

30th June 2015

A Place to Stand: e-Infrastructures and Data Management for Global Change Research

**Belmont Forum e-Infrastructures
& Data Management Community
Strategy and Implementation Plan**

"Give me a place to stand, and I will move the world"

- Archimedes

**Steering Committee, Belmont Forum e-Infrastructures
and Data Management Collaborative Research Action**

Lee Allison, Arizona Geological Survey (Co-Chair)

Robert Gurney OBE, University of Reading (Co-Chair)

OVERVIEW

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

“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*

An e-infrastructure that supports data-intensive, multidisciplinary 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 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) presents an initial path forward.

Recommendations

The Belmont Forum is urged to adopt 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.

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 Belmont Forum agency-funded research to enable harmonization of the *e-infrastructure data layer* through enhanced project data planning, monitoring, review and sharing.
4. **Determine international and community best practice** to inform Belmont Forum research e-infrastructure policy, 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 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.

REPORT STRUCTURE

The recommendations and the rationale for the Belmont Forum to implement them are described in the main section of this report. Appendix 1: Recommendation *Action Themes* describes specific actions to implement these recommendations (Action Theme 1: Coordination Office, Action Theme 2: Data Planning, Action Theme 3: e-Infrastructure, Action Theme 4: Human Dimensions). Appendix 2: Action Dependencies - describes dependencies among the four Action Themes. Appendix 3: Transition and Implementation - describes proposed next steps to assist the Belmont Forum in preparing for its annual meeting in October 2015 and with implementing the recommendations. Appendix 4: Relevant Organizations and Programs - describes the organizations and programs cited throughout the report. Appendix 5: Acronyms and Glossary - lists the acronyms commonly used in the report and defines several key terms. Such terms are italicized the first time they appear.

TABLE OF CONTENTS

Motivation	1
A New Data Literacy for the 21 st Century	1
Unique Challenges in Global Change Research	2
Importance of Overcoming Historical Barriers to Interoperability	2
Reproducibility in Science	2
Findings and Recommendations.....	3
Our Vision	3
Adopt Data Principles	3
Foster Communication, Collaboration and Coordination.....	4
Promote Effective Data Planning and Stewardship	4
Determine International and Community Best Practice to Inform Belmont Forum Research e-Infrastructure Policy	4
Support the Development of a Cross-Disciplinary Training Curriculum to Build Capability	5
Leveraging the Power of the Belmont Forum	5
Shared Responsibilities.....	6
Broader Impacts	6
Benefits of Acting.....	6
The Consequences of Not Acting.....	7
Next Steps.....	8
Acknowledgements	8
Appendix 1: Recommendation Action Themes.....	A1-1
Action Theme 1: Coordination Office.....	A1-1
Action Theme 2: Data Planning	A1-7
Action Theme 3: e-Infrastructure.....	A1-13
Action Theme 4: Human Dimensions.....	A1-20
Appendix 2: Action Dependencies.....	A2-1
Appendix 3: Transition and Implementation	A3-1
Appendix 4: Relevant Organizations and Programs.....	A4-1
Appendix 5: Acronyms and Glossary	A5-1
Acronyms and Abbreviations	A5-1
Glossary	A5-1

MOTIVATION

e-Infrastructures and Data Management Collaborative Research Action

The Belmont Forum initiated the e-Infrastructures and Data Management Collaborative Research Action (CRA) in recognition that:

“...the need to address global environmental challenges requires a more coordinated approach to the planning, implementation, and management of data, analytics and e-infrastructures through international collaboration.”
- Belmont Forum, New Delhi, February 2013

This Community Strategy and Implementation Plan results from a series of collaborative scoping activities, including a survey of the global change research community, conducted as part of an extensive 18-month international review by domain scientists, computer and information scientists, legal scholars, social scientists and other experts on the state of global data management, networking, computing, legal issues and governance. The result is a set of recommendations on how the Belmont Forum can leverage existing resources and investments.

“There is no turning back the clock on our interconnected world, but we could jeopardize its benefits if we fail to invest in a trusted data environment”.
- Ellen Richey, Chief Enterprise Risk Officer, Visa, USA, in WEF Blog on Big Data

CRA participants focused on assessing existing capability in the areas of technical architecture, governance standards and research skills. The goal of this work was to determine how best to promote interoperability of data and e-infrastructure, and identify a more coordinated, holistic, and sustainable approach to the funding and support of global environmental change research. Our evaluation identified gaps and barriers in the wider dissemination and use of the critical data generated and used by global change researchers.

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.¹ 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.

¹ A World That Counts: Mobilising The Data Revolution for Sustainable Development. undatarevolution.org/report/.

Unique Challenges in Global Change Research

Global change research is a crucible for shaping e-infrastructure technologies and research practices more generally. 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 multidisciplinary 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.

Importance of Overcoming Historical Barriers to Interoperability

Major regional, national and international e-infrastructure efforts² 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 balanced and agile aligned support for bridging 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.

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 in this report but is interwoven into its

² COOPEUS, RDA, ICSU-WDS, DataONE, DIAS, ESIP, EarthCube, GBIF, GEOSS, iCORDI, INSPIRE and OneGeology. For more information, please see Appendix 4: Relevant Organizations and Programs.

conclusions and recommendations. The term “reliability” of data is emerging as a possible alternative descriptor of the issues involved in reproducibility of science.

FINDINGS AND RECOMMENDATIONS

Our Vision

Our 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 by 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.***

Adopt Data Principles

Adopting the ***five data principles*** listed below, through the authority and reputation of the Belmont Forum, will help establish a global and interoperable e-infrastructure to widen access to data and ensure its long-term preservation in global change research.

Research data must be:

1. ***Discoverable*** through catalogues and search engines, with data access and use conditions, including licenses, clearly indicated. Data should have appropriate persistent, unique and resolvable identifiers.
2. ***Accessible*** by default, and made available with minimum time delay, except where international and national policies or legislation preclude the sharing of data as *Open Data*. Data sources should always be cited.
3. ***Understandable and interoperable*** in a way that allows researchers, including those outside the discipline of origin, to use them. Preference should be given to non-proprietary international and community standards via data e-infrastructures that facilitate access, use and interpretation of data. Data must also be reusable and thus require proper contextual information and metadata, including provenance, quality and uncertainty indicators. Provision should be made for multiple languages.
4. ***Manageable*** and protected from loss for future use in sustainable, trustworthy repositories with data management policies and plans for all data at the project and institutional levels. Metrics should be exploited to facilitate the ability to measure return on investment, and can be used to implement incentive schemes for researchers, as well as provide measures of data quality.
5. ***Supported*** by a highly skilled workforce and a broad-based training and education curriculum as an integral part of research programs.

The development of these principles was informed by previous 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.

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.***

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 will 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, enabling co-evolution of research needs with e-infrastructure, increasing data usefulness, building trust among stakeholders, and reducing 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.***

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 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 foster best practice (in data sharing stewardship, analysis,***

modeling and workflows, and implementation of e-infrastructures) to promote efficiencies and trust in data and e-infrastructure solutions.

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 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 also 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.***

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.

"If you want to go fast, go alone. If you want to go far, go together."
- African proverb

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

also are 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.

Shared Responsibilities

We 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.³ 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.

BROADER IMPACTS

Benefits of Acting

This proposed set of initiatives will better enable the Belmont Forum to fulfill its charge 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.

“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

³ See Appendix 4: Relevant Organizations and Programs.

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.

The 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

environmental change and extreme hazards 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 sciences research needs.

NEXT STEPS

The Secretariat will widely disseminate this report to foster dialog in the scientific and technical communities on its findings and recommendations. We look forward to working with Belmont Forum members to evaluate these recommendations and assist with their implementation should the Belmont Forum choose to act.

ACKNOWLEDGEMENTS

The project Secretariat gratefully acknowledges the contributions of the Steering Committee in the organization and conclusions of this report, with invaluable guidance and insights from the Group of Program Coordinators and the project Assembly for this CRA. Supporting documents and evidence can be found in the appendices to this report and on the project *Knowledge Hub* at www.bfe-inf.org. This CRA was supported by the US National Science Foundation (NSF) under Grant No. 1358690 and the UK Natural Environment Research Council (NERC) under Grant Ref NE/LO14391/1. Members of the Steering Committee were supported by their individual Belmont Forum participating agencies. Opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF or NERC.

APPENDIX 1: RECOMMENDATION ACTION THEMES

Action Theme 1: Coordination Office

Foster Communication, Collaboration and Coordination through the Establishment of a Data and e-Infrastructure Coordination Office

Executive Summary

- 1. Action Theme Summary:** The Belmont Forum should establish a data and e-Infrastructure Coordination Office within the proposed Belmont Forum Secretariat to foster communication, collaboration and coordination across Belmont Forum funded projects, engage with the wider global e-infrastructure community and carry out duties described below.
- 2. Objectives, Justification, Impact:** The Coordination Office will support communication, collaboration and coordination among Belmont Forum-funded activities, e-infrastructure organizations, projects and experts, to promote greater awareness of the wider landscape of activity, facilitate access to external resources, and foster cooperation among and other projects. Activities include: 1) implement, monitor, support and evaluate interoperability capabilities at the data and system level among Belmont Forum-funded projects and participants; 2) reduce or avoid duplication; 3) leverage resources and ongoing Belmont-Forum funded projects and activities (including those in the other Action Themes); 4) map the data and e-infrastructure landscape; and 5) facilitate global interoperability convergence and coordination between key services.

There are numerous national and international research initiatives⁴ seeking to facilitate scientific discovery and improve predictions within individual scientific domains by removing barriers to interoperability. Working closely and coordinating activities with these initiatives will be imperative. Members of the global change research, information communication science communities actively engage across many of these efforts and recognize the value of combining efforts to maximize the efficient use of resources and the productivity of working groups. However, there are technical, organizational, social and cultural challenges that must be addressed in order to successfully establish a framework for information interoperability on an international scale and across disciplines. Fostering communication, coordination and collaboration among the diverse scientific, data and computational communities will create a more sustainable and cohesive international community of practice, and will better support data-intensive research by the Belmont Forum and others by more effectively reducing barriers to data sharing and interoperability.

Description of Actions

Near-term (start within 0-2 years) and Medium-term (start within 1-5 years):

- 1. The Belmont Forum should establish a Coordination Office to foster ongoing communication, collaboration and coordination.** The Coordination Office role could be undertaken by a proposed Belmont Forum Secretariat, or it could be outsourced to an external group (decided by a funding call or an invitation to tender). This report recommends the overall Coordination Office be housed internally to the Belmont Forum because it will develop and integrate e-infrastructure and data management activities crucial to the success of all Belmont Forum projects, by better leveraging projects, sharing of

⁴ See Appendix 4: Relevant Organizations and Programs.

best practice, and promoting interoperability. However, some tasks could be outsourced, such as the mapping the landscape activities and others.

We understand that there is a proposal to establish an overall Belmont Forum Secretariat: these functions should sit in that office, be funded in the same way as the office overall, and be co-located. **The Belmont Forum should initiate the Coordination Office by appointing staff**, including a Data e-Infrastructure Officer (DIO) and a Communication, Collaboration and Coordination Officer (CCCO), and staff support as needed, along with appointed “champions” from the other Action Themes (Data Planning, e-Infrastructure and Human Dimensions) to liaise with this Coordination Office, to:

- a. **Establish (near-term) and provide support (medium-term) to a Data Policy Advisory Board (DPAB) and Security Advisory Board (SAB)** to oversee relevant legal and security issues, and contribute toward the development of the Enhanced Data Plan Template. These Boards will require staff assistance, and possible travel support, to carry out their charges, preparing reports, gathering information and responding to requests from other Belmont Forum groups (e.g., TPO, GPC, Principals, Assemblies). More information on the Advisory Boards can be found in the Data Planning Action Theme.
- b. **Oversee an initial (near-term) and ongoing (medium-term) mapping of organizations and best practice activities**, to identify the people, projects, programs and organizations working toward data and e-infrastructure interoperability in global change research, the specific roles these individuals and groups are playing, and the best practices they employ. These activities will additionally focus on mapping and compiling existing e-infrastructure and data management capacity-building activities and best practice, as well as identify critical participants for initial human dimensions workshops (see Human Dimensions Action Theme).
- c. Initiate the **establishment (near-term) and management (medium-term) of a Strategic Coordination Network (SCN)** of organizations and entities to identify collaborative strategies to address Belmont Forum objectives and actions to implement a more coordinated, holistic, and sustainable approach to the funding and support of global environmental change research. The intent is to establish necessary, trusted relationships with relevant entities, harmonize data practices, share information about and align funding mechanisms, raise awareness of and access to each other’s work, and develop a process for fostering more cooperative and targeted efforts.

