



Research Data for a Changing World

26th January 2016

Lee Allison, Arizona Geological Survey [1],

lee.allison@azgs.az.gov

Belmont Forum E-Infrastructures Secretariat [2]

Rowena Davis, Arizona Geological Survey [1]

Belmont Forum E-Infrastructures Secretariat [2]

Robert Gurney, University of Reading [3],

r.j.gurney@reading.ac.uk

Belmont Forum E-Infrastructures Secretariat [2]

Erica Key, Belmont Forum [4]

Kim Oakley, University of Reading [3]

Belmont Forum E-Infrastructures Secretariat [2]

“Why should we care? Because, just as the World Wide Web has transformed our lives and economies, so this new data wave will matter eventually to every one of us, scientist or not.”
- *The Data Harvest, RDA Europe, 2014*

1. Introduction

Global environmental change is one of the most pervasive concerns of the 21st century. Scientists throughout the world are undertaking research to determine the nature and extent of these changes, and their impacts on humans and the environment. Global change research enables scientists to understand and predict how our planet functions and evolves and to investigate responses to those changes. This 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 such data and the supporting e-infrastructure, policy makers and scientists will be forced to feel our way into the future without the benefit of new scientific understanding, unfocused and ill-prepared.

Environmental change is most rapidly occurring in the Arctic, where additional pressures for development and commerce are speeding the need for informed decision making. Data committees, such as the Arctic Data Committee of the Sustaining Arctic Observing Networks, are working quickly to craft policy for the Arctic that would provide measured access to information – “ethically open data” – while giving careful consideration to both the potential benefit and harm that data could cause to northern communities. The Arctic also leads the way for valuation of local knowledge and community or citizen science, creating opportunities for new data management approaches that connect users with knowledge holders.

Gaining support for these data initiatives requires the concerted efforts of many organizations involved in the funding, implementation, and evaluation of data and data management. Over the past two years, a team of domain experts and representatives from international data governance bodies came together under the Belmont Forum auspices to forward a set of global data principles. These principles target not only best practices for open access, but also communication, training, and stewardship to grow an informed community of data providers and users. These principles were adopted by the Forum in October 2015. Many of the members in the Forum hail from Arctic and Arctic-interested nations; thus, the implementation of these principles has the potential to meaningfully improve data transparency in all regions, including the Arctic.

The Belmont Forum seeks to establish a cooperative approach to developing **sustainable practices within the global change research community** for data discovery, management and curation. The goal is to streamline the dissemination of global environmental change information and maximize the opportunities for effective action.

1.1 About the Belmont Forum

Established in 2009, the Belmont Forum¹ comprises the world's major funding agencies of global environmental change research and also international science councils. It is guided by the Belmont Challenge, which aims: *"to deliver knowledge needed for action to avoid and adapt to detrimental environmental change including extreme hazardous events."*² The Belmont Forum serves as a round table for these agencies to address issues related to global environmental change collectively. To meet the goals of the Belmont Challenge, the Belmont Forum coordinates funding for Collaborative Research Actions (CRAs), which are high-priority research activities designed to improve the way funding agencies collaborate with each other and develop opportunities for multi-national research. Initially, priority focus has included Arctic observing and science for sustainability, coastal vulnerability, freshwater security, food security and land use change, mountains as sentinels of change, biodiversity and ecosystem services, and climate predictability.

1.2 E-Infrastructures and Data Management Collaborative Research Action

Accurate and reproducible science requires comprehensive and verifiable data that are appropriately documented and accessible. As researchers strive to understand the vast and varied systems comprising the global environment, a large factor determining their success is access to robust and reliable data as a foundation and reference point for their own conclusions. The Belmont Forum initiated the E-Infrastructures and Data Management CRA to collectively develop achievable and sustainable e-infrastructures and data management practices in recognition that:

"...the need to address global environmental challenges requires a more coordinated approach to the planning, implementation, and management of

¹ For more information on current membership, please visit <http://www.belmontforum.org>.

² The Belmont Challenge: A Global, Environmental Research Mission for Sustainability, (March 2011) <http://belmontforum.org/sites/default/files/documents/belmont-challenge-white-paper.pdf>

data, analytics and e-infrastructures through international collaboration.”

