a 'Knowledge Hub' community-building and strategy development program. The CRA will address the Belmont Challenge priorities, i.e., societally relevant global environmental change challenges. This effort represents an important step to coordinate and streamline international efforts on community governance, interoperability and system architectures, so that environmental data and information can be exchanged internationally and across subject domains easily and efficiently. The Knowledge Hub builds on existing programs and consortia, bringing together the leading national natural and social scientists, users, and data and computational scientists that are utilising advanced cyber-infrastructure for environmental data management, modelling and services. The project will deliver a community-owned Strategy and Implementation Plan, which will provide recommendations to prioritize international funding opportunities for Belmont Forum members to build pilots and exemplars in order to accelerate delivery of end-to-end global change decision support systems.

¹ Institute of Earth Physics of Paris, representative of the French National Research Agency in the Steering Committee of E-Infra CRA

² Head of ICT Department of the French National Research Agency (ANR), Representative of ANR for E-Infra CRA

³ Deputy Head of Environment and Biological Resources Department of the French National Research Agency (ANR), Co-chair of Belmont Forum

⁴ Program Officer, ICT Department of the French National Research Agency (ANR),

⁵ Belmont Forum website: <http://igfagcr.org/index.php/belmont-forum>

⁶ Belmont Forum members: <http://igfagcr.org/index.php/bf-members>

⁷ E-Infra CRA website: <http://bfe-inf.org/>

This Knowledge Hub community serves as a searchable, frequently-updated digital communication centre to allow participants to share ideas, collaborate to achieve project goals, and connect with researchers around the world. The Knowledge Hub is guided by a Steering Committee, comprised of a member from each participating Belmont Forum country and organization. Steering Committee members are responsible for leading one or more work packages to collectively assess existing international E-infrastructure capabilities, gaps and overlaps, prioritize challenges, and provide recommendations on how to best address the Belmont Challenge⁸, focusing on the areas of standards (WP1), computation/data infrastructure (WP2), harmonization of global data infrastructure (WP3), data sharing (WP4), open data (WP5), and capacity building (over-arching) (WP6). A National Delegation Assembly is composed of approximately 10 researchers and users from each Belmont Forum nation, each of whom participate in scoping activities via the work packages.

The draft work programme of the six work packages is presented below.

Work Programme (October 2013 – March 2015)

The majority of the investigations, research and other activities as part of the E-Infrastructures and Data Management CRA are being carried out over 12 months by a series of six work packages, each of which is led by one or several Steering Committee members. A 3-day Steering Committee meeting was held in Windsor, United Kingdom, October 15-17, 2013, to identify and determine leadership of WPs in the areas of Governance, Architecture, and Interoperability. The WPs are divided into the following areas of focus:

1. Standards
2. Improved interface between the computation and data infrastructures
3. Harmonization of global data infrastructure for sharing environmental data
4. Data Sharing
5. Open Data
6. Capacity-Building (over-arching activity)

Members of the National Assembly Delegations (approximately 10 people from each Belmont Forum member country or organization) populate the WPs. WP activities will culminate in a Community Strategy and Implementation Plan, with the intent to provide recommendations on how to best address the Belmont Challenge and identify where future activities might most productively be steered to promote better e-infrastructure and data management practices that can accelerate delivery of end-to end global change decision support systems. The final Plan will be delivered to the Belmont Forum in early 2015.

WP 1 - Standards for Sustainable Global Research

WP leaders: Roberto Cossu (ESA/ESRIN) and Mustapha Mokrane (ICSU)

⁸ Belmont Challenge : <http://igfagcr.org/index.php/challenge>

The WP aims at enabling data integration of heterogeneous multidisciplinary datasets (including datasets from social sciences and humanities), the development of specialized thematic platforms allowing data manipulation and visualization, reuse of data and derived data such as models, and at ensuring credibility of the derived research findings.

This WP will address the following areas: ontologies and semantics; discoverability and annotation; data provenance identifiers; minimum standards for interoperability and reuse. The WP will work in synergy with the Future Earth initiative.

WP2 - Improved Interface between the Computation and Data Infrastructure

WP leaders: Jean-Pierre Vilotte (Institute of Earth Physics of Paris) and Toshio Koike (University of Tokyo)

The WP aims at fostering interaction between users and data and computation e-infrastructures providers; and at providing recommendations and implementation strategy plan for R&D initiatives, integrated e-infrastructures services and support to orchestrated data-intensive and extreme-computing workflows. It narrows down the targeting issues through a clear separation between community building, high-level strategy and coordination, service provision and standardization, innovation

This work package should address the following areas: enable the exploitation and the valorization of massive and complex data generated by high-throughput instruments, observational and monitoring systems; bringing together users and experts on data and HPC infrastructures; good practices and potential multi-disciplinary use cases or proxy mini-apps; existing synergistic gaps and barriers between data and HPC infrastructures in support of orchestrated data-intensive and extreme-computing workflows; barriers and gaps in innovative tools and methods for Big Data analytics; barriers and gaps in integrated services between data and compute infrastructures including data life-cycles; user-driven performance and quality indicators for Data and HPC infrastructures interface.

WP3 - Harmonization of Global Data Infrastructure for Sharing Environmental Data

WP leader: Christoph Waldmann (Centre for Marine Environmental Sciences)

The WP aims at fostering interaction of data infrastructures and identifying minimal requirements for interoperability to enable cooperation across disciplines in the framework of environmental sciences, identifying synergies with Future Earth, RDA and GEO/GEOSS.

This WP should address the following area: priority actions to foster data sharing, establish data policy principles, data provenance and identifiers, and ensure adequate data archiving/management structures.

WP4 – Data Sharing

WP leaders: Dale Peters (University of KwaZulu-Natal) & Andrew Treloar (Australian National Data Service)

The WP aims at identifying the values and incentives that encourage the deposit and sharing of data; including the user perception of trust in data curation practice, infrastructure management, and organizational continuity; as well as the legal issues pertaining to the protection of data privacy and the IP rights of data owners.

This WP should address the following areas: data authenticity and quality; data provenance (certification, standards); incentives to share data; cultural issues; trust across spectrum of sources; preservation of data context; legal issues (including liability); long term sustainability and organisational continuity; community approach to standards.

WP5 – Open Data

WP leaders: Andrew Treloar (Australian National Data Service) and Birgit Gemeinholzer (Justus Liebig University Giessen)

The WP aims at developing a strategy to promote/implement open data across groups with an interest in environmental data who are at various stages of technical maturity. The WP also ensures that all the data providers into Future Earth have a secure and sustainable location into which they can lodge their data for access and use by others; and enable research communities to draw on work done by civil society/crowdsourcing, and enable civil society to analyse results generated by research community.

This WP should address the following areas: open access data availability and data use; integration between different scientific research communities and the civil society; strategies to promote/implement open data across groups with an interest in environmental data who are at various stages of technical maturity; match-making services linking data providers with suitable repositories to support open data; uses of open data, liability/uncertainty in decision making.

WP6 – Capacity Building

WP leaders: Lee Allison (Arizona Geological Survey) and Robert Gurney (University of Reading)

The WP aims at developing a common understanding and protocols about building capability in training, security issues and legal frameworks. The WP works closely with other WPs to identify where they may require support in these areas.

This WP should address the following areas: resource gaps, including data infrastructure, legal and security issues; cross-disciplinary education and training and cross WP issues identified as work progresses; sustainable human resources.