Mechanisms will need to be developed and maintained to foster communication and information sharing across these efforts in the medium to long-term. Types of entities include funders (e.g., Research Data Alliance Colloquium), pan-national bodies (e.g., G7), umbrella organizations (e.g., Group on Earth Observations), publishers, learned societies, libraries, institutional repositories, data centers and relevant national and international initiatives (e.g., Research Data Alliance, EarthCube, GEOSS, OneGeology, ESFRI, GFBio)⁵. The SCN could initially be formed at the level of the Belmont Forum Principals or the Group of Program Coordinators.

- d. **Oversee the implementation, monitoring, and evaluation (near term)** of Belmont Forum e-infrastructure activities aligned around the other Action Themes in this report (Data Planning, e-Infrastructure and Human Dimensions). **Regularly report results, challenges and findings** to the Belmont Forum and stakeholder communities (medium-term).
- e. **Foster coordination (near-term) among Belmont Forum-funded projects** to share and disseminate best practices in e-infrastructure and data management, leverage work and results,

⁵ See Appendix 4: Glossary for a list of relevant organizations and programs.

and participate in case studies and workshops (medium and long-term). This serves two purposes: 1) it focuses development on projects relevant to the Belmont Forum goals; and 2) leverages these multidisciplinary international collaborations as exemplars in e-infrastructure and data management.

- f. **Provide an auditing function for Data Plans** submitted by Belmont Forum-funded projects (medium-term), as described in the Data Planning Action Theme.
- g. **Organize initial scoping workshops** to inform future case studies, human dimension capacity-building curriculum and data plans (see the e-Infrastructure Action Theme).
- h. **Coordinate with external bodies** regarding any functions and tasks of the Coordination Office that are outsourced (near-term and medium-term).

Shared Responsibilities, Anticipated Duration, Resources and Dependencies:

1. **What the community can do (who, what), how Belmont Forum can leverage or support community activities:** The Coordination Office is a shared community responsibility: we cannot assume external entities will participate unsolicited just because the Office is established. They will need to appreciate the value of enhanced communication, collaboration and coordination, and commit resources to be active participants. Leaders in the global change community have already expressed their willingness and desire to participate under Belmont Forum guidance. Community input drove the inclusion of this Action Theme in this report.
2. **Anticipated duration:** This report recommends that the Coordination Office exist throughout the lifespan of Belmont Forum e-infrastructure activities, although it may adapt over time. This Office is the overarching hub of activity which will actively implement Belmont Forum e-infrastructure activities.
3. **Resources (people, effort):** The Coordination Office requires a dedicated Data e-Infrastructure Officer (DIO) to act as the data and e-infrastructure resources and services expert within the Belmont Forum as a whole. The DIO would be the staff-lead to develop and ensure delivery of the data and e-infrastructure strategy for the Belmont Forum. Duties will include liaising with the proposed Data Policy and Security Advisory Boards, and other Data Planning activities as described in the Data Planning Action Theme.

The Coordination Office also requires a dedicated Communication, Collaboration and Coordination Officer (CCCO) responsible for coordinating among Belmont Forum projects and with external initiatives. Thematic “champions” representing each of the other Action Themes in this report (Data Planning, e-Infrastructure and Human Dimensions) will ensure a synergy among individual these efforts, and will produce a coordinated approach to improved data stewardship and e-infrastructure building. Either the DIO or CCCO position could serve as overall coordinator of Belmont Forum e-infrastructure activities, reporting to the head of a Belmont Forum Secretariat Office.

The thematic champions are anticipated to be part-time assignments to members of the Belmont Forum Coordination Office (including possibly the DIO and CCCO), to staff, or to representatives of Belmont Forum member agencies.

Dependencies: This is the overarching activity to implement Belmont Forum e-infrastructure activities under a holistic strategy and vision. As such, all of the proposed activities in this and the other action themes are dependent on the establishment of the Coordination Office.

Relationships with other activities

1. **Relationships with other Action Themes:** Communication, collaboration and coordination are central to integrating the wide range of activities proposed in the Action Themes (Data Planning, e-Infrastructure, Human Dimensions). These will be connected via the Coordination Office.
2. **Relationships with other Belmont Forum projects:** A major role of the Coordination Office will be to implement, support, monitor and evaluate synergies among Belmont Forum-funded projects and participants, with the goal of promulgating best practices in data stewardship and promoting interoperability of data and e-infrastructures in support of Belmont Forum-related research.
3. **Relationships with other activities:** The mapping of organizations, best practices and capacity-building initiatives, as well as the formation of the Strategic Coordination Network, will foster relationships with related bodies and projects to identify collaborative strategies to achieve Belmont Forum e-infrastructure and data management goals. This also will ensure that activities are not duplicated elsewhere.

Justification in Context: *Why are we recommending the Belmont Forum address this issue? Why is the Belmont Forum well or best suited to act in this way?*

The Belmont Forum should undertake this Action Theme because no other national or international bodies view this as part of their mandates or authorities, and there are no viable alternatives identified to-date for other bodies to assume responsibility.

1. **Relevance:** The Belmont Forum should establish a Coordination Office function to foster ongoing communication, collaboration and coordination. Establishing the Coordination Office will achieve the vision in the original Belmont Forum e-Infrastructures and Data Management Collaborative Research Action scoping document that “it may be more productive for the Belmont Forum to put its resources into creating the ‘space’ to bring communities together for further integration and to build a community strategy rather than to focus these resources on a competitive call for several small grants and a single strategic consortia.” This approach will also foster a more comprehensive model of a globally-coordinated e-infrastructure where data is an active component that enables wider potential applications, including for individual Belmont Forum members.

Establish and support both a Data Policy Advisory Board and a Security Advisory Board: Legal and Security areas were identified as very important in the Open Data Survey. Legal and security experts in the current CRA recommended that Legal and Security bodies need be established early, due to the rapid evolution and understanding of Data Policy and Security issues. The interaction of these issues within and among Belmont Forum member institutions also is critical in establishing community trust in overall Belmont Forum e-infrastructure and data management activities (Data Planning Action Theme).

Oversee an initial (near-term) and ongoing (medium-term) mapping of the landscape: Mapping the landscape will ensure that Belmont Forum research activities can be planned to take advantage of existing investments and that a coordinated e-infrastructure can be developed. It will also help the Belmont Forum recognize how any infrastructure elements it might fund fit into the global infrastructure landscape. Finally, it will help the Belmont Forum avoid duplication of efforts already underway.

Establish and manage a Strategic Coordination Network: Establishing the SCN will reduce risks of duplication and divergence of activities that result in lack of interoperability. Moving sooner rather than later in establishing the SCN allows greater opportunities for collaboration and cooperation, leading to wider promulgation and adoption of community best practices.

Oversee the implementation, monitoring and evaluation of Belmont Forum e-infrastructure initiative activities: This is a cross-cutting activity of the Belmont Forum that differs from traditional research projects and requires active management and participation by the funding agencies.

Foster coordination among Belmont Forum-funded projects: This is, by definition, a Belmont Forum effort to leverage existing investments and develop capabilities to enhance not only those projects but provide a coordination framework for ongoing and future activities.

Provide an auditing function for Data Plans: This provides a measure of accountability for investigators, and provides feedback and guidance on creating and implementing effective Plans. The promulgation of common Data Plans across collaborating partners in Belmont Forum projects will improve e-infrastructure and promote sustainable and reusable data.

Anticipated Impact: *For example: influence on policy and practice; impact of adopting or not adopting the recommendation, including institutionally, nationally and internationally; reproducible science, data principles.*

1. **Scale, Scope:** The communication, collaboration and coordination activities could start on a small scale by focusing on the existing Belmont Forum funded projects, in order to support them better with shared data capabilities, and to develop and test *data-intensive science* methodology and an implementation addressing Global Change research challenges. The scope of influence will grow as the role of the Coordination Office matures. Participants in similar activities in other fields warn that the magnitude of such activities and rapidity of change and new development often surprises the entities carrying them out. This kind of activity needs to be designed for medium to long term operation to be effective.
2. **Influence:** The mapping of organizations and the Strategic Coordination Network will enhance collaboration and general practice among Belmont Forum members and beyond to those institutions involved in the global change research, data management and stewardship, and e-infrastructure development. It could harmonize duplicative efforts and organizations, lessen volunteer fatigue, reduce redundancy and increase the impact of funding initiatives. Overall it should lead to a broader, and more effective e-infrastructure integrated with network and computational elements.

Evidence Bank:

The need for the proposed Actions has been widely made, both geographically and across many scientific disciplines. Two such examples follow:

1. Several of the proposed coordination activities are drawn from the EarthCube Initiative, funded by the US National Science Foundation: a community-led cyberinfrastructure in the process of developing and evaluating a test governance framework that aims to foster communication, collaboration and coordination. Relevant documents include:

- EarthCube Enterprise Governance Critical Functions (2014), available at: earthcube.org/sites/default/files/doc-repository/EarthCubeGovernanceFunctions_21July2014.pdf

- EarthCube Demonstration Governance Charter (2014), available at: earthcube.org/document/2015/earthcube-charter

2. Many of the external coordination and collaboration components in this action (mapping the landscape and the formation of the Strategic Coordination Network of organizations) are based on a scoping paper authored by several individuals, including several members of the CRA Steering Committee and in the US Delegation (“Increasing the effectiveness of international scientific research data interoperability efforts” (2014) by Lindsay Powers et al.).

The scoping paper was born out of a breakout session on international e-infrastructure at the Earth Science Information Partnership (ESIP) 2014 Summer Meeting, which brought together leaders from a wide range of regional and global organizations and projects engaged in e-infrastructure for the environmental and Earth sciences. The group concluded that while coordination was a critical community need, none of the existing bodies were capable of taking it on, that it could be a significant longer term undertaking, and that the Belmont Forum appeared to be uniquely positioned to provide this capability. References cited in the scoping paper include:

- 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

Action Theme 2: Data Planning

Promote effective data planning and stewardship in all Belmont Forum agency-funded research

Executive Summary

1. **Action Theme Summary:** The Belmont Forum should promote active and effective data stewardship in all Belmont Forum funded research, and enable harmonization of e-infrastructure through enhanced project data planning, monitoring, review and sharing. Current Data Management Plan (DMP) requirements vary greatly across Belmont Forum agencies in their level of detail and compliance with general stewardship principles, including use of *certified, trusted data* repositories. DMPs do not generally include sufficient guidance on current best practice and available e-infrastructure, and how to enable data exploitation. Furthermore, there is often little or no formal accountability to ensure DMPs accomplish what is proposed or to learn from their success or failure.

This Action Theme proposes asking funders to require funded researchers to provide and comply with an Enhanced Data Plan (EDP). The EDP should cover the management of data, connection of the data to associated context (research project, resulting publications, researcher details, data provenance, etc.), facilitating the discoverability of the data, and ensuring that the data is capable of reuse/exploitation. Depending on the discipline area, this might also require the use of *Certified Data* repositories and connections to relevant tools and services.

2. **Objectives, Justification, Impact:** Establishing Belmont Forum good practice through common Enhanced Data Plan requirements and monitoring will improve data availability, interoperability, enable co-evolution with e-infrastructure to ultimately help increase data usefulness, build trust among stakeholders, and reduce costs. The Belmont Forum is ideally positioned to achieve significant impact by changing grant funding policies and reward systems to promote better data planning and implementation.

Description of Actions

Near-Term Actions (start within 0-2 years):

1. Belmont Forum members to **appoint a Data e-Infrastructure Officer (“Champion”) as part of the Coordination Office to oversee Data Plan activities and liaise with the Belmont Forum Principals and appropriate members of the scientific community.** See also Coordination Office Action Theme.
2. Belmont Forum to **adopt a common minimum Enhanced Data Plan (EDP) template developed in conjunction with all Belmont Forum agencies. The template should include evolving guidelines which cover not only data stewardship but also e-infrastructure needs for effective data exploitation.**
3. Coordination Office to **establish a Security Advisory Board SAB to oversee relevant security issues and contribute to the development of the Enhanced Data Plan template.** The SAB would consist of security professionals, covering a broad spectrum of cyber disciplines (including, but not limited to, cloud security, *big data* analytics, user and endpoint security). Ideally, SAB members also would have experience covering government, enterprise and user security, as well as environmental and social sciences.

4. Coordination Office to establish a **Data Policy Advisory Board (DPAB) to oversee relevant legal issues and contribute toward the development of the Enhanced Data Plan template.** The DPAB should consist of funder representatives, experts in data policy, global change researchers, and representatives of data repositories. The DPAB should develop, monitor and review a Data Sharing Policy and Data Usage Policy for the Belmont Forum, taking into account privacy, security, intellectual property and other legal issues. These policies would draw upon the work already undertaken by various organizations, including GEO, ICSU and RDA. It is anticipated that legal issues will arise in three principal areas: 1) the terms on which data will form part of (or will benefit from) e-infrastructure development activities by Belmont Forum members; 2) the manner in which such data will be handled on an ongoing basis; and 3) the terms and conditions on which users will access and use these data.
5. Coordination Office to **feed inputs from the Data Policy Advisory Board and Security Advisory Board to Enhanced Data Plan template** guiding its evolution and development.
6. Belmont Forum members to work with the Coordination Office to **establish mechanisms for regular review and monitoring of Enhanced Data Plan accomplishments and effectiveness.**
7. Belmont Forum members to **assign budget allocation for Enhanced Data Plan implementation, including data curation and publishing costs, as part of their standard grant policy.** This action will incentivize data sharing and reuse, which, in turn, will optimize return on investments made in global research activities for enhanced social and economic development.
8. Belmont Forum members to **extend the current reward system associated with grant awards to include a conditional sharing and sustainable archiving requirement to complete the data lifecycle, and to reinforce successful implementation of Enhanced Data Plans.**

Medium-Term Actions (start within 1-5 years):

9. **Develop and maintain a set of metrics to assess compliance with the Belmont Forum-funded Enhanced Data Plans.**
10. Belmont Forum members to **review and monitor implementation of the Enhanced Data Plans, starting with funded research projects under the Future Earth initiative.**
11. Coordination Office (Data e-Infrastructure Officer) to **identify key repositories - preferably certified, trusted repositories - relevant to the Belmont Forum research agenda to serve as exemplars for the Enhanced Data Plans.** The ICSU World Data System and re3data initiatives were identified as two of the useful resources to identify trusted digital repositories.
12. Coordination Office (Data e-Infrastructure Officer) to initiate a **gap analysis to identify data from Belmont Forum-funded research not stored in trusted repositories, and work with Belmont Forum agencies and other stakeholders, including the research community, to explore ways of filling gaps.**
13. Coordination Office (Data e-Infrastructure Officer) to support the **development of a Belmont Forum research register of software, workflows, policies and standards (including vocabularies, ontologies, data models, data structures and computational interfaces), for improved guidance within the Enhanced Data Plans.** Where existing domain- or disciplinary-specific initiatives have already begun activity in this area, it is advised that they are supported and that duplication of effort is minimized.

14. Coordination Office (Data e-Infrastructure Officer) to **improve Enhanced Data Plan template guidance over time using inputs from: 1) the Data Policy Advisory Board (including legal) and Security Advisory Board; 2) the mapping activities of data repositories and services, standards, tools and best practice; and 3) the e-infrastructure case studies and scoping workshops.**
15. Coordination Office (Data e-Infrastructure Officer) to **maintain Enhanced Data Plans in a freely accessible, searchable resource.**

Relationships with other activities

1. **Relationships with other Action Themes:** Many of the actions suggested under this Action Theme are dependent upon an ongoing Coordination Office, recommended to be housed within the Belmont Forum Secretariat. Without an executive body empowered to initiate activity, there is no clear mechanism for enforcing recommended actions.
2. **Relationships with other Belmont Forum projects:** All projects funded under Belmont Forum CRAs have a relationship with this action, as it is recommended that the Enhanced Data Plans be initiated and harmonized across all Belmont Forum projects.
3. **Relationships with other activities:** There are a variety of Data Management Plan templates already available that could be used as a starting point, with some funding agencies making these mandatory. However, there may be no agency provision for future funding to be dependent on active Enhanced Data Plan development nor for following-up on implementation. Additionally, gaps in requirements for e-infrastructure and exploitation of data, and the lack of guidance provided to proposers, have been identified in existing Data Management Plan templates.

Belmont Forum funding agencies need to work collectively to engage with organizations already involved in this area, such as the ICSU Committee on Data for Science and Technology (CODATA), World Data System (WDS), Research Data Alliance (Interest Group on Active data management plans), the GEO Data Management Principles Task Force, and others including, but not limited to, Force11, FITS, DCC, EarthCube, Journal of Open Research Software, Earth Science Information Partners Federation, and NISO. They should also collaborate and contribute to the RDA Working group on Repository Audit and Certification DSA–WDS Partnership, and the Registry of Research Data Repositories (re3data) in cooperation with ICSU WDS to provide a subset of the registry or an independent extension that aggregates quality and certification properties (WDS, Data Seal of Approval, ISO 16363, etc.).

The UK Natural Environment Research Council (NERC) Rural Economy and Land Use Project produced a report on its Data Support Service that incorporated both natural and social science data that would be a very useful starting point to “develop and learn about the mechanics of data integration and create a legacy, not only of substantive findings, but of methods that can be applied in future research”. The European Commission’s Horizon 2020 open data pilot is a further initiative which will be very helpful in shaping this work ([europa.eu/rapid/press-release IP-13-1257 en.htm](https://europa.eu/rapid/press-release_IP-13-1257_en.htm)).

Justification in Context: *Why are we recommending the Belmont Forum address this issue? Why is the Belmont Forum well or best suited to act in this way?*

1. **Relevance and History:** The research needed to address global change challenges requires integrating vast amounts of heterogeneous data across scientific disciplines and domains. This can happen only if data sets generated are discoverable, accessible, usable, curated and preserved for the long-term,

within a supporting e-infrastructure framework that enables their exploitation and evolves in response to research needs and technological innovation.

Although these issues are variously addressed by individual agencies, coordination at the level of the Belmont Forum would have a strong impact, both for harmonization of activities for tracking best practice through the ongoing improvement, review and sharing of Data Plans. Further, alignment of international and cross-domain efforts in interoperability will support data use today and in the future, and promote new interdisciplinary and international scientific understanding relevant to the Belmont Forum research agenda.

Data Policy Advisory Board: The Belmont Forum needs to minimize legal barriers to facilitate cross-border data sharing and access. Data Sharing must be made easy and be encouraged, minimizing or eliminating any risk of disclosing prohibited information on the part of its users and providers. Ongoing decisions will need to be made concerning metadata, licenses, upkeep, obligatory contributions and other commitments, compliance (with those commitments and internal policies), disputes, costs and funding (including possible cost recovery).

Security Advisory Board: There is a clear, universal need to ensure that data be secured from unauthorized access, tampering and destruction. The Security Advisory Board can ensure security considerations are built into Enhanced Data Plan templates for Belmont Forum-funded research, and more widely. Agencies already operate within national security frameworks, but international exchanges need to consider harmonized best practice if international data and information exchanges are to be trusted and secure.

Anticipated Impact: *For example: influence on policy and practice; impact of adopting or not adopting the recommendation including institutionally, nationally and internationally; reproducible science, data principles.*

1. **Scale and Scope:** Many funding agencies have established and, on occasion, made mandatory, practices for submitting Data Management Plans (DMPs) as part of the research proposal process. However, these practices currently do not include guidance on, or requirements for, e-infrastructure and data exploitation needs. By removing the emphasis on ‘management,’ the scope for these plans (now called ‘Enhanced Data Plans’) is widened to include other important aspects related to more general data requirements. The scale of this action is both international, in terms of the research and its applications, and national, in terms of the anticipated funding of e-infrastructure by individual Belmont Forum members.
2. **Influence:** Actions undertaken by the Belmont Forum to adopt and use community good practices for research data planning will ultimately benefit existing and future Belmont Forum-funded research:
 - a. Improved data management and exploitation practices will enable international multi- and cross-disciplinary approaches needed to respond to the global change challenges as well as unlocking additional economic value from data holdings by facilitating data reuse.
 - b. Recording and reviewing successful Data Plan implementation over time will lead to improved Belmont Forum data processes and advise the development of future e-infrastructures.
 - c. Greater trust in the outputs of research and their applications will be facilitated because data will be more accessible and useful for validation of research results.

- d. Funder mandates can be a useful driver of compliance, by supporting coordinated engagement of various stakeholders across the complex scholarly acknowledgements and rewards system.
- e. Taking an active role in promoting active data stewardship will minimize the current perceived risk that journals remain the sole or defining gatekeepers to research outputs, with the reward system for researchers being regulated by publishers alone.
- f. The Belmont Forum can contribute to addressing the current ineffective data management situation which, at best, leads to research projects delivering unusable data or developing their own data management, exploitation systems and standards (thus increasing the fragmentation). At worst, it leads to making data holdings completely unavailable. All of these current practices seriously limit the impact of the original investment.

Evidence Bank:

Many reports, studies and existing policies support this action and call on funders to play an active role to promote good practices toward open and reusable research data. Below we list the most important ones, starting with the OECD Principles of 2007 and including the G8 Open Data Charter and US White House Executive Order in 2013.

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/open-data-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

The following articles were additionally identified as supporting the recommended actions:

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_fspress

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. This is a useful study on the Value of Research Data, highlighting the benefits of data sharing for scientific progress and the advancement of knowledge. Recommendations of the study call for investigation in the near future of a reward system that includes data metrics, and proposes solutions to the development of data metrics.
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: nbn-resolving.de/urn:nbn:de:bsz:291-universaar-873

The **Open Data Survey** (completed as part of this CRA) included the following relevant question: *“What services or functionalities do you expect from infrastructures?”*

1. 89% of respondents listed ‘Repository certified’ as of either intermediate or highest importance.
2. 89% of respondents also listed ‘Information about tools’ as of either intermediate or highest importance.
3. 92% of respondents listed ‘Code published or accessible’ as of either intermediate or highest importance.

The wish list for open data from the survey additionally included the need for long-tail research data sets to be supported (via repositories) and highlighted a concern that data from smaller projects (e.g., citizen science) was not shared because they do not have the technology.

Several reports and existing policies make evident the need for a Data Policy and Security Advisory component in this Action Theme to inform decisions concerning data sharing, metadata, licenses and other legal issues, and security.

1. **Data Sharing:** EU jurisdictions (Directive 96/9/EC, 1996); Creative Commons (e.g., the Creative Commons Public Domain Waiver (CC0) or the Creative Commons Attribution Only license (CC BY 4.0)), and data usage agreements employed by other data sharing projects, such as the Database of Genotypes and Phenotypes (dbGaP) of the US National Institutes of Health.
2. **Metadata and Interoperability:** Smart Grid Interoperability Panel (SGIP) chartered by the US National Institute of Standards and Technology
3. **Data Security and Integrity:** the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) and national security mandates; The Security White Paper, which was produced by the Environmental Virtual Observatory project, explores the key considerations which are directly relevant to this Belmont Forum initiative: bfe-inf.org/document/evop-cyber-security-whitepaper.

Action Theme 3: e-Infrastructure

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

Executive Summary

- 1. Action Theme Summary:** We need to determine the best ways of sharing, shaping and using collaborative data-intensive e-infrastructure across the Belmont Forum to address global change research challenges. This needs to be driven by scientific requirements; the best way of focusing investments across the Belmont Forum is by using case studies. Case studies of the way researchers are using data and e-infrastructures in large interdisciplinary investigations will determine best practice. Reports from the case studies will be used to inform the development of Enhanced Data Plans, to illustrate weaknesses in existing e-Infrastructures and so provide evidence to advise on the technical implementation of future e-infrastructures.

The Belmont Forum should implement iterative cycles of scoping workshops and international calls for case studies, linked to improved sharing e-infrastructure. The workshops will identify the most critical issues to address, including interoperability at system and service levels, for a variety of cross-disciplinary research applications and communities. This Action Theme will start by using existing cross-disciplinary research projects already being carried out under the Belmont Forum, Future Earth, GEO or other international initiatives and build on them.

Objectives, Justification, Impact: Examples of good practice in data and information sharing, and in developing and using e-infrastructure, are relatively new, uncoordinated and, only rarely have long-term support. Important science, societal and economic questions require changing methods and practices to facilitate cross-disciplinary research conducted around real world problems rather than in silos. Modern ways of carrying out research allow many domains of environmental and social sciences come together in much larger and more diverse research teams than in earlier endeavors. New global change research problems are at the international forefront of research and often push the boundaries of the use of data and e-infrastructure.

This Action Theme will guide cross-disciplinary e-infrastructure development to ensure best practice in sharing data and e-Infrastructure. International case studies and scoping workshops will also be used to determine good-better-best practice across natural, social, health, engineering and economic domains.

Description of Actions

Description of the Action Theme:

Case studies can serve a multitude of purposes. Case studies effectively become high-level workflows that illustrate potential solutions to research and e-infrastructure challenges. It is important to adopt a consistent methodology to select a portfolio of case studies that is responsive to emerging needs as expressed in the Future Earth research programs and funded by the Belmont Forum. *The Belmont Challenge: A Global, Environmental Research Mission for Sustainability* lists a number of challenges and desired responses (referenced in Evidence section) and the *Future Earth 2025 Vision* lists eight focal challenges. A methodology selecting a portfolio of case studies combining the challenges listed in these two documents, together with challenges from other relevant international initiatives, will ensure that investments meet scientific challenges and are not investments for their own sake..

An initial set of **scoping workshops** will be used to identify a set of cross-disciplinary case studies of how researchers are using data and e-infrastructures to ensure that a breadth of science, data-intensive methods and informatics techniques are used. Policies will be developed based on the case studies to define and refine good practice for e-infrastructure and data management, informing e-infrastructure development and national investments. A fast **iterative call-review-call cycle** will cultivate good-better-best practice as well as ensuring results are up-to-date, and illustrating how e-infrastructure investments are best made. Existing Belmont Forum-funded projects will likely meet many of the early case study needs, but further calls may be required to fill gaps identified through the case studies, perhaps through suitable future Belmont Forum program calls. Other programs, such as GEO flagships, may also produce suitable case studies, and the insights of GEO, RDA and other groups can be useful in determining best practice. The set of case studies should include projects demonstrating the integration of both social and natural science, and data and model inter-comparison.

The dynamic human responses to opportunities raised by discoverable, accessible and understandable data are hard to predict, especially as environmental challenges and technologies are all changing rapidly. This Action Theme must therefore be agile and facilitate co-evolution of research needs, methods, tools and services, including the integration of social science constructs.

Near-term (start within 0-2 years)

1. Coordination Office to **appoint an e-Infrastructure Champion** to oversee activities and liaise with the Coordination Office. The Champion will drive the following actions and analyze the results from the scoping workshops and reports from the case studies to enhance the Evaluation Matrix and inform the directions of further Case Studies.
2. e-Infrastructure Champion to **convene a series of scoping workshops under the authority of the Coordination Office** (recommended in this report to be housed under the Belmont Forum Secretariat). This will inform the development of the Evaluation Matrix, which will be used to analyze, score, and identify cross-disciplinary case studies, and the broader scope of the relationship between environmental science, data, and e-infrastructure. The workshops should include a mix of environmental, social and health research scientists, data scientists, data-aware engineers, critical stakeholders (including e-infrastructure providers for data, networking and computing) as well as international policy makers, decision-makers and advisory bodies. The first workshop/call cycle could remain focused on a single or a few selected topics, such as climate change, natural hazards or food security, and related Belmont Forum-funded project topics. Pertinent external stakeholders, such as the Integrated Research on Disaster Risk (IRDR) and the Intergovernmental Panel on Climate Change (IPCC), should be identified by the e-Infrastructure Champion, and be involved as appropriate. The resulting Evaluation Matrix will be used to inform the analysis action.

Example of scoping workshop: Bring together existing Future Earth and other interdisciplinary research projects that currently utilize, or plan to utilize, data from two or more natural, social, health and engineering science domains, and/or created value-added integrated products. Workshop goals are to:

- *Assess the types of data integration, the data analysis workflows used, and the range of archival and dissemination arrangements underway or planned.*
- *Identify gaps, barriers and opportunities to promote more efficient and effective data acquisition, use and integration across domains, and facilitate orchestrated data analysis workflows and modeling across data and computing infrastructures.*

- *Identify ways to improve access, archiving and stewardship of integrated data products working with existing e-infrastructure elements or by developing new elements.*
 - *Identify and assess the social, environmental and economic benefits of improved data interoperability and preservation for the various stakeholders.*
 - *Identify and assess challenges and beacons of good practice in conducting international data and model inter-comparisons.*
3. e-Infrastructure Champion to convene a series of scoping workshops under the authority of the Coordination Office to **analyze e-infrastructure applications of existing Belmont Forum-funded projects and other international initiatives, using the derived Evaluation Matrix to identify critical gaps and barriers and define any required funding calls and priorities for case studies.** The analysis is required for the longer-term actions and will be incorporated into the training developed under the Human Dimensions Action Theme.

Medium-term (start within 1-5 years):

1. Belmont Forum to **hold a three-year competitive funding call for case studies** to establish good practice in prioritized Evaluation Matrix scenarios.
2. e-Infrastructure Champion to convene additional scoping workshops or smaller scoping sessions under the authority of the Belmont Secretariat to **review and analyze reports from case studies to identify strategies to implement best e-infrastructure practices.** Analyses will be used to inform the evolution of e-infrastructure and national investments and will be reflected in subsequent iterative calls, where further strategies may be identified or existing strategies may be enhanced.
3. Belmont Forum members to **define current best practice policy to implement the strategies determined from the analyses** for subsequent cycles of scoping workshops and calls for case studies.

Relationships with other activities

1. **Relationships with other Action Themes:** Analysis will be performed by the e-Infrastructure Champion in collaboration with the Coordination Office. Results will be used to inform the Data Planning and Human Dimensions Action Themes.
2. **Relationships with other Belmont Forum projects:** Other Belmont Forum-funded projects are considered a rich resource from which to derive early case studies.
3. **Relationships with other activities:** Existing international projects funded under other programs may be readily identified (by the Coordination Office or other working group members) to complete further case studies. These links will facilitate acceptance and adoption of Belmont Forum practices in other initiatives.

Other projects could include Future Earth, IRDR, IPCC, Group on Earth Observations (GEO), World Meteorological Organization (WMO), Intergovernmental Oceanographic Commission (IOC), International Council for Science World Data System (ICSU-WDS), funding group on Collaborative Research e-Infrastructures (European Commission) and the Research Data Alliance (RDA).

Justification in Context: *Why are we recommending the Belmont Forum address this issue? Why is the Belmont Forum well or best suited to act in this way?*

1. **Relevance:** Important environmental research and societal questions demand research conducted around interdisciplinary research problems rather than in scientific discipline silos. In many cases data and knowledge available in one discipline is not understood and therefore cannot be applied in other disciplines. The data may not even be discoverable outside the holder's organization or discipline. Decisions are often made without using relevant data that are already available. Urgency frequently forces the use of imperfect and incomplete data, and to trade accuracy for speed, whereas monitoring needs continuous-streaming of data and data analysis methods. Reducing risk by facilitating better use of existing data could yield significant socio-economic benefits.

Integrating data from the natural and social sciences, as well as from the health and engineering domains, is an essential step in delivering information and knowledge about risks, impacts and vulnerabilities associated with global change and potential mitigation and adaptation options, and their potential impacts. The development and orchestration of relevant and shared data-intensive workflows across data and computing infrastructures are critical to make better-informed decisions. All research disciplines will benefit from the improved exploitation of all available data. The Belmont Forum is uniquely placed to define and endorse best practice across disciplines and regions.

2. **History:** Past projects have considered some of the issues mentioned above. However these have been limited in scope, context or geography. For example, within climate modeling, the provision of large global and regional data sets from different institutions in standard formats and through standard gateways is organized through the Earth System Grid Federation (ESGF, esgf.llnl.gov).

Similar projects in other domains include the Data Integration and Analysis System (DIAS, editoria.u-tokyo.ac.jp/projects/dias), Virtual Earthquake and Seismology Research Community in Europe e-science environment (VERCE, verce.eu), Ocean Data Interoperability Platform (ODIP, odip.eu), Connecting research infrastructures in environmental sciences (COOPEUS, coopeus.eu) and ICSU World Data System (ICSU-WDS, icsu-wds.org).

Few organizations or networks exist with the mission, resources or experience to deal with interdisciplinary, integrated data and support innovative use of these data in research, as well as societal applications. There are large gaps in the existing e-infrastructure, but the communities of data users who seek interdisciplinary data and services for research, decision making, policy development, and so on, is expanding rapidly.

Beacons of best practice should be shared more widely and be used to benchmark performance more generally, supporting the full life-cycle of data, and to promote more sophisticated internationally coordinated efforts addressing today's gaps and barriers.

Anticipated Impact: *For example: influence on policy and practice; impact of adopting or not adopting the recommendation including institutionally, nationally and internationally; reproducible science, data principles.*

1. **Scale and Scope:** Global access to data is changing the ways in which we think about and address problems on climate change and environmental hazard and risk. Shared data is catalyzing change in global collaboration and in businesses. The scale of this action is, therefore, global and requires international coordination and collaboration. Even when causes and consequences of environmental

change are expressed at regional, national or local scales, they are of relevance globally, given their potential for triggering global impacts.

Interdisciplinary research and applications developed at local and regional scales will benefit from improved access to data products generated in other regions. International research will additionally benefit from opportunities to compare their more homogeneous approaches with local and regional scale efforts.

A user-driven sustainable data-and-services e-infrastructure needs to be application-oriented, easily accessible, open and agile so that it can continuously adapt to changes in technology and research practice. Case studies must, therefore, include the relevant natural, social, health and engineering elements while identifying gaps in missions, resources and capabilities, and also involving appropriate business models to explore how to sustain long-term stewardship.

2. **Influence:** This cross-disciplinary e-infrastructure action is proposed so the majority of researchers in most disciplines benefit from the wealth of data now being produced. Adoption of best practices will transform the way research is conducted, 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 by tackling a large increase in diversity, volume and rate of multidisciplinary data. It is critically important to establish and enable a cross-disciplinary framework and data-intensive e-infrastructure so that scientific knowledge can transcend disciplines and address new problems.

Global change encompasses both human and natural change. Therefore, we need to combine knowledge of natural sciences, socio-economics and humanities as well as natural environmental sciences. This can only be undertaken through data-intensive e-infrastructures and methods that are scalable via application to other contexts. Cross-disciplinary environmental research challenges will drive the coevolution of data-intensive e-infrastructure, in harmony with research practice and societal needs.

The greatest positive impact of undertaking these actions will be the increase in the number of people accessing data. This should be encouraged by engaging ‘citizen scientists’ and using social media in collecting and annotating data, for example. When data, services and tools that researchers need are discoverable and accessible, they are more likely to be used in problem solving and decision-making.

Evidence Bank:

Supporting publications, reports and data:

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%20Rev2.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.

3. The Belmont Challenge: A Global, Environmental Research Mission for Sustainability, 2011:

igfagcr.org/sites/default/files/documents/belmont-challenge-white-paper.pdf

This document sets out the Belmont Forum's vision for the priority knowledge, capabilities, underpinning research and organizational mechanism needed to deliver effective solutions to the challenges of global environmental change. The vision provides a basis for research funders to broker new partnerships with international stakeholders from the science community, operational service providers and users, to align and mobilize our collective resources and expertise toward a global environmental research mission for sustainability.

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.

New data collected during/for this CRA (e.g. The Open Data Survey, Interviews, Scoping papers etc.)

The following recommendations were made in the paper: "Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey" which arose from the work of this CRA and which reinforce the recommendations presented as part of this action:

1. Funders should make open data archiving mandatory. This has to be enforced at national levels while taking into account that the implementation of policies relies on adequate infrastructures and human support. Therefore infrastructures for data archiving have to be sufficiently supported to secure high-quality data stewardship. This in turn demands international consultations on who is archiving what kind of data, which for global environmental change research could be best achieved through the Belmont Forum. This would help to prevent inefficiencies and multiplication of efforts related to public infrastructures and guidelines for data archiving.
2. Scientific merits as well as accelerating research and applications are still the main motivators for publishing data; thus ethics of data sharing and reuse should be taken into account when proposing guidelines for open data sharing and reuse.
3. Support and training activities should be supported in concerted ways, targeting researchers as well as current and future data and information professionals. Interoperability between infrastructures should be further facilitated which is best achieved via the Belmont Forum. Interoperability should take into account generic requirements (e.g. providing links to publications and funder information) as well as disciplinary norms and standards (e.g. vocabularies, metadata standards).

bfe-inf.org/document/open-data-global-environmental-research-belmont-forum%E2%80%99s-open-data-survey

Action Theme 4: Human Dimensions

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

Executive Summary

1. **Action Theme Summary:** The Belmont Forum should: 1) support the development of a holistic training and education curriculum in data-intensive environmental science, for delivery to environmental, social and computer scientists; and 2) launch the curriculum through the creation of a number of international short courses and immersive winter/summer schools. Whilst there is existing training in a number of areas and domains, it is frequently too restricted in its scope and there is considerable evidence of the shortage of skilled people worldwide who have both a well-developed understanding of environmental, social, and health science and skills and knowledge in data science, data-handling and computational methods and technologies. This curriculum will provide a focus and toolkits for training and knowledge dissemination from international and national experts to those who can further apply the knowledge and experience locally. The courses developed under this curriculum could be 'Belmont Certified', raising awareness of the minimum data and observation standards adopted by the Belmont Forum.
2. **Objectives, Justification, Impact:** The objective of this action is to develop a curriculum that can be replicated to: 1) ensure Belmont Forum researchers incorporate data plans into their work; and 2) attract and prepare a new wave of researchers to conduct data-intensive environmental change research to meet global change challenges, and facilitate their ability to develop new opportunities for cross-disciplinary research through collaboration with existing best practice in data management and data-intensive research.