— Belmont Forum, New Delhi, February 2013

A report with priorities was one output from the CRA, and is now being implemented. This is the result of activities conducted over an 18-month period by an international Assembly of more than 120 domain scientists, computer scientists, information scientists, social scientists and legal scholars. Their task was to survey the state of current practices and establish recommendations on how the Belmont Forum can leverage existing resources and investments to foster a more coordinated, holistic and sustainable approach to the funding and support of open and effective data management practices. The Assembly was guided by an international Steering Committee which consisted of experts from research and user communities from participating Belmont Forum member countries. Members of the Steering Committee were responsible for leading one or more Assembly working groups (Work Packages) in order to collectively assess existing international e-infrastructure capabilities, identify gaps and overlaps, prioritize challenges, and provide recommendations on how to best address the Belmont Challenge. Logistical and administrative support was provided by a joint US-UK Secretariat.

1.3 E-Infrastructures and Data Management CRA Members

The main sections of the report were written by the project Secretariat with guidance and significant input from the Steering Committee, with review and edits from the Assembly. This final report is a synthesis of:

- 1) Comprehensive reports by each Work Package on the state of the art, barriers, gaps and best practices;
- 2) Steering Committee contributions from a series of in-person and virtual meetings, and;
- 3) Feedback from meetings of national delegations of the participating Belmont Forum countries.

This report prioritized actions best suited for the Belmont Forum collaboratively to address *interoperability* and organizational challenges in data management and e-infrastructure, and to identify existing national and international initiatives which demonstrate good practice to create a global momentum toward thoughtful data management. Therefore, this report aimed to:

- Identify strategic science policies, outlining what can be done better, in a multilateral way, to support global change research;
- Clearly express global e-infrastructure needs, barriers and gaps;
- Inform stakeholders;
- Prioritize actions to address interoperability challenges.

2. Report Summary

An e-infrastructure that supports data-intensive, multi-disciplinary research is needed to facilitate new discoveries and accelerate the pace of science to address 21st century global change challenges. Data discovery, access, sharing and interoperability collectively form core elements of an emerging shared vision of

e-infrastructure for scientific discovery. These elements further depend on building relationships among data sets, people, systems, organizations and networks. However, the pace and breadth of change in data and information management across the *data lifecycle* means that no one country or institution can unilaterally provide the leadership and resources required to use data and information effectively, or to establish and maintain the relationships needed to support a coordinated, global e-infrastructure.

The Belmont Forum represents many of the world's largest and most influential funders of environmental and social science research. It is uniquely capable of catalyzing international collaboration and leveraging existing national programs to effectively initiate and guide best practice in data stewardship, data sharing and e-infrastructure development to meet global change research needs. Furthermore, alignment of international and cross-domain efforts in interoperability will promote new interdisciplinary and international scientific understanding relevant to the Belmont Forum research agenda. As such, ***the Belmont Forum is ideally poised to play a vital and transformative leadership role in establishing a sustained human and technical international data e-infrastructure to support global change research.*** This *Community Strategy and Implementation Plan* (CSIP)³ proposed an initial path forward.

2.1 Recommendations

The Belmont Forum was presented with the overarching and synergistic recommendations listed below, through its unique role in global research collaboration, to: fill critical global e-infrastructure gaps; improve data management and exploitation; coordinate and integrate disparate organizational and technical elements; share best practices; and foster new *data literacy* to enable actionable and societally beneficial science. These recommendations have the potential to transform the way data are used and research is conducted by accelerating discovery, increasing the value of research in decision-making, and catalyzing changes throughout the economy and society that are of value to all citizens.

The five recommendations are:

1. **Adopt Data Principles** that establish a global, interoperable e-infrastructure with cost-effective solutions to widen access to data and ensure its proper management and long-term preservation. Researchers should be aware of, and plan for, the costs of data intensive research.
2. **Foster communication, collaboration and coordination** between the wider research community and the Belmont Forum, and across Belmont Forum projects through a Data and e-Infrastructure Coordination Office established within a Belmont Forum Secretariat.
3. **Promote effective data planning and stewardship** in all research funded by Belmont Forum agencies to enable harmonization of the *e-infrastructure data layer* through enhanced project data planning, monitoring, review and sharing.

