

Belmont Forum and the e-Infrastructures & Data Management Project

SUGI Nexus Data Management Requirements

London, 12 June 2018

Clement Ng (belmont@jst.go.jp)
Japan Science and Technology Agency



SUGI Nexus

Data Management Requirements

Session Outline

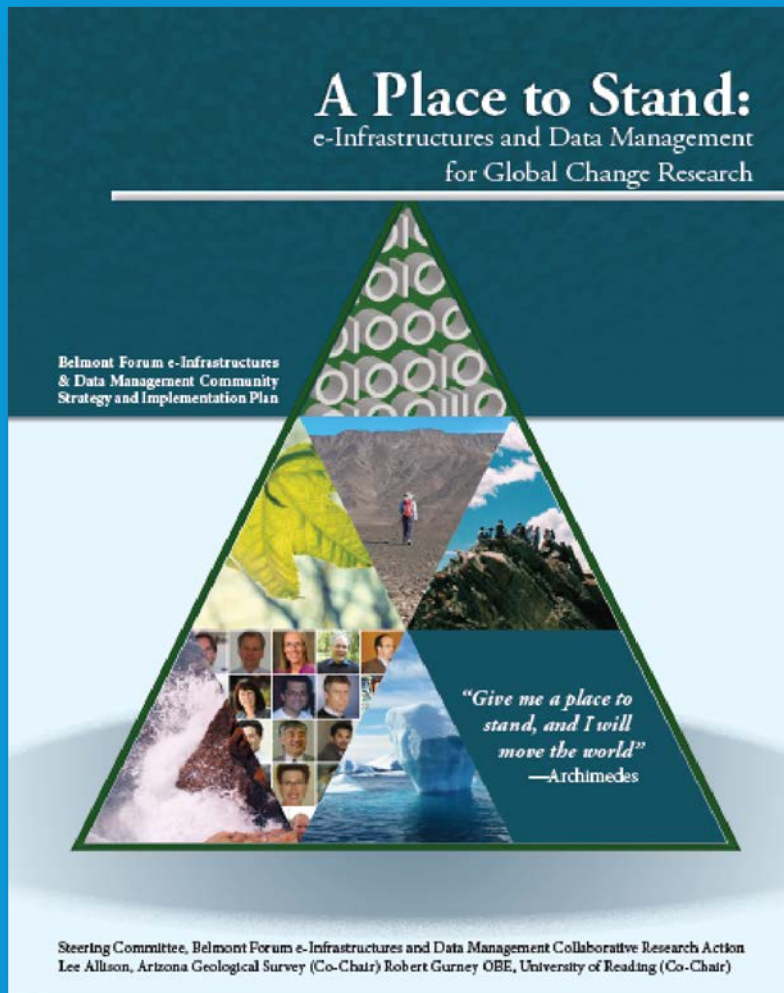
1. Introduction of the Belmont Forum e-Infrastructures and Data Management Project
 - Data Management
 - Training and Capacity Building
2. Workshop Discussion
 - Feedback on Data Management Plan annex
 - Review of feedback received in workshop survey

International Partnership of Funding Agencies and Science Councils



- *More than 25 national science-funding agencies and science councils + some NGOs*
- *Increase impact of environmental research and data : transnational, transdisciplinary CRAs*
- *Promote and leverage policies, methods and systems for: transnational data FAIRness, sharing research products and data stewardship*

e-Infrastructures & Data Management Project

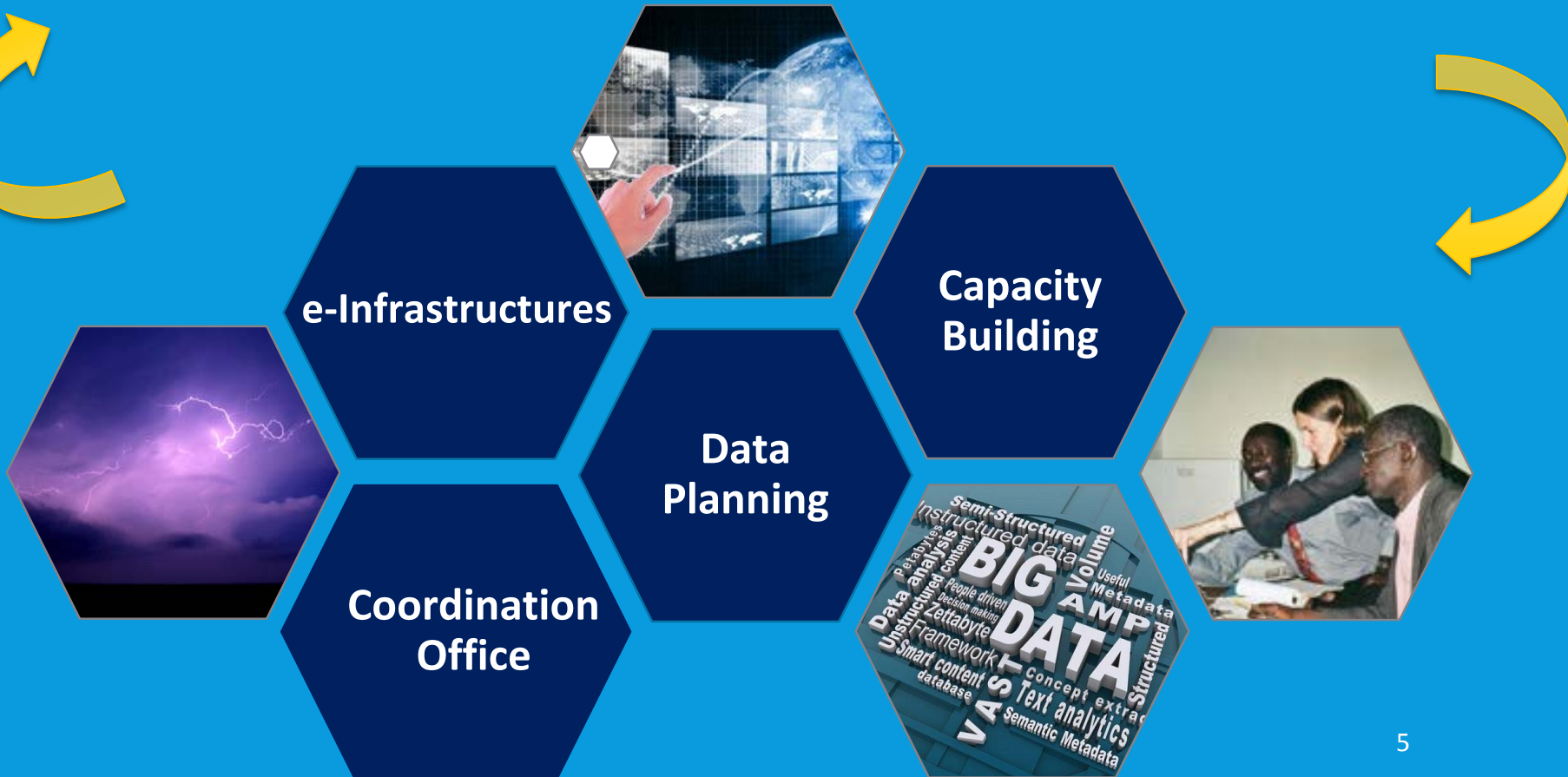


Belmont Forum approved e-
I&DM Project October 2015

<http://www.bfe-inf.org>

e-I&DM Implementation

Belmont Forum Collaborative Research Actions (CRAs – Calls for Proposals)



Data Management Plan: SUGI Nexus Application Annex

The template questions:

- A) *What data sets of **long-term value** do you expect that the project will produce?*
- B) *How do you intend to **manage these data** during the life of the project?*
- C) *How will the data be managed **after the end of the project**?*
- D) *What **supporting documentation and other information** do you plan to make available?*
- E) *Do you envisage there being any **restrictions** on how the data can be accessed or reused?*
- F) *Will there be **other types of material** of long-term value produced?*

Changing Culture and Behaviours

Domain Scientists (CRAs)

Curricula Framework

- Making non-data scientists data savvier

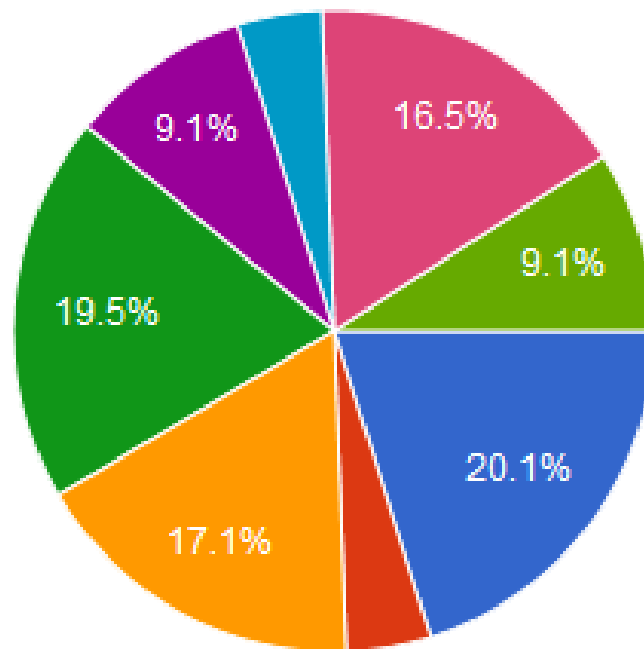
FAIR data principles, services and tools

Training on

- transdisciplinary science, science communication, full life cycle data management and stewardship

Skills gap survey

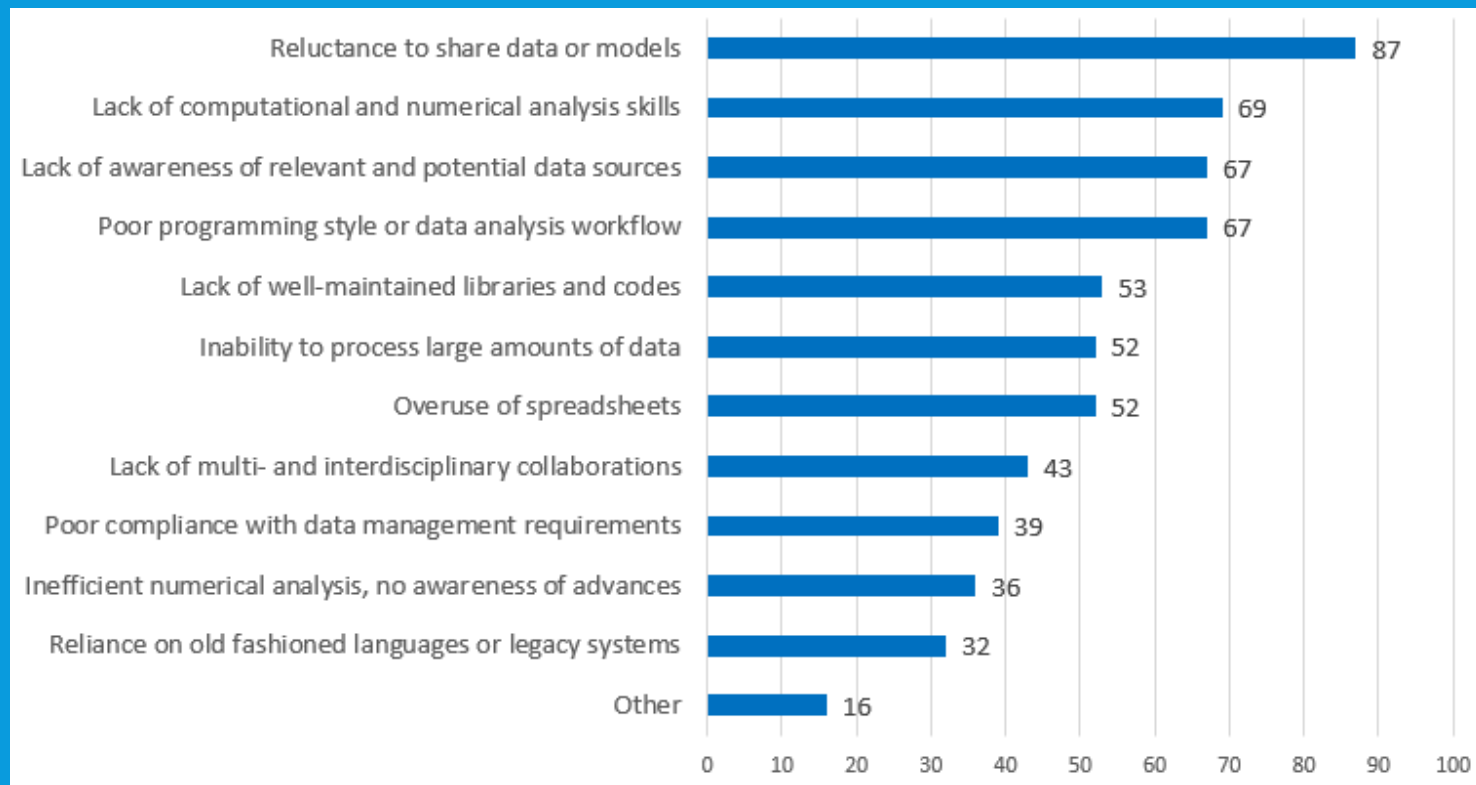
Largest data use challenge



- Data complexity
- Data volumes
- Finding relevant existing data - knowing what's out there
- Lack of data standards and exchange standards
- Dealing with multiple data types
- Data access and file transfer
- Data management and storage
- Other

<http://bfe-inf.org/document/skills-gap-analysis>

Needing most improvement



We now welcome your frank discussion!

Workshop survey questions:

1. How clear and relevant were the questions stated in the proposal application's Data Management Plan?
2. Could you suggest any changes/improvements to the DMP template?
3. In regards to the outcomes included in your DMP, what challenges do you expect to achieve them?
4. What kind of support from the Belmont Forum/funding agencies would assist you with your DMP?
5. Would your project's team benefit from specific member(s) trained to solve data problems?

1. How clear and relevant were the questions stated in the proposal application's Data Management Plan?

- The question were clear and also helped to structure and form the DMP.
- Questions were quite clear and consistent with those available from other guides or supporting tools to DMP

2. Could you suggest any changes/improvements to the DMP template?

- More questions on the data protection and the treatment of trans-national data set would be good to include. E.g. highlight of different data protection policies form different countries (e.g. US vs. EU)
- Open format with own categories, interactive format
- No, not in general. We found to get helpful support by the online tool DMPonline (<https://dmponline.dcc.ac.uk/>) while elaborating the first stage DMP

3. In regards to the outcomes included in your DMP, what challenges do you expect to achieve them?

- Varying international practice standards
 - International data standard practices vary greatly. I found myself explaining the US perspective several times
 - The project intends to experiment with different formats that make