Description of Actions

1. A scoping workshop will be held to design an overall Short Course Program curriculum. The curriculum will focus on developing skills which support the theory and methods for handling and analyzing environmental, social, and health science data through local and cloud-based tools, data standards and data management best practices for reproducible environmental science, linked to relevant environmental questions. This training will build upon courses delivered in national contexts (such as degrees), and will not only teach the needed skills and ensure common implementation of standards, but will also provide all attendees with a toolkit they can use to deliver and cascade the knowledge and skills through their own communities (both geographical and discipline).
2. The individual courses will be aimed at practicing scientists and graduate students engaged in Belmont Forum associated projects. In the first instance, the aim will be to bring together both young and established scientists from diverse fields (e.g. hydrosphere, atmosphere, geosphere, biosphere social, and health sciences) with information scientists and engineers under the guidance of acknowledged leaders in the fields involved. The courses will also be available to mainstream scientists to improve their use of research results and maintenance of observational data and to those who are stewards of the data in data centers and the like. Policy makers could also be involved in project work, or have access to particular modules, to help them understand and promote the importance of interoperability throughout the whole research and data lifecycle.

Near Term Actions (start within 0-2 years):

1. Coordination Office to **appoint a Human Dimensions Champion** to oversee activities, coordinate with activities to implement the other Action Themes, and liaise with the Belmont Forum funded projects and

the Belmont Forum Secretariat. The **Human Dimensions Champion** should work closely with the Coordination Office to identify existing training initiatives being carried out by Belmont Forum members and others around the globe which address training in this area. Belmont Forum members should agree to a **Policy** of incorporating the evolving short courses into their own funded projects and, therefore, into any training that they are offering. The Champion, along with the Coordination Office, should communicate those courses or materials that are open to external attendees across the Belmont Forum community.

2. Coordination Office to **create and maintain a database, accessible via a website, of Belmont Forum member and other training initiatives in this area** to ensure that those which are already accessible to international students are not duplicated and are promoted more widely.
3. The Human Dimensions Champion and the Coordination Office to **organize a scoping workshop to design the overall** curriculum for a program of short courses as part of what should become a larger global “virtual university of e-infrastructure” which will deliver skills and knowledge in ‘Informatics for Human and Environmental Science’.

The Scoping Workshop should address theory and methods for responsibly producing, handling and analyzing environmental, social and health sciences data through local and cloud-based computing tools, data standards and data management best practices for reproducible sciences. The curriculum should be modular, so that while the whole scheme could be delivered as one course online, in-person delivery could be regional, and individual students could take one module at a time, if desired. The curriculum should draw on the experience of the limited number of successful existing short courses in this area. It is also possible that graduates could take their training from the course, and using online tools, deliver elements to research colleagues in their area or discipline.

4. Belmont Forum to **initiate a competitive funding call** for the delivery of training courses against this curriculum based on existing successful exemplars and on newly proposed courses.
5. An additional **competitive funding call** should be initiated for the design and delivery of short courses which incorporate the recommendations of this report, including both environmental and relevant aspects of social sciences, as well as informatics. The purpose of the call would be to make nationally-funded short courses Belmont Forum compliant, and then allow people in other Belmont Forum member countries to attend those courses if they are qualified and would benefit from the training. Each participant should be responsible for their own costs, whether supported by their Belmont Forum member or from other sources. It would also be beneficial to include funding for students from non-Belmont Forum countries, particularly those from developing countries with little existing national training. An aspect of the training will be to provide data management skills for environmental scientists generally and to support the general responsibility of all scientists to archive data for long-term interoperability. Participants in the short courses could then use the materials and online tools to run other courses to promote good practice and cascade the skills more widely in their respective communities. It could be a condition of funding that they promulgate the training locally.

Medium and Long-term Actions (Begin within 1-5 years):

1. Belmont Forum to initiate a subsequent **competitive funding call to develop online training programs**, building on the existing exemplars and results from previous funding calls, and incorporating ongoing developments in data plans and e-infrastructure, to ensure that trainees have the capabilities and tools to pass on their skills to their local communities. These programs should also be aimed at data stewards in order to share best practice internationally.

Belmont Forum funding should be provided for the continuing provision and updating of the courses, particularly to utilize new and innovative delivery mechanisms. Individual Belmont Forum member agencies may separately allocate funds for appropriate students and scientists to attend international short courses and summer schools to prepare them better to compete in data-enabled activities.

Relationships with other activities

1. **Relationships with other Action Themes:** This activity should build on, use and propagate best practices highlighted by other Action Themes. Data plans are an underpinning element, and the utilization of the case studies in the e-Infrastructure Action Theme as exemplars for students to employ and learn from will be very desirable. The appointed Human Dimensions Champion will work closely with the Coordination Office to identify existing training initiatives being carried out by Belmont Forum members around the globe which address training in this area.
2. **Relationships with other Belmont Forum projects:** The training developed under this action (and best practices utilized) should be made available throughout all Belmont Forum organizations and projects.
3. **Relationships with other activities:** Working with external organizations should ensure that initiatives are not duplicated and training best practice is promulgated. Participants from non-Belmont Forum organizations should be welcomed in these courses, subject to their costs being covered. Conversely, if external organizations already offer high quality training in relevant areas, then Belmont Forum members can make funding available for their researchers to attend.

Justification in Context: *Why are we recommending the Belmont Forum address this issue? Why is the Belmont Forum well or best suited to act in this way?*

1. **Relevance:** Researchers are fully stretched just trying to address challenges in their own field. They are naturally focused on their own scientific problems and are competitive with colleagues as well as collaborative. Researchers are, therefore, reluctant to be diverted into understanding new methods and research practices, particularly if they are unsure of their benefits. Consequently, the Belmont Forum must invest in education that allows international researchers to engage incrementally. Examples include: best-practice guides for those planning data-intensive research including use of High Performance Computing and *commodity computing*; packages for generating metadata; automated metadata and provenance generation; frameworks for discovering and exploring data; aids for extracting selected data and delivering them in standard forms and working among differing ontologies; and facilities for data archiving and publishing. This Action Theme addresses a major gap identified in the Open Data Survey questionnaire. The Belmont Forum can target a large international and interdisciplinary pool of leading scientists and educators, thus obtaining an impact that no single national funding agency can achieve. Cascading the training from Belmont Forum courses nationally will have long lasting impacts that will influence a much wider population.
2. **History:** Successful exemplars exist (e.g., the Venice Summer School on Biogeodynamics and Earth System Science (BESS): istitutoveneto.org/bess/program/; The San Diego Supercomputer Center (UCSD) Summer Institute: sdsc.edu/Events/summerinstitute/; UNITAR's Multilateral Diplomacy and Climate Change Programs that provide an online training environment for diplomats and policy makers preparing for the UN Framework Convention on Climate Change; the GEO-CAB web portal: geocab.org/. [DataONE webinars and seminars on best practices and investigator toolkits: dataone.org/best-practices](http://dataone.org/best-practices), and others). These exemplars have demonstrated that a large cross-section of diverse educators, policy makers and young scientists active in environmental sciences, social and health sciences can be reached,

sometimes with a lasting impact in terms of peer-reviewed publications which can reach an even wider community of practice. While the whole curriculum may be of interest and benefit to early career scientists, specific modules may offer opportunities for skills development in particular areas to more established professionals.

Anticipated Impact: Example: influence on policy and practice; impact of adopting or not adopting the recommendation including institutionally, nationally and internationally; reproducible science, data principles.

1. **Scale and Scope:** This action will have international reach through wide participation at local, but interconnected, sites. Initial delivery of short course training will be face-to-face, but by using web-based delivery methods, these efforts can be shared with students in widely distributed venues. Alumni will also be provided with toolkits and access to online material that will enable them to pass on their knowledge and training to local colleagues. This online material also will enable those unable to travel to international venues to access the training and knowledge.

National initiatives cannot reach a large enough group of scientists, nor can they involve a wide enough population of appropriate international subject matter leaders. Only a coordinated action of multinational funding agencies can create the critical mass of competent scientists and ensure the initiative can have a significant and measurable impact. Existing exemplars have relied on funding by learned societies or by national or university sources. The Belmont Forum can ensure a systematic action which is otherwise impossible, as a resource to complement widely scattered and uneven training activities.

2. **Influence:** Responses to the Open Data Survey, carried out as part of this CRA, indicate that suitably trained people are recognized as a key element in improving global data management, stewardship, preservation and e-infrastructure interoperability. . This Action Theme will have a long lasting influence as it affects people's career trajectories. Some existing short course programs include projects which lead to joint publications across nations and disciplines. Through their readership, these publications can promote the knowledge and good practice and have far reaching impacts beyond simply those who attended the courses. The face-to-face courses should, if possible, utilize this mechanism to leverage the influence on the fields involved, in terms of publication products and in interdisciplinary and international collaborations. Policy makers can be involved through targeted activities with students or by taking particular modules within the proposed short courses. For continuity, students from previous sessions (alumni) can revisit subsequent courses to help with projects and instruction, and to cascade to nationally delivered courses building on the Belmont Forum program. This also has the benefit of allowing further cross communication among disciplinary areas.

Evidence Bank:

A number of countries have carried out skills gap reviews amongst their environmental scientists to consider their long term strategic requirements. They include:

- 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)

- 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)

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.

The Venice Summer Schools on Biogeodynamics and Earth System Science (BESS) have been running since 1989, involve both environmental scientists and quantitative computer scientists and have evolved a very successful format to address the problems considered here while attracting high quality students. Over 500 students from 36 different countries have participated, supported by 60 leading researchers from around the world. Following their participation, evidence shows that over 60% of attendees are active in research, with the remainder in operational agencies, and with much unmet demand from potential students.

The UN, through UNOSAT and UNITAR, deliver both online and face to face training to diplomats, policy makers, educators and young scientists. UNOSAT’s more technical courses are delivered face to face, in a diverse range of locations around the globe. Between 2012 and 2014, 30 courses were given to over 400 students in 10 different locations, with considerable unmet demand. 35% of participants are from Africa, 31% Asia, 6% S America, 28% Europe with a small, growing interest from the Middle East.

APPENDIX 2: ACTION DEPENDENCIES

All Action Themes proposed in this *Community Strategy and Implementation Plan* include elements which are dependent upon a central coordinating body for implementation, which is referred to in this report as the Coordination Office. It was recognized that the Belmont Forum principals will require a separate operational agent or agency housed within the Belmont Forum Secretariat to undertake internationally coordinated action on behalf of the Belmont Forum.

The Data Planning Action Theme will influence all future funding calls under the Belmont Forum, and will, therefore, have an impact on the other Action Themes, which include recommendations for funding call actions.

The e-Infrastructure Action Theme will advise training developed under the Human Dimensions Action Theme, although is not dependent upon it. It will also inform the Data Planning Action Theme.

The Human Dimensions Action Theme will be informed by, and is therefore partially dependent upon, the case studies identified in the e-Infrastructure Action Theme. The Data Planning Action Theme additionally underpins recommendations within the Human Dimensions Action Theme.

APPENDIX 3: TRANSITION AND IMPLEMENTATION

July 2015-July 2016 and beyond

Purpose, Goals and Organization

A transition period lasting from the delivery of this *Community Strategy and Implementation Plan* (CSIP) on June 30, 2015 through to the end of December 2015 is recommended to ensure that the hard work of the scoping process carried out as part of the e-Infrastructures and Data Management Collaborative Research Action (CRA) does not simply stall at the end of June. This report recommends an implementation period beginning in January 2016 to move the current e-Infrastructures and Data Management CRA Secretariat functions to a Coordination Office housed within the Belmont Forum Secretariat, to be completed by the end of July 2016. This report recommends that while some of the tasks under the purview of the Coordination Office could be outsourced externally, the overall 'function' should be housed internally to the Belmont Forum because it will develop and integrate e-infrastructure and data management activities critical to the success of all Belmont Forum projects and goals.

These tasks in the initial transition period would be carried out by the existing project Secretariat funded by the US National Science Foundation (NSF) and the UK Natural Environmental Research Council (NERC), with guidance and assistance from the Steering Committee (SC) and Belmont Forum Group of Program Coordinators (GPC). Participation of SC members is contingent on each Belmont Forum member continuing their involvement and any required support. It is not expected that the Steering Committee will need to meet in person during the transition period, but we recognize the Belmont Forum may request additional tasks which may require such meeting(s). It is proposed that the Assembly should dissolve and be thanked at the end of June 2015, having completed its task.

Any work carried out by the CRA Secretariat and SC beyond December 2015 is dependent on the outcomes of the Belmont Forum Annual Meeting in October 2015 in Oslo, Norway.

Near-Term Transition Period: July 2015 - December 2015

A transition period is needed to conduct community outreach and engagement to inform the Belmont Forum community and others about the CSIP, solicit broader feedback on the CSIP, assist the Belmont Forum (as a group and as individual members) in preparing for the Belmont Forum Annual Meeting, update and maintain the Knowledge Hub, and assist the Belmont Forum in the implementation of the actions recommended in this CSIP.

Our recommendation is that the current Secretariat, SC, and GPC for the e-Infrastructure and Data Management CRA carry out transition tasks as described in the next section. However, we recommend that the Assembly (composed of the national delegations) dissolve after the delivery of the CSIP on June 30, 2015. Assembly members have participated based on goodwill, and their mandate expires at the end of June. There is not

currently a clear role for the Assembly as a whole in the transition phase, but they could play a role nationally, depending on the needs of individual Belmont Forum member agencies.

Current Secretariat for the e-Infrastructure and Data Management CRA

Collectively, NSF and NERC (as the CRA Theme Program Office, TPO), and the Arizona Geological Survey and University of Reading (as the joint US-UK Secretariat), are in a position to continue to support some level of activity in the transition period. Anticipated tasks include the following:

1. Conduct active community outreach and engagement (virtually and in-person)
 - a. Create a public version of the CSIP for publishing on the Knowledge Hub
 - b. Publish articles about the CSIP, and the opportunity for public feedback, in scientific periodicals, such as AGU's Eos
 - c. Describe and promote the CSIP and associated products on the Knowledge Hub
 - d. Create mechanisms to compile and synthesize public feedback on the CSIP
 - e. Create outreach materials that SC members and GPC members can use to engage their respective communities
 - f. Host engagement activities at professional conferences, such as held by AGU, EGU, RDA and other relevant events, including COP 21, Future Earth and others; and engage with stakeholders individually, as appropriate. Activities may include town halls, technical sessions, exhibit booths, webinars, video presentations, meetings and others, as appropriate
 - g. Leverage social media, including Twitter, to draw traffic to the project Knowledge Hub, and solicit broader feedback on the CSIP
2. Assist the Belmont Forum in preparing for the Oslo Meeting in October 2015
 - a. Assist the TPO in preparing and circulating briefing materials, PowerPoint presentations and other materials for use prior to or during the Oslo Meeting
 - b. Prepare slides and other informational products for Belmont Forum members to use with their stakeholders and constituencies
 - c. Work with the SC and GPC to assist individual Belmont Forum members in preparing position papers on how CSIP actions could be implemented nationally; these will be prepared and circulated prior to the Oslo Meeting
 - d. Work with the SC and GPC to identify estimated timelines, resources, and potential implementation mechanisms for the Belmont Forum to carry out the proposed actions
 - e. Provide coordination mechanisms for individual Belmont Forum agencies to indicate their interest in supporting specific actions, and connect with other agencies also interested in supporting those actions
3. Plan Project Coordination Events
 - a. Organize virtual and in-person (if needed) meetings of the SC and GPC
4. Maintain and update the project Knowledge Hub

Steering Committee for the current e-Infrastructure and Data Management CRA

In the fifth Steering Committee meeting held in Tokyo, Japan in April 2015, SC and GPC members agreed that the SC could continue in an advisory capacity during the transition and implementation period (to end no later than June 30, 2016). Continued participation of SC members and other particularly well engaged assembly members is dependent on their personal availability, and availability of funds to cover any travel expenses (should the need for in-person SC or other meetings arise). Each SC member is encouraged to have a conversation with her or his respective GPC member regarding her or his continued involvement. Anticipated tasks include the following:

1. Work with the e-Infrastructure and Data Management CRA Secretariat and GPC to prepare for the annual Belmont Forum Annual Meeting
2. Serve as subject matter experts and points of contact in their respective countries or organizations
3. Participate in virtual meetings with the Secretariat and/or GPC as needed
4. Assist the Secretariat in conducting community outreach and engagement
5. Attend in-person events as needed
6. Work with GPC members to identify what actions each member agency is willing to support

Group of Program Coordinators for the current e-Infrastructure and Data Management CRA

1. Advise the Belmont Forum on addressing issues identified through public feedback on the CSIP
2. Work with the Secretariat and SC to prepare for the Belmont Forum Annual Meeting
3. Work with the Secretariat and SC to assist Belmont Forum members in preparation of position papers on implementing the CSIP actions nationally
4. Begin to act on Annual Meeting outcomes (after October 2015)
5. Potentially initiate national meetings to inform the communities and stakeholders about the outcome of the Belmont Forum Annual Meeting, depending on the needs of individual Belmont Forum members and associated GPC representatives.
6. Work with Belmont Forum Principals and GPC to identify and share which actions individual funding agencies are willing to support; connect with other agencies that are interested in supporting similar actions

Implementation Period: January - June 2016

Depending on the outcomes of the 2015 Belmont Forum Annual Meeting, the original CRA Secretariat and SC may be needed to assist and advise the GPC and the Belmont Forum Principals on implementing the actions agreed upon at this meeting. Activities would likely focus on setting up and transferring responsibilities to the proposed Coordination Office, and filling leadership positions and associated advisory boards and subgroups. The Secretariat and SC would gradually transfer ownership of tasks (community outreach, managing the Knowledge Hub and social media, etc.) to the Coordination Office.

Medium-Term: July 2016 - June 2020

By this time, the Coordination Office should be operational and should be coordinating activities under the purview of the Belmont Forum, including activities outlined in the Action Themes (Coordination Office, Data Planning, e-Infrastructure and Human Dimensions) described in detail in Appendix I.

APPENDIX 4: RELEVANT ORGANIZATIONS AND PROGRAMS

These organizations and programs are referenced previously in this report.

AGU - American Geophysical Union - sites.agu.org

The purpose of the AGU is to promote discovery in Earth and space science for the benefit of humanity. It was established in 1919 by the U.S. National Research Council and operated as an unincorporated affiliate of the U.S. National Academy of Sciences for more than 50 years. The AGU is now an international non-profit scientific association with more than 62,000 members.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities - openaccess.mpg.de/Berlin-Declaration

Open access means that scientific literature should be publicly available, free of charge on the internet so that those who are interested can read, download, copy, distribute, print, search, refer to, and, in any other conceivable legal way, use full texts without encountering any financial, legal or technical barriers other than those associated with Internet access itself. The Max Planck Society is a co-founder of the international Open Access movement. The publication of the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" in October 2003 and the subsequent annual conferences heralded the introduction of a process that heightened awareness around the theme of accessibility to scientific information.

BESS - Summer School on Biogeodynamics and Earth System Sciences - www.istitutoveneto.org/bess/program.

The Summer School on Biogeodynamics and Earth System Sciences is organized by the Venice Institute of Science, Letters and Arts, and addresses the scientific issues posed by a deep understanding of the coupled dynamics of biotic and abiotic components of the Earth System, e.g. related to climate-biosphere interactions, the bio-geomorphic processes that shape the landscape, the two-way interaction between atmospheric and biospheric processes at the surface. The Summer School continues a tradition of Summer Schools on the Environmental Sciences organized by IVSLA for more than 20 years, and has seen the participation of about 60 high-level scientists and of more than 500 attendees.

Bouchout Declaration for Open Biodiversity Knowledge Management - bouchoutdeclaration.org/declaration/

The Bouchout Declaration is an opportunity for those organizations, initiatives and individuals who create, manage and use biodiversity information, and who believe in the opportunities and potential of the big data world, to declare their support of the Open Access agenda. By endorsing the principles of Open Access and discoverability of data, the signatories strengthen the arguments that will be put to governments and funding bodies, and will accelerate the maturation and evolution of Open Biodiversity Knowledge Management, making biodiversity sciences more relevant, innovative, and responsive to societal needs. The project is funded under the European Union's Seventh Framework Program.

CODATA - Committee on Data for Science and Technology - codata.org

CODATA was established in 1966 by the International Council for Science (ICSU) to promote and encourage, on a worldwide basis, the compilation, evaluation and dissemination of reliable numerical data of importance to science and technology. It works to improve the quality, reliability, management and accessibility of data of importance to all fields of science and technology. It is a resource that provides scientists and engineers with access to international data activities for increased awareness, direct cooperation and new knowledge.

COOPEUS - Cooperation EU/US, Connecting research infrastructures in environmental sciences - coopeus.eu

COOPEUS is a project funded under the Research Infrastructures action of the 7th Framework Program for Research and Innovation of the EU, with the goal of bringing together scientists and users being involved in Europe's major environmental related research infrastructure projects, namely EISCAT, EPOS, LifeWATCH, EMSO, and ICOS, with their US counterparts that are responsible for the NSF funded projects AMISR, EarthScope, DataONE, OOI and NEON.

COP21 - 21st Conference of the Parties to the United Nations Framework Convention on Climate Change, or the "2015 Paris Climate Conference" - cop21.gouv.fr/en

The 21st session of the Conference of the Parties to the United Nations Framework Convention on Climate Change (UNFCCC) will be held November/December 2015 in Paris, bringing together around 40,000 participants in total including delegates representing each country, observers and civil society members. The aim of the conference is to reach a universal, legally binding agreement that will enable the UNFCCC to combat climate change effectively and boost the transition towards resilient, low-carbon societies and economies.

Creative Commons - <http://creativecommons.org/>

Creative Commons is a nonprofit organization that enables the sharing and use of creativity and knowledge through free legal tools. It provides free, easy-to-use copyright licenses to provide a simple, standardized way to give the public permission to share and use creative work — on conditions of the content generator's choice.

DataONE - Data Observation Network for Earth - dataone.org

DataONE is the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data. DataONE is a collaboration among many partner organizations, and is funded by the U.S. National Science Foundation under a Cooperative Agreement.

Data Seal of Approval (DSA) - datasealofapproval.org/en

The Data Archiving and Networked Services (DANS) is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and is supported by the Netherlands Organization for Scientific Research (NWO). When DANS was established in 2005 it was assigned the task of developing a Seal of Approval for data, to ensure that archived data can still be found, understood and used in the future. The first edition of the DSA was presented in an international conference in 2008. The objectives of the DSA are to safeguard data, ensure high quality and guide reliable management of data for the future without requiring the implementation of new standards, regulations or high costs.

dbGaP - The database of Genotypes and Phenotypes - ncbi.nlm.nih.gov/gap

The dbGaP database was developed by the U.S. National Center for Biotechnology Information (NCBI) to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype. NCBI was established in November 1988, as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH).

DCC - Digital Curation Centre - www.dcc.ac.uk

The DCC is a center of expertise in digital information curation with a focus on building capacity, capability and skills for research data management across the UK's higher education research community. It provides expert advice and practical help to anyone in UK higher education and research wanting to store, manage, protect and share digital research data. It also provides access to a range of resources, including How-to Guides, case studies and online services, and training programs.

DIAS - Data Integration and Analysis System - editoria.u-tokyo.ac.jp/projects/dias

DIAS was launched in 2006 as part of the Earth Observation and Ocean Exploration System, which is one of five National Key Technologies defined by the 3rd Basic Program for Science and Technology of Japan. The mission of DIAS is: to coordinate the cutting-edge information science and technology and the various research fields addressing the Earth's environment; to construct data infrastructure that can integrate earth observation data, numerical model outputs, and socio-economic data effectively; to create knowledge enabling us to solve the Earth's environment problems; and to generate socio-economic benefits.

EarthCube - earthcube.org

EarthCube enables geoscientists to address the challenges of understanding and predicting a complex and evolving solid Earth, hydrosphere, atmosphere, and space environment systems by fostering community-governed efforts that develop a common cyberinfrastructure for the purpose of collecting, accessing, analyzing, sharing and visualizing all forms of data and related resources, through the use of advanced technological and computational capabilities. EarthCube began in 2011 as a joint initiative between the U.S. National Science Foundation Directorate for Geosciences (GEO) and the Division of Advanced Cyberinfrastructure (ACI).

EGU - European Geosciences Union - egu.eu

The EGU is Europe's premier geosciences union, dedicated to the pursuit of excellence in the geosciences and the planetary and space sciences for the benefit of humanity, worldwide. It was established in September 2002 as a merger of the European Geophysical Society (EGS) and the European Union of Geosciences (EUG). It is a non-profit international union of scientists with over 12,500 members from all over the world.

e-IRG - e-Infrastructure Reflection Group - e-irg.eu

The e-IRG is a coordination and support action (CSA) funded within the H2020 framework. The e-Infrastructure Reflection Group was founded to define and recommend best practices for the pan-European electronic infrastructure efforts. It consists of official government delegates from all the EU countries. The e-IRG produces white papers, roadmaps and recommendations, and analyzes the future foundations of the European Knowledge Society.

ESFRI - European Strategy Forum on Research Infrastructures - ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

ESFRI was established in 2002 to support a coherent and strategy-led approach to policy-making on research infrastructures in Europe, and to facilitate multilateral initiatives leading to the better use and development of research infrastructures, at the EU and international levels. ESFRI's strategy aims to overcoming the fragmentation of individual policies and provides Europe with the most up-to-date Research Infrastructures, responding to the rapidly evolving Science frontiers, advancing also the knowledge-based technologies and their extended use. It published a Roadmap for pan-European research infrastructures in 2006 (updated 2008 and 2010).

ESGF - Earth System Grid Federation - esgf.llnl.gov

The ESGF Peer-to-Peer (P2P) enterprise system is a collaboration that develops, deploys and maintains software infrastructure for the management, dissemination, and analysis of model output and observational data. ESGF's primary goal is to facilitate advancements in Earth System Science. It is an interagency and international effort led by the US Department of Energy (DOE) and co-funded by other US and non-US organizations. ESGF P2P was developed by computer and climate scientists to serve climate data from multiple locations and sources (such as model simulations, satellite observations, and reanalysis products) for worldwide distribution.

ESIP- Earth Science Information Partnership - esipfed.org

The ESIP Federation is an open, networked community that brings together science, data and information technology practitioners. Practitioners work together on interoperability efforts across Earth and environmental science allowing self-governed and directed groups to emerge around common issues, ebbing and flowing as the need for them arises. ESIP was founded in 1998 by NASA in response to a National Research Council (NRC) review of the Earth Observation System Data and Information System (EOSDIS).

European Commission - ec.europa.eu/index

The European Commission is the European Union's executive body. It represents the interests of the European Union as a whole (not the interests of individual countries).

European Union- europa.eu/about-eu/index

The EU is an economic and political partnership between 28 European countries that together cover much of the continent to create a single market for Europe and reduce barriers to economic growth.

EVO - Environmental Virtual Observatory Pilot - evo-uk.org

EVOp was a UK NERC-funded initiative to apply modern cloud computing to illustrate how data, models and tools could be served to environmental scientists and other users in the soil and hydrology community. The pilot was completed in 2014, and results are now being implemented in the UK.

FITS - File Information Tool Sets - projects.iq.harvard.edu/fits/tools

FITS identifies, validates and extracts technical metadata for a wide range of file formats. It was produced at Harvard University.

Force11 - The Future of Research Communications and e-Scholarship - force11.org

Force11 is a community of scholars, librarians, archivists, publishers and research funders from around the world that has arisen organically to help facilitate the change toward improved knowledge creation and sharing. They aim to bring about a change in modern scholarly communications through the effective use of information technology.

Future Earth - futureearth.org

Future Earth is a major international research platform providing the knowledge and support to accelerate our transformations to a sustainable world. It will be an international hub to coordinate new, interdisciplinary approaches to research as well as being a platform for international engagement to ensure that knowledge is generated in partnership with society and users of science. The Governing Council

of Future Earth is composed of members of the Science and Technology Alliance for Global Sustainability, an informal international partnership that is committed to making full use of science and technology to inform equitable, sustainable solutions to the most pressing issues currently confronting humankind. Future Earth is open to scientists of all disciplines, natural and social, as well as engineering, the humanities and law.

Future Earth 2025 Vision - futureearth.org/media/future-earth-2025-vision

The Future Earth 2025 Vision outlines what Future Earth will contribute over the coming decade to achieve its vision for people to thrive in a sustainable and equitable world. It sets out an ambitious, holistic framework for research and capacity mobilization. Central to achieving the vision is a commitment to co-design and co-produce knowledge in collaboration with societal partners, in order to develop solutions-oriented research that responds to the sustainability challenges facing society.

G8 - The Group of Eight (Industrialized Nations) - g8.co.uk

The G8 is an assembly of world leaders who meet annually to discuss global issues and discusses and creates global policies. Adherence to these policies is not obligatory, and other countries can decide whether or not to conform. Russia is temporarily suspended as a member, so the current active membership comprises seven countries (Canada, France, Germany, Italy, Japan, United Kingdom and United States) and the European Union.

G8 Open Data Charter - gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

The G8 recognized that open data are an untapped resource with huge potential to encourage the building of stronger, more interconnected societies that better meet the needs of their citizens and allow innovation and prosperity to flourish. In June 2013, they published a set of principles that will be the foundation for access to, and the release and reuse of, data made available by G8 governments.

GBIF - Global Biodiversity Information Facility - gbif.org

GBIF is an international open data infrastructure, funded by governments. It allows anyone, anywhere to access data about all types of life on Earth, shared across national boundaries via the Internet.

GEO - Group on Earth Observations - earthobservations.org

Established in 2005, GEO is a voluntary partnership of governments and organizations that envisions “a future wherein decisions and actions for the benefit of humankind are informed by coordinated, comprehensive and sustained Earth observations and information.” GEO membership includes 96 nations, the European Commission, and 87 participating organizations comprised of international bodies with a mandate in Earth observations.

GEO-CAB Web Portal - Earth Observation Capacity Building Portal - <http://www.geocab.org/>.

The main objective of this portal is to provide freely accessible information through a web resource facility to help decision makers, policy makers and users of earth observation products to become more familiar with earth observation applications and opportunities.

GEOSS - The Global Earth Observation System of Systems - earthobservations.org/geoss.php

The GEO community is creating a Global Earth Observation System of Systems (GEOSS) that will link Earth observation resources worldwide across multiple societal benefit areas - agriculture, biodiversity, climate,

disasters, ecosystems, energy, health, water and weather - and make those resources available for better informed decision-making.

GFBio - German Federation for Biological Data - gfbio.org

The goal of GFBio is to provide a sustainable, service oriented, national data infrastructure facilitating data sharing and stimulating data-intensive science in the fields of biological and environmental research. The federation will build on proven data archiving infrastructures and workflows, and the resources of Germany's major natural history collection data repositories. The new infrastructure will improve and integrate these existing components within a common technological and organizational framework.

HIPAA - Health Insurance Portability and Accountability Act of 1996 (US) - <http://www.hhs.gov/ocr/privacy/index.html>

HIPAA is the US federal Health Insurance Portability and Accountability Act of 1996. The primary goal is to make it easier for people to keep health insurance, protect the confidentiality and security of healthcare information and help the healthcare industry control administrative costs.

Horizon 2020 - <http://ec.europa.eu/programmes/horizon2020/>

Horizon 2020 is the European Union Research and Innovation program, with nearly €80 billion of funding available over 7 years (2014 to 2020). The goal is to ensure Europe produces world-class science, removes barriers to innovation and makes it easier for the public and private sectors to work together in delivering innovation. Many of its programs concern global change and involve large amounts of data.

iCORDI - International Collaboration on Research Data Infrastructure - epcc.ed.ac.uk/projects-portfolio/icordi-international-collaboration-research-data-infrastructure

iCORDI aims to be the premium global forum driving convergence between emerging global data infrastructures with a particular focus on Europe and the US. Its prime objective is to serve as an EU-US coordination platform chartered with improving the interoperability of today's and tomorrow's scientific data infrastructures on both continents.

ICSU-WDS - International Council for Science World Data System - icsu-wds.org

The WDS is an Interdisciplinary Body of the International Council for Science (ICSU) created by its 29th General Assembly in Maputo, Mozambique in 2008. Its goals include: enabling universal and equitable access to quality-assured scientific data, data services, products and information; ensuring long term data stewardship; fostering compliance to agreed-upon data standards and conventions; and providing mechanisms to facilitate and improve access to data and data products.

IEAG - Independent Expert Advisory Group - undatarevolution.org

In 2014, UN Secretary-General Ban Ki-moon named an Independent Expert Advisory Group on the Data Revolution for Sustainable Development. The 24 experts from civil society, private sector, academia, governments and international organizations is expected to assess new opportunities linked to innovation, technical progress and the surge of new public and private data providers to support and complement conventional statistical systems and strengthen accountability at the global, regional and national levels.

INSPIRE - Infrastructure for Spatial Information in the European Community - inspire.ec.europa.eu

The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure. This will enable the sharing of environmental spatial information among public sector organizations and better facilitate public access to spatial information across Europe.

IOC - Intergovernmental Oceanographic Commission - ioc-unesco.org

The purpose of the Commission is to promote international cooperation and to coordinate programs in research, services and capacity-building, in order to learn more about the nature and resources of the ocean and coastal areas and to apply that knowledge for the improvement of management, sustainable development, the protection of the marine environment, and the decision-making processes of its Member States.

IPCC - Intergovernmental Panel on Climate Change - ipcc.ch

The IPCC was established by the United Nations Environment Program (UNEP) and the World Meteorological Organization (WMO) in 1988 to provide the world with a clear scientific view on the current state of knowledge in climate change and its potential environmental and socio-economic impacts. The IPCC reviews and assesses the most recent scientific, technical and socio-economic information produced worldwide relevant to the understanding of climate change. Because of its scientific and intergovernmental nature, the IPCC embodies a unique opportunity to provide rigorous and balanced scientific information to decision makers.

IRDR - Integrated Research on Disaster Risk - irdrinternational.org

IRDR is a global, multi-disciplinary approach to dealing with the challenges brought by natural disasters, mitigating their impacts, and improving related policy-making mechanisms. It is a research program co-sponsored by the International Council for Science (ICSU), the International Social Science Council (ISSC), and the United Nations International Strategy for Disaster Reduction (UNISDR).

ISO 16363 - www.iso.org/iso/catalogue_detail.htm?csnumber=56510

ISO 16363:2012 defines a recommended practice for assessing the trustworthiness of digital repositories. It is applicable to the entire range of digital repositories and can be used as a basis for certification.

Journal of Open Research Software - openresearchsoftware.metajnl.com

The Journal of Open Research Software (JORS) features peer reviewed software papers describing research software with high reuse potential. They are working with a number of specialist and institutional repositories to ensure that the associated software is professionally archived, preserved, and is openly available. The software and the papers will be citable, and reuse will be tracked. JORS also publishes full-length research papers that cover different aspects of creating, maintaining and evaluating open source research software. The aim is to promote the dissemination of best practice and experience related to the development and maintenance of reusable, sustainable research software.

NASA - National Aeronautics and Space Administration - nasa.gov

The National Aeronautics and Space Administration (NASA) is the agency responsible for the US civilian space program and aeronautics and aerospace research.

NERC - Natural Environment Research Council (United Kingdom) - nerc.ac.uk

NERC is the UK's largest funder of independent environmental science, training and innovation, delivered through universities and research centers. It is one of the sponsors of this CRA.

NISO - National Information Standards Organization - niso.org

Since its inception in the 1930s (as a committee created by the American Standards Association and the American Library Association), NISO is where content publishers, libraries, and software developers turn for information industry standards that allow them to work together. Through NISO, all of these communities are able to collaborate on mutually accepted standards. In 2013, a grant has enabled NISO to undertake further work to explore, identify, and advance standards and/or best practices related to a new suite of potential metrics in the community.

NSF - National Science Foundation (United States) - nsf.gov

The NSF is an independent federal agency created by Congress in 1950 to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense. It funds approximately 24% of all federally supported basic research conducted by America's colleges and universities. In many fields such as mathematics, computer science and the social sciences, NSF is the major source of federal backing. It is one of the sponsors of this CRA.

ODIP - Ocean Data Interoperability Platform - odip.eu

ODIP includes all the major organizations engaged in ocean data management in EU, US, and Australia. It: contributes to the removal of barriers hindering the effective sharing of data across scientific domains and international boundaries; organizes international workshops to foster the development of common standards and develop prototypes to evaluate and test selected potential standards and interoperability solutions; provides a forum to harmonize the diverse regional systems; and facilitates the dissemination of best practice and the transfer of technology, through the development of international cooperation.

OECD Principles of 2007 - Organization for Economic Co-operation and Development Principles and Guidelines for Access to Research Data from Public Funding – oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm

The mission of the OECD is to promote policies that will improve the economic and social well-being of people around the world. It provides a forum in which governments can work together to share experiences and seek solutions to common problems. Science and technology ministers called on the OECD in 2004 to develop a set of guidelines based on commonly agreed principles to facilitate cost-effective access to digital research data from public funding. The “Principles and Guidelines for Access to Research Data from Public Funding” are the outcome of this request. They are intended to assist all actors involved when trying to improve the international sharing of, and access to, research data.

OneGeology - onegeology.org

The OneGeology concept emerged in 2006. Geological surveys and geoscientists around the world have a responsibility to: make accessible the best geological map and other geoscience data; work towards consistent standards for data access – schematic interoperability; and enhance and increase the use and usability of their data.

OSTP - The Office of Science and Technology Policy - whitehouse.gov/administration/eop/ostp

The US Congress established the Office of Science and Technology Policy in 1976 with a broad mandate to advise the US President and others within the Executive Office of the President on the effects of science and technology on domestic and international affairs.

PCAST - President's Council of Advisors on Science and Technology
whitehouse.gov/administration/eop/ostp/pcast

PCAST is an advisory group of the United States' leading scientists and engineers who directly advise the President and the Executive Office of the President (OSTP). PCAST makes policy recommendations in areas at the intersection of science, technology, and innovation that are critical to strengthening the economy and making sound policy.

RDA - Research Data Alliance - rd-alliance.org

The RDA builds the social and technical bridges that enable open sharing of data. Its vision is researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society. The RDA is supported by the European Commission, the U.S. NSF and other US agencies, and the Australian Government.

re3data - Registry of Research Data Repositories - re3data.org

re3data.org is a global registry of research data repositories that covers research data repositories from different academic disciplines. It presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. re3data.org promotes a culture of sharing, increased access and better visibility of research data. The registry went live in autumn 2012 and is funded by the German Research Foundation (DFG).

The Royal Society- royalsociety.org/

The Royal Society, based in the United Kingdom, is a self-governing fellowship, founded in 1660, of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose is to recognize, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

SDSC - San Diego Supercomputer Center - sdsc.edu

SDSC provides data-intensive computing and cyberinfrastructure (an accessible, integrated network of computer-based resources and expertise, focused on accelerating scientific inquiry and discovery), providing resources, services, and expertise to the national research community including industry and academia. SDSC is a partner in XSEDE (Extreme Science and Engineering Discovery Environment), the most advanced collection of integrated digital resources and services in the world. SDSC was established as one of the nation's first supercomputer centers.

SGIP - Smart Grid Interoperability Panel - sgip.org

SGIP is a consortium established to support the U.S. National Institute of Standards and Technology (NIST) in its fulfilment of its responsibilities pursuant to the Energy Independence and Security Act of 2007. It is a non-regulatory agency dedicated to promoting US innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve quality of life.

RELU - Rural Economy and Land Use program - relu.ac.uk

RELU was a NERC program aimed to advance understanding of the challenges caused by environmental change. Interdisciplinary research was funded between 2004 and 2013 in order to inform policy and practice with choices on how to manage the countryside and rural economies. RELU enabled researchers to work together to investigate the social, economic, environmental and technological challenges faced by rural areas.

UN - United Nations - un.org/en

The UN is an international organization founded in 1945. It is currently made up of 193 Member States. The mission and work of the UN are guided by the purposes and principles contained in its founding Charter.

UNITAR - United Nations Institute for Training and Research - unitar.org

The UNITAR is a principal training arm of the UN, working in every region of the world. They empower individuals, governments and organizations through knowledge and learning to effectively overcome contemporary global challenges. Their training targets two key groups of beneficiaries: the delegates to the UN and others who develop intergovernmental agreements establishing global norms, policies and programs, and the key national change agents who turn the global agreements into action at the national level.

UNOSAT - UNITAR's Operational Satellite Applications Program - unitar.org/unosat/

UNOSAT is a technology-intensive program delivering imagery analysis and satellite solutions to relief and development organizations within and outside the UN system to help make a difference in critical areas such as humanitarian relief, human security, strategic territorial and development planning. UNOSAT develops applied research solutions keeping in sight the needs of the beneficiaries at the end of the process.

VERCE - Virtual Earthquake and Seismology Research Community in Europe - verce.eu

VERCE is developing a data-intensive e-science environment to enable innovative data analysis and data modelling methods that fully exploit the increasing wealth of open data generated by the observational and monitoring systems of the global seismology community.

WDS - World Data System - see ICSU-WDS

WMO - World Meteorological Organization - wmo.int

The World Meteorological Organization is a specialized agency of the United Nations. It is the UN system's authoritative voice on the state and behavior of the Earth's atmosphere, its interaction with the oceans, the climate it produces and the resulting distribution of water resources. It has 191 member states and territories.

APPENDIX 5: ACRONYMS AND GLOSSARY

Acronyms and Abbreviations

CCCO - Communication, Collaboration and Coordination Officer

CRA - Collaborative Research Action

CSIP - Community Strategy and Implementation Plan

DIO - Data e-Infrastructure Officer

DMP - Data Management Plan

DP - Data Plan

DPAB - Data Policy Advisory Board

EDP - Enhanced Data Plan

GPC - Group of Program Coordinators

SAB - Security Advisory Board

SC - Steering Committee

SCN - Strategic Coordination Network

TPO - Theme Program Office

WP - Work Package

Glossary

Action Theme

Each of the six Work Packages (working groups) within this CRA recommended a number of actions to make progress towards the Vision. These actions were collated into four Action Themes to aid readability.

Big Data

Big Data can be characterized by 3Vs: Volume, Variety and Velocity. Although Big Data does not refer to any specific quantity, the term is often used when speaking about petabytes and exabytes of data, much of which cannot be integrated easily. Challenges with Big Data include analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy.

Certified Data

Certified Data has been subjected to a structured quality process to ensure that it meets or exceeds the standards established by its intended consumers. Certified Data is an extension of data quality. Since the quality of data can be measured, standards can be set in a way that data can be certified based on its intended use.

Commodity computing

Commodity computing refers to the use of architectures composed of distributed and heterogeneous computing and data resources providing capacity computing (throughput computing), performance and best serving a large number of applications characterized by a loosely coupled execution mode (often referred as perfectly parallel or embarrassingly parallel applications). The concept of commodity computing services encompass solutions offered by grid and data grid computing, cloud computing. Resilience and fault tolerance are important challenges. Users value commodity-computing services based on their flexibility and ability to offer rapid access to data and computing resources. Architectures and technologies of commodity computing are rapidly changing providing data and computing infrastructures as a service through diverse virtualization mechanisms.

Cloud computing

Cloud computing is an on-demand service offering a large pool of usable and accessible virtualized resources, such as hardware (storage, computing, networking), platforms (development and production) and system of software stack (data management, submission, etc.) in infrastructures that provide those services. The hardware and software is what is called Cloud. Public clouds provide those services to external users, usually in a pay-as-you-go manner. Public research clouds are emerging in the landscape today as an alternative to commercial cloud providers.

Data lifecycle

Data often have a longer lifespan than the research project that creates them. Researchers may continue to work on data after funding has ceased, follow-up projects may analyze or add to the data, and data may be reused by other researchers. Well organized, well documented, preserved and shared data are invaluable to advance scientific inquiry and to increase opportunities for learning and innovation. The data lifecycle includes creating data, processing them, analyzing them, preserving them, giving access to them, and then reusing them.

Data literacy

This is the ability to read, create and communicate data as information.

Data-intensive science

Data-intensive science uses advanced mathematical and statistical methods, computing, and information and communication capabilities to help researchers explore and manipulate massive data sets and enable new insights, not previously possible. In consequence, the speed at which a scientific discipline advances will depend on how well its researchers collaborate with one another and with technologists in areas such as databases, workflow management, visualization, and hybrid and cloud computing technologies.

e-Infrastructure

e-Infrastructure is a combination and interworking of distributed hardware resources (storage systems, computing, networking), Information and Communication Technology (ICT) architectures (e.g. Cloud and virtualization, service-oriented architectures are among the most promising architectures today) and digitally-based technology stack (software, middleware), services and tools (data management and access services, High Performance and commodity computing services, virtualization services, networking services, security services), together with the people and organizational structures and policies (access, authorization) needed to support it all.

e-infrastructures enables public and private resource infrastructures with ICT-enabled methods to achieve new, better, faster, and more efficient research, innovation, decision making. There is a growing trend away from delivery of e-Infrastructure as a technology or product in favor of delivery as a service.

Data-intensive e-infrastructure describes a coherent model for e-infrastructure as a service driven by new data-intensive science that makes data an active component of the e-infrastructure, and requires close coordination and interoperability between key services that were historically divided into components: data management, HPC and commodity computing, networking. Data-intensive e-infrastructure is based on access to and analysis of large amounts and of large diversity of new and existing multidisciplinary data in innovative combinations. Long-data storage, curation, interoperability and certification are just the tip of the iceberg.

Data-intensive e-infrastructures can be described as a new scientific instrument, sometimes referred as a Data-Scope. They are specially designed to provide a seamlessly interacting service allowing users to fluently “observe”, analyze and exploit large amounts of data (created not only by scientific instruments and computers but also by processing and collating archived data) from various disciplines such as environment, social, health sciences for fundamental research and society’s urgent research applications.

e-Infrastructure data layer

The data layer is specifically the data element of any e-infrastructure, such as the functions to prepare and preserve data.

High Performance Computing (HPC)

High Performance Computing refers to the use of high-end parallel computing architectures combining capability computing performance (serving a coarse number of specialized computing applications requiring extremely powerful parallel execution, during which computing, communication, and data throughput tasks are tightly coupled) and capacity computing performance (serving an extremely large number of concurrent parallel tasks on a large-scale computing architecture). HPC technology is rapidly changing toward massively parallel multicore and hybrid architectures to deliver extreme scale performance (in computing, data storage and management systems) which requires a change of paradigm in terms of methods and algorithms.

Today, HPC is not only measured by the number of floating-point operations per seconds (flops) at the scale of the petaflop (10^{15} flops) and soon the exaflop (10^{18} flops) but also in terms of data-handling capacity (at the petabyte scale and soon the zetabyte scale) and capability (high performance parallel data management systems). HPC infrastructures and services are provided in a relatively centralized manner from a limited number of large-scale installations (HPC centers). Those infrastructures have diverse geographical constraints, networking environments, organizational models and authorization policies which vary between countries.

Interoperability

Interoperability is the ability of a computer system or software to work with other systems or products without special effort on the part of a user.

Knowledge Hub

The Knowledge Hub (bfe-inf.org) is the legacy repository for documents produced during the term of this CRA.

Metadata

Metadata describes other data. It provides information about a certain item's content. For example, a text document's metadata may contain information about how long the document is, who the author is, when the document was written and a short summary of the document. A metadata on a data set will include information about the source of the data, and when and where it was captured or generated, while the data's metadata will describe the content of the data fields.

Ontology

In information technology, an ontology is a set of concepts - such as things, events, and relations - that are specified in some way (such as specific natural language) in order to create an agreed-upon vocabulary for exchanging information.

Open Data

According to the Open Data Handbook (opendatahandbook.org/guide/en/what-is-open-data/), "Open Data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike". Open Data must be available as a whole, and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form. Open Data must be provided under terms that permit reuse and redistribution including the intermixing with other data sets. Everyone must be able to use, reuse and redistribute (there should be no discrimination against fields of endeavor or against persons or groups). Open Data enables interoperability: the ability of diverse systems (and organizations) to work together.

Open Science

Open Science is the movement to make scientific research, data and dissemination accessible to all levels of society. Much of the work of science depends on having appropriate tools available to analyze experimental data and to interact with models. Powerful computers are now cheap enough such that significant processing power is within reach of many people. What is often missing is software that lets a researcher choose between models and make sense of the observations, and to test reproducibly the observations and models of other researchers.

Provenance

Provenance of data refers to the process of tracing and recording the origins of data and their movement.

Trusted Data

Trusted Data are drawn from carefully selected sources, transformed in accordance with the data's intended use, and delivered in formats and time frames that are appropriate to specific consumers of reports and other manifestations of that data. Trusted Data should have the following six properties: be complete, be current, be consistent, be clean, be compliant and be collaborative. There must be a complete description of the data available and their provenance.

Trusted Repositories

Trusted Repositories have a mission to provide reliable, long-term access to managed digital resources both now and in the future. Trust is important in two important areas:

- Access to data which may be sensitive in some way or another*
- Preservation services which allow the long-term curation and reuse of any data, despite their potentially disclosive nature*

In the first case, data owners have to trust a repository not only to provide access solely to authorized users, but also to carry out services such as ingest processing (which includes ensuring the data are appropriately anonymized and internally consistent) and data archiving (managing the data within a secure environment) without disclosing any sensitive information. Trust ought to be transitive: data subjects who trust data owners to look after information about them appropriately should implicitly trust the data archives and repositories who become custodians of these data.

In the second case, data users have to trust the data held in the archive are the same data that have been deposited by the data owners. The data should remain so as repositories migrate to new standards and formats to support long-term preservation. Data provided should be not only usable but also authentic and reliable versions of the data. Data users also have the right to know whether the reproducibility of results will be affected by changes to the data. Increasingly, and especially with international access to data, repositories which have been licensed to provide access to data and are able to assign access to these data to other archives have to trust each other. Data creators, repositories and users also increasingly rely upon each other for services as well as data.

Data has to be stored and distributed together with meta-information (including provenance) for long periods, and data integrity must be proven.

Vocabulary

See Ontology.