³ Available online: http://www.bfe-inf.org/sites/default/files/doc-repository/A_Place_to_Stand-Belmont_Forum_E-Infrastructures_Data_Management_CSIP.PDF

4. **Determine international and community best practice** to inform e-infrastructure policy for all Belmont Forum research, in harmony with evolving research practices and technologies and their interactions, through identification and analysis of cross-disciplinary research case studies.
5. **Support the development of a cross-disciplinary training curriculum** to expand human capacity in technology and data-intensive analysis methods for global change research, and increase the number of scientists with cross-cutting skills and experience in best practice.

These recommendations have been adopted by the Belmont Forum, and the remainder of this paper describes what has been adopted and agreed.

2.2 Motivation

2.2.1 A New Data Literacy for the 21st Century

The United Nations noted that the world needs a new data literacy that enables actionable and socially-beneficial science to address environmental change affecting disaster mitigation, resilience, water and other natural resources.⁴ Broader and more effective development of best practice in data stewardship, sharing and cross-disciplinary use are pillars of the new data literacy and the basis of *Open Science* and, more generally, of the direction of science itself. Global access to data will change the ways we address environmental change problems and also change our behavior; mastery in the management and exploitation of data is key to successful collaboration and future research.

2.2.2 Unique Challenges in Global Change Research

Global change research is a crucible for shaping e-infrastructure technologies and research practices. Free and open exchange of data, methods and results, as well as effective data stewardship, are central to advancing scientific enquiry in all fields but there are particular challenges and needs in cross-disciplinary research areas. Challenging multi-disciplinary research questions relating to the Earth system span physical (e.g. atmosphere, land, and oceans), political, social and geographical boundaries, requiring data and information to be interoperable and exchangeable worldwide. Global change research also integrates diverse observations, data-intensive analytical methods and numerical models across numerous scientific domains. It requires extensive data storage and movements, including emerging capacities in *cloud computing* and *High Performance Computing*. In addition, there is a need to preserve historical, often “small” and disparate data, as much of global change research relies on observations that by definition cannot be repeated. Both the public and commerce have a high level of interest in the results, leading to an increasing demand for veracity, dissemination and citizen involvement.

⁴ A World That Counts: Mobilizing The Data Revolution for Sustainable Development. undatarevolution.org/report/.

2.2.3 Importance of Overcoming Historical Barriers to Interoperability

Major regional, national and international e-infrastructure efforts⁵ have noted that cultural, social and organizational barriers to global data sharing and interoperability generally exceed technical barriers. These non-technical aspects are easily overlooked or considered outside the scope of domain and of information and communication science and technology programs. Funding strategies by research agencies have also inadvertently bolstered these barriers by supporting investigator- or discipline-generated projects that are generally disconnected from each other and are typically independent of an overarching, integrated framework that would contribute to a coordinated e-infrastructure. Similarly, policy has often focused independently on particular segments of the data lifecycle (such as data acquisition, storage and distribution or data-intensive High Performance simulation) whereas a policy which bridges the whole data lifecycle is required for a healthy data-intensive e-infrastructure environment. Thus, the emphasis in this report is to integrate across the technical and non-technical aspects of interoperable data and e-infrastructure.

2.2.4 Reproducibility in Science

In October 2014, the Belmont Forum Principals requested that this CRA consider issues regarding reproducibility in science. Elements of reproducibility underpin all science, including global change research. They include: reuse of data and code; need for data repositories and sharing platforms; standards required for sharing code and data effectively and accurately; citation, *provenance*, *metadata*, tools and incentive mechanisms; capture and sharing of workflows; and ensuring domain-specific statistical reproducibility in the computational and data science software stack. Accurate capture and free exchange of data and information is inherent in this. Reproducibility is thus not drawn out separately but is interwoven into its conclusions and recommendations. The term “reliability” of data is emerging as a possible alternative descriptor of the issues involved in reproducibility of science.

2.3 Findings and Recommendations

2.3.1 Vision

The Belmont Forum vision is of high quality, reliable and multidisciplinary global change research enabled by a sustained human and technical, internationally coordinated and data-intensive e-infrastructure able to process a continuous increase in the diversity and volume of data generated. In such a research-driven e-infrastructure, data should be discoverable, reusable, open and accessible by default as far as possible. In addition, the data’s fitness-for-purpose should be assessed using transparent metadata relating to trustworthiness and quality. To realize this vision and maximize the return on public investments in research, all stakeholders need appropriate incentives to contribute to and support this vision. ***The Belmont Forum can blaze a path towards achieving this vision by implementing the recommendations outlined below.***

⁵ COOPEUS, RDA, ICSU-WDS, DataONE, DIAS, ESIP, EarthCube, GBIF, GEOSS, iCORDI, INSPIRE and OneGeology.

- [Adopt Data Principles and Policy](#)

The Belmont Forum has adopted a common data policy and the following principles to widen access to data and promote its long-term preservation in global change research; help improve data management and exploitation; coordinate and integrate disparate organizational and technical elements; fill critical global e-infrastructure gaps; share best practices; and foster new data literacy.

The Belmont Forum recognizes that significant advances in open access to data have been achieved and implementation of this policy and these principles requires support by a highly skilled workforce. The Belmont Forum recommends a broad-based training and education curriculum as an integral part of research programs and encourages researchers to be aware of, and plan for, the costs of data intensive research. The Belmont Forum's ambition is that this policy and these principles will take positive steps toward establishing a global, interoperable e-infrastructure based on cost-effective solutions that can help enable actionable and societally beneficial science.

Data should be:

- **Discoverable** through catalogues and search engines
- **Accessible** as open data by default, and made available with minimum time delay
- **Understandable** in a way that allows researchers—including those outside the discipline of origin—to use them
- **Manageable** and protected from loss for future use in sustainable, trustworthy repositories

The Belmont Forum and its members will support and promote this data policy and principles with the intent of making these data principles enforceable over time.

The development of these principles was informed by data principles generated and recommended by many international programs, such as the G8. These principles underpin the recommendations in this report as they inform the nature of the data plans and help identify best practice.

- [II. Foster Communication, Collaboration and Coordination](#)

An appropriate organizational and community-building environment is necessary to: resolve barriers and gaps in global data sharing and interoperability; build relationships; distill information from data; and align incentives for effective and collaborative data management.

Otherwise, the current trend of competing or conflicting technology development and agency policies will endure. While this work is, and will continue to be, undertaken largely in a national context, the

Belmont Forum can place it into a global context by fostering the appropriate coordination and collaboration environment. *The Belmont Forum can and must champion the organizational, community-building and*



technical framework needed to facilitate the international and interdisciplinary exchange of global change information through its member organizations, both individually and collectively.

- **III. Promote Effective Data Planning and Stewardship**

Communicating best practice in data and information stewardship and sharing will not only help to improve collaborative efforts but also reduce the associated risks and costs of data management. This involves: paying attention to the full lifecycle of data use and the rates at which information is gleaned from data; changing policies to promote better and more effective data planning; adopting data stewardship principles; and implementing incentives for their adoption, similar to the ways in which scientists are incentivized to publish research results. Establishing good practice is fundamental to improving data availability and interoperability. It will enable co-evolution of research needs with e-infrastructure, increase data usefulness, build trust among stakeholders, and reduce overall costs resulting from ineffective data management. ***The Belmont Forum is ideally positioned to achieve significant impact by collectively changing grant funding policies and reward systems to promote more effective data planning and stewardship.***

- **IV. Determine International and Community Best Practice to Inform Belmont Forum Research e-Infrastructure Policy**

Individual research domains successfully exchange best practice, either through scholarly publishing or increasingly through exchanging information via the Internet using a variety of mechanisms and applications. While there are beacons of good practice, there are inconsistencies in the exchange of information and the shaping and sharing of data-intensive e-infrastructure between nations and across domains and users. The rapid pace of change in technology and its adoption makes the normal development of good practice difficult and it is unclear whether the market will produce suitable solutions without intervention. Environmental and social sciences have a strong need to preserve and exchange information globally and all Belmont Forum members have examples of good practice to share. ***The Belmont Forum is uniquely placed to review worldwide and discipline-specific current practice and to foster best practice (in data sharing stewardship, analysis, modeling and workflows, and in the implementation of e-infrastructures) to promote efficiencies and trust in data and e-infrastructure solutions.***

- **V. Support the Development of a Cross-Disciplinary Training Curriculum to Build Capability**

E-Infrastructures globally lack enough skilled people who understand data management and data intensive methods in environmental, social and health sciences, and in engineering to effectively drive this area forward. While training exists in a number of domains, it is frequently restrictive in scope. In addition, formal training is typically aimed at university students and early career researchers but there is a strong need for established scientists to become more data-enabled and data-proficient. Significant progress in building this capability can be achieved through cataloguing, accrediting and enhancing existing training efforts, filling critical gaps in a nascent global curriculum, and sharing methods for interdisciplinary and transdisciplinary exploitation of data.

The Belmont Forum is well placed to stimulate new ways of thinking and working amongst distributed and diverse researchers, data and information scientists and data-enabled domain scientists, enabling them to better address global change research challenges.

2.3.2 Leveraging the Power of the Belmont Forum

If the planet were a patient in a modern intensive care hospital unit, there would be a coordinated set of sophisticated monitors and instruments, rapid analysis and presentation of test results, a team of medical professionals coordinating diagnosis and treatment according to proven medical principles and best practices, and a set of available experts from different specialties drawing on the best available medical research and data. The Belmont Forum is in a unique position to develop key pieces of a comparable global e-infrastructure. It can act as a catalyst for promoting dialogue and collaboration, and leverage - but not replace - existing national programs. It also provides a synergistic, top-down approach that complements bottom-up activities carried out by individual nations and organizations across the globe.

Implementation of these recommendations could include adopting internal actions and policies to align Belmont Forum efforts with external developments, influencing research investments judiciously, targeting limited resources where they are uniquely or best suited, or issuing funding calls (such as a networking or community-building action, a call to run a summer school or develop training materials, small-scale priming activities, large-scale research activities, or whatever is most appropriate to address the issue in question). For some actions, the Belmont Forum could identify that a CRA or invitation to tender would be the best funding mechanism to address an issue.

The challenges and opportunities in creating coordinated, global and interoperable e-infrastructure are complex but addressing them will result in tremendous benefits to stakeholders at all levels. These challenges are also clearly outside the ability of any single entity to attempt to control or implement, both in terms of resources and authority. Development of an e-infrastructure capable of supporting the existing and emerging global change research agenda has been, and will likely continue to be, organic with many aspects unpredictable and disruptive. It must therefore be agile and adaptable to meet changing research needs and technology development. Shared responsibilities are a key to success.

2.3.3 Shared Responsibilities

We have described the rationale for the Belmont Forum to undertake the recommended actions but have not discussed what the larger research and computing communities should do for Belmont Forum e-infrastructure and data management actions to be successful. Do individual Belmont Forum members take independent action? What should external entities and funding agencies do to support these activities? How does the Belmont Forum respond to external dynamics?

Globally, researchers and governments alike are recognizing the importance of data discovery, access, information sharing and interoperability. These collectively form core elements of an emerging shared vision

of e-infrastructure for scientific discoveries, governance and resource management. There are numerous challenges to achieving these ambitious goals, many of which have been identified through existing Earth and related science informatics community initiatives. This broad, loosely-coupled community has identified many of the technical and social challenges to e-infrastructure but developing solutions that are adopted and collectively enhanced by the scientific community is still difficult. By building a cohesive international community committed to this e-infrastructure vision, the Belmont Forum can create opportunities for shared and more sustainable efforts toward removing barriers to interoperability on a global scale. GEO is one of the key international organizations involved with this initiative, together with ICSU and Future Earth, and the Research Data Alliance. This collaboration means that this initiative will build on the excellent work already in progress, including in the Arctic, and ensure no loss of momentum and engagement with the wider scientific community.

3. Actions to Catalyze Recommendations⁶

- **Action Theme 1: Coordination Office**

Foster communication, collaboration and coordination through the establishment of a Data and e-Infrastructure Coordination Office

- **Action Theme 2: Data Planning**

Promote effective data planning and stewardship in all research funded by Belmont Forum agencies

- **Action Theme 3: e-Infrastructure**

Determine international and community best practice in order to inform e-infrastructure policy for all Belmont Forum research

- **Action Theme 4: Human Dimensions**

Support the development of a cross-disciplinary training curriculum to build capability

“Too often, development efforts have been hampered by a lack of the most basic data about the social and economic circumstances in which people live... We must also take advantage of new technologies and access to open data for all people.”
 - Bali Communiqué of the High-Level Panel, March 2013

⁶ Please see Appendix 1: References for supporting documentation

3.1 Broader Impacts

3.1.1 Benefits of Acting

This proposed set of initiatives will enable the Belmont Forum to fulfill its charge better to “*to deliver knowledge needed for action to avoid and adapt to detrimental environmental change including extreme hazardous events.*” In addition, through internal adoption by individual Belmont Forum members, these recommendations will have much broader impacts for disciplines and programs outside of environmental change research and for organizations engaged in scientific and technical research and operations worldwide.

Accelerate the Pace of Scientific Discovery

The recommendations have the potential to transform the way research is conducted by accelerating discovery, increasing the value of research decision-making, and catalyzing changes throughout the economy and society that are of value to all citizens. New scientific discoveries and socio-economic innovation will emerge from tackling the large increase in diversity, volume and rate of growth of multidisciplinary data. Establishing and enabling a cross-disciplinary framework and data-intensive e-infrastructure, with network and computational elements, will allow scientific knowledge to transcend disciplines and address new environmental change problems. Acting now, at a stage early in the development of distributed network solutions and similar elements of e-infrastructure, means that the Belmont Forum can have extraordinary influence on those specialized developments.

Broaden Dissemination of Best Practice

Actions to adopt and use best practices for research data and e-infrastructure planning and development will ultimately benefit current and future Belmont Forum-funded research, and the general research landscape. This could foster greater trust in research outputs, because data are available for validation and reuse.

Enhance Coordination

Developing coordinated and interoperable data and e-infrastructure includes mapping relevant activities in and among organizations. Mapping will enhance collaboration and general practice within the Belmont Forum, across activities within member agencies and countries, and in institutions involved in the global coordination of environmental and social science information. It will harmonize efforts and organizations, lessen volunteer fatigue, reduce redundancy and duplication of effort, and increase the impact of funding initiatives.

Build Capability

Facilitating international, cross-disciplinary training will increase the potential for broader, global participation in research, and expand human capability and competitiveness. This will result in products and publications of greater benefit to the international community. Students and researchers, especially from developing nations, will also benefit from the opportunity to present their research problems and materials, compare best practice, and network with contemporaries in other countries and disciplines. In itself, this will be an important legacy of the investments described here. Taking all these investments together, they will be transformative.

3.1.2 Consequences of Not Acting

Impaired Ability to Respond to Detrimental Effects of Environmental Change

Global change research is extremely time-critical. Given the immediate and long-term risks of environmental change, together with the ever-increasing amounts of research data being generated, much damage would be done to the field of study (Earth) and our ability to start formulating meaningful evidence-driven actions if delays force us to start again or backtrack. Not acting may limit our options and ability to respond to crises, since avoidable errors in decisions occur daily. Decision makers may not know about reasonable options for adaptation and mitigation because data and knowledge were not shared, or Earth system models will incorrectly assess impacts because they did not incorporate realistic or current data. We can also lose visibility of existing data if they are not curated and made accessible to modern e-infrastructures. Avoiding such errors and loss of data by promoting better access, preservation and use of existing data would yield significant financial savings, reduce distress and save lives.

Lost Opportunities and Squandered Valuable Resources

Not acting will create lost opportunities, delays in achieving Belmont Forum goals, squandering of valuable resources in the form of increased costs to retrofit incompatible data, software and scientific results, and losing data irretrievably. Not acting could also result in losing momentum in the application of globally integrated e-infrastructure for research, which has potentially profound economic and societal consequences. Not acting also means that, in the void of truly globally accepted agreements, special interest developers may be the only option and may drive solutions that are incompatible with environmental and social science research needs.

Appendix 1: References

The following references support the recommendations described in the Community Strategy and Implementation Plan:

A. Action Theme 1: Communication, Coordination, and Collaboration

Deese, B. C. and J. P. Holdren. 2014. Memorandum for the heads of Executive Departments and Agencies: Science and Technology Priorities for the FY 2016 Budget. Executive Office of the President of the United States. whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

European Commission, Commission to the Council and the European Parliament. 2012. On the Experience Gained in the Application of Directive 1003/4/EC on Public Access to Environmental Information. eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0774:FIN:EN:PDF

Holdren, J., T. Dickenson, G. Paulson, et al. 2014. National Plan for Earth Observations. National Science and Technology Council, Executive Office of the President. 71 pp. whitehouse.gov/sites/default/files/microsites/ostp/NSTC/2014_national_plan_for_civil_earth_observations.pdf

Office of Science and Technology Policy. 2013. Memo: Increasing access to the results of federally funded research. whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

President's Council on Advancing Science and Technology (PCAST) 2011. Working Group on Biodiversity Preservation and Ecosystem Services. In: Sustaining Environmental Capital Report. whitehouse.gov/sites/default/files/microsites/ostp/pcast_sustaining_environmental_capital_report.pdf

B. Action Theme 2: Data Planning

1. OECD Principles and Guidelines for Access to Research Data from Public Funding (OECD, June 2007) oecd.org/sti/sci-tech/38500813.pdf

2. Riding the wave: How Europe can gain from the rising tide of scientific data (EU, October 2010) cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

3. The Data Harvest: How sharing research data can yield knowledge, jobs, and growth (RDA, December 2014) europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf

4. Science as an open enterprise (The Royal Society, UK, June 2012) royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

5. Increasing Access to the Results of Federally Funded Scientific Research (US OSTP, February 2013)
whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
6. Making Open and Machine Readable the New Default for Government Information (US White House, May 2013) whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-
7. G8 Open Data Charter and Technical Annex (G8, June 2013) gov.uk/government/publications/opendata-charter/g8-open-data-charter-and-technical-annex
8. Today's Data, Tomorrow's Discoveries (US NSF, March 2015)
nsf.gov/publications/pub_summ.jsp?ods_key=nsf15052
9. Increasing Access to the Results of Scientific Research (US NASA, November 2014)
science.nasa.gov/media/medialibrary/2014/12/05/NASA_Plan_for_increasing_access_to_results_of_federally_funded_research.pdf
1. ncbi.nlm.nih.gov/pmc/articles/PMC3817176/
2. datapub.cdlib.org/2014/03/03/finding-disciplinary-data-repositories-with-databib-and-re3data/
3. docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1001&context=lib_fspres

Relevant work on data publishing has been conducted by RDA/WDS and CODATA working groups, amongst others, which can easily be adapted and updated. The following texts were additionally identified as relevant evidence of the need for this action;

1. Costas, R., Meijer, I., Zahedi, Z. and Wouters, P. (2013). The Value of Research Data - Metrics for data sets from a cultural and technical point of view. A Knowledge Exchange Report, available from knowledge-exchange.info/datametrics.
2. Mooney, H, Newton, MP. (2012). The Anatomy of a Data Citation: Discovery, Reuse, and Credit. Journal of Librarianship and Scholarly Communication 1(1):eP1035. dx.doi.org/10.7710/2162-3309.1035. (Motivates reward structures to encourage data publication).
3. Klump, J. (2012), Offener Zugang zu Forschungsdaten: Open Data und Open Access to Data – Die ungleichen Geschwister, in Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft, edited by U. Herb, pp. 45–53, universaar, Saarbrücken, Germany. [online] Available from: nbnresolving.de/urn:nbn:de:bsz:291-universaar-873

C. Action Theme 3: e-Infrastructure

1. A Science Plan for Integrated Research on Disaster Risk, 2008:
wcdrr.org/uploads/Sendai_Framework_for_Disaster_Risk_Reduction_2015-2030.pdf The ICSU Priority Area

Assessment on Environment and its Relation to Sustainable Development (2003) and the ICSU Foresight Analysis (2004) both proposed 'Natural and human-induced hazards' as an important emerging issue. Responding to this proposal, the Science Plan of Integrated Research on Disaster Risk (IRDR) aims to generate new information and data, and leave a legacy of coordinated and integrated global data and information sets across hazards and disciplines, with unprecedented degrees of access.

2. GEOSS Data Sharing Action Plan, 2010:

[earthobservations.org/documents/geo_vii/07_GEOSS%20Data%20Sharing%20Action%20Plan%20Re v2.pdf](http://earthobservations.org/documents/geo_vii/07_GEOSS%20Data%20Sharing%20Action%20Plan%20Re%20v2.pdf)

The "GEOSS Data Sharing Principles" is one of the first accomplishments of the Group on Earth Observations (GEO). It states: a. There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation; b. All shared data, metadata and products will be made available with minimum time delay and at minimum cost; and c. All shared data, metadata and products should be free of charge or no more than cost of reproduction will be encouraged for research and education.

4. Future Earth 2025 Vision, 2014: futureearth.org/sites/default/files/future-earth_10-year-vision_web.pdf.

The vision of Future Earth is for people to thrive in a sustainable and equitable world. It says that Future Earth is contributing to improved modes of sharing data about environmental change and progress towards sustainability in order to support policy and practice at different levels. The outputs include science-based data, tools and resources to support improved resilience of people, communities and economies, including disaster risk reduction.

5. The Road to Dignity by 2030:

un.org/disabilities/documents/reports/SG_Synthesis_Report_Road_to_Dignity_by_2030.pdf Ending Poverty, Transforming All Lives and Protecting the Planet, 2014 calls for inclusive, agile and coordinated action to usher in an era of sustainable development for all. Secretary-General Ban Ki-moon presented the United Nations (UN) General Assembly with this document, which will guide negotiations for a new global agenda centered on people and the planet, and underpinned by human rights. It emphasizes the role of data in the new agenda by saying that the world must acquire a new 'data literacy' in order to be equipped with the tools, methodologies, capacities, and information necessary to shine a light on the challenges of responding to the new agenda.

6. A World that Counts, 2014: undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf.

This document was published by the UN Secretary-General's Independent Expert Advisory Group (IEAG) on a Data Revolution for Sustainable Development. It emphasizes that data are the lifeblood of decision-making and the raw material for accountability and that effective policies become almost impossible without high-quality data providing the right information on the right things at the right time.

7. Sendai Framework for Disaster Risk Reduction 2015-2030:

wcdrr.org/uploads/Sendai_Framework_for_Disaster_Risk_Reduction_2015-2030.pdf Representatives from 187 UN member states have adopted the first major agreement of the Post-2015 development agenda, a far

reaching new framework 'Sendai Framework for Disaster Risk Reduction 2015-2030', for disaster risk reduction with seven targets and four priorities for action. It also promotes and enhances, through international cooperation (including technology transfer), access to and the sharing and use of non-sensitive data, information, communications and geospatial and space-based technologies and related services.

8. Sustaining Domain Repositories for Digital Data: A White Paper 2013:

datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf This paper addresses some of the common needs of domain repositories across the natural, social, and health sciences, though not explicitly the issue of how to fill gaps between/across domains.

D. Action Theme 4: Human Dimensions

UK: "Most Wanted II - Postgraduate and Professional Skills Needs in the Environment Sector"

nerc.ac.uk/skills/postgrad/policy/skillsreview/2012/

UK: Employer Engagement - enhancing HEI engagement with the Satellite Industry Final Report

[hestem.ac.uk/resources/outputs/projects?keys=Space&x=0&y=0&field_author_date_value\[value\]\[year\]=&field_discipline_value_many_to_one=All&field_activity_project_nid=All](http://hestem.ac.uk/resources/outputs/projects?keys=Space&x=0&y=0&field_author_date_value[value][year]=&field_discipline_value_many_to_one=All&field_activity_project_nid=All) < A1-24 > Appendix 1, Action Theme 4:

Human Dimensions

- E. EU: (SIM4RDM produced an EU landscape report in 2013, which describes the need for data management plans and how researchers said they would benefit from face-to-face support and training sim4rdm.eu/documents/project-outputs)

Schmidt, B., Gemeinholzer, B., Treloar, A (2016): Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. [PLOS One, DOI: 10.1371/journal.pone.0146695](https://doi.org/10.1371/journal.pone.0146695)

The Open Data Survey (completed as part of this CRA) with the responses of more than 1000 global participants, cites only 23% of respondents having any awareness of data publishing guidelines of any type. The report strongly recommends the support of training activities in this area, noting particularly, "Support and training activities should be supported in concerted ways, targeting researchers as well as current and future data and information professionals." Crucially, all evidence identifies the global shortage of researchers who are literate in cross-cutting and interdisciplinary skills - environmental and social scientists who are also skilled in informatics, or information technologists who have environmental and social science expertise.