A funder's approach to more open data and better data management

Belmont Forum e-Infrastructures & Data Management Project

The Belmont Forum partnership of funding organizations, and international and regional science councils, is committed to accelerating open-data sharing and reuse by improving researchers' data-management practices, solving e-infrastructure challenges and improving the data skills of global environmental-change scientists.

The Belmont Forum¹ is a partnership of national science funding organizations, international science councils, and regional consortia across the world committed to the advancement of global environmental science (Fig. 1). The partnership aims to accelerate delivery of data-driven environmental research to remove critical barriers to sustainability by aligning and mobilizing international resources.

The Belmont Forum activities are driven by the Belmont Challenge² that encourages international transdisciplinary research to provide knowledge for understanding, mitigating and adapting to global environmental change. The Belmont Forum supports multinational and transdisciplinary collaborative research through Collaborative Research Actions (CRAs)³, bringing together natural sciences, social sciences and the humanities, as well as stakeholders, to co-create knowledge and solutions for sustainable development.

Global environmental-change research increasingly requires integrating large amounts of diverse data across scientific disciplines to deliver the policy-relevant and decision-focused knowledge that societies require to respond and adapt to global environmental change and extreme hazards, to manage natural resources responsibly, to grow our economies, and to limit or even escape the effects of poverty. To carry out this research, data need to be discoverable, accessible, usable, curated, and preserved for the long term. This needs to be done within a supporting data-intensive e-infrastructure framework that enables data exploitation, and that evolves in response to research needs and technological innovation. Without open data and the supporting e-infrastructure, policy makers and scientists will be forced to feel their way into the future without the benefit of new scientific understanding; unfocused and ill-prepared.

To accelerate the openness, accessibility and reuse of data from CRA projects, the Belmont Forum adopted an Open Data Policy and Principles⁴ to stimulate new approaches to the collection, reuse, analysis, validation, and management of data, digital outputs and information, thus increasing the transparency of the research process and robustness of the results. In 2015, the Forum established the e-Infrastructures & Data Management (e-I&DM) Project⁵ to help implement the Open Data Policy and reduce barriers to data sharing and interoperability. e-I&DM is promulgating procedures, standards, workflows, and other elements critical to identifying a path toward cooperative e-infrastructures and data-management policies and practices that enable and accelerate open access to, and reuse of, transdisciplinary research data.

Implementing data management for openness and reuse

The Belmont Forum is gradually implementing its Open Data Policy through its CRA funding process. All CRA calls now require a data management plan (Data and Digital Outputs Management Annex⁶) to ensure that project teams will meet both the Open Data Policy and Principles and the Force11 FAIR (Findable, Accessible, Interoperable and Reproducible) Data Principles⁷, and adhere to relevant standards and community best practices. Belmont Forum researchers must consider data-management issues from the inception of a project in order to plan and budget appropriately for data curation, management and sharing. Data-management plans should also comply with public-access

policies and applicable national laws of the respective funding agencies supporting CRA awards.

Research data and digital outputs are expected to be open by default and publicly accessible, possibly after a short period of exclusivity, unless there are legitimate reasons to constrain access. Data and digital outputs must be discoverable through machine-readable catalogues, information systems and search engines. A full Data and Digital Outputs Management Plan for an awarded Belmont Forum project is expected to be a living, actively updated document that describes the data-management lifecycle for the data and other digital outputs collected, processed, or reused.

A related e-I&DM initiative is a collaboration between Belmont Forum funding agencies and science publishers to articulate a coherent set of data and digital-outputsmanagement expectations for published research, with the ultimate result of improved sharing and data reuse. Now approved by the Belmont Forum Plenary, the Data Accessibility Statement language will be incorporated into the Data and Digital



Figure 1: Belmont Forum: An International Partnership of Funding Agencies and Science Councils.

Key SEI CRA features: implementation



Figure 2: A "Task Force" will steer the SEI CRA implementation in synergy with the Belmont Forum e-Infrastructures & Data Management Project, continuously monitoring the progress and activities of the funded projects through regular workshops, fostering collaboration and maximizing the outcomes across projects. This will contribute to a knowledge Hub catalyzing research efficiently through sharing of best practices, methods and software implementations, to allow delivery of recommendations and priorities to the Belmont Forum for transnational federated data e-infrastructures, data policies and capacity building.

Outputs Management Annex, so researchers will understand the end-to-end expectations of both funders and publishers regarding management of their research data to maximize openness, accessibility, and reuse.

Addressing the barriers to transnational data sharing and reuse

The capability is emerging to bring computer science and technology, as well as large and complex data sets, to bear on interdisciplinary and transdisciplinary science. It is therefore critically important to establish and enable transnational frameworks so that data-driven scientific knowledge can transcend both disciplinary and geographical borders, ultimately increasing the scientific underpinnings of policy and action. International collaboration within the Belmont Forum research priorities holds the potential to establish international foundations for federated data integration and analysis systems with shared services. It can also bring together best practices from the public and private sectors, foster open-data and open-science stewardship among the science communities, including related areas such as publishing, and encourage data and cloud providers and others to adopt common standards and practices for the benefit of all.

For these reasons, the Belmont Forum recently closed a four-year competitive call on Science-driven e-Infrastructure Innovation (SEI) for the Enhancement of Transnational, Interdisciplinary and Transdisciplinary Data Use in Environmental Change⁸. The SEI call will fund initiatives that bring together environmental, social, and economic scientists with data scientists, computational scientists, and e-infrastructure and cyberinfrastructure developers and providers to solve methodological, technological and/or procedural challenges currently facing interdisciplinary and transdisciplinary environmental-change research.

The SEI call is being implemented under a "task force" structure (Fig. 2) that requires all funded projects to share results, participate in annual steering workshops, and contribute to a knowledge hub that will catalyze efficient research through sharing of best practices, methods and software implementations. Information in the knowledge hub may also be used to deliver research-driven recommendations to the Belmont Forum to address needs or enhance current strategies for transnational federated data e-infrastructures, data policies and capacity building.

Building researchers' data skills

The Belmont Forum e-I&DM strategy document, 'A Place to Stand'⁹, recommended that a "cross-disciplinary training curriculum was required to expand human capacity in technology and data-intensive analysis methods for global change research" and that a new data literacy was required for the 21st century. Consequently, the e-I&DM Project developed the Data Skills Curricula Framework¹⁰, based on a global survey¹¹ (Skills Gap Analysis), data skills workshop¹² and extensive consultation with data-management experts and trainers.

The Curricula Framework outlines core modules to enhance the skills of domain scientists specifically to make data handling more efficient, research more reproducible and data more shareable - including visualizations for end-users. The five core skills comprise programing, particulars of environmental data, visualization, data management, and interdisciplinary data exchange. Further, a number of optional modules are suggested for more-established researchers as useful introductions to widen their data skills, such as machine learning and object-oriented programing. Two additional modules aim to provide Principal Investigators with an overview of data management and skills needed for open data.

Of the core curricula, the two skill areas addressed least by existing training are 'Environmental data: expectations and limitations' and 'Interdisciplinary data exchange'. Since materials on the former are likely to exist in university courses, 'Interdisciplinary data exchange' is the current focus of the Belmont Forum, to be taught in a mixed class of environmental scientists, social scientists and engineers.

To build on existing capabilities, e-I&DM is investigating the training activities currently available from Belmont Forum member agencies. In addition, e-I&DM is working closely with the data-science community to identify existing training offerings available from around the world and augment content and provision of courses as needed.

Taken as a whole, the Belmont Forum's focus on data management, e-infrastructures and data skills is a critical step forward in advancing open-data sharing, data accessibility and data reuse.

AFFILIATION

Belmont Forum e-Infrastructures and Data Management Project, Montevideo, Uruguay

CONTACT

c3o@bfe-inf.org

LINKS

¹belmontforum.org ²belmontforum.org/wp-content/uploads/2017/04/

belmont-challenge-white-paper.pdf ³bfgo.org ⁴belmontforum.org/about/

open-data-policy-and-principles/

⁵bfe-inf.org

- ⁶bfe-inf.org/sites/default/files/doc-repository/ CRA_Data_Digital_Outputs_Management_ Annex_20180501.pdf
- ⁷force11.org/group/fairprinciples
- ⁸bfgo.org/opportunity/index.jsp#sei2018
- [°]bfe-inf.org/sites/default/files/doc-repository/ Community_Edition_A_Place_to_Stand-Belmont_
 - Forum_E-Infrastructures_Data_Management_ CSIP_0.pdf
- ¹⁰bfe-inf.org/sites/default/files/doc-repository/Outline_ Data_Skills_Curricula_Framework.pdf
- ¹¹bfe-inf.org/sites/default/files/doc-repository/BF-Skills-Gap_Analysis-2017_0.pdf
- ¹²drive.google.com/file/d/0B-s5Of-9WNvBQd2d6ZnpxNXZIZDA/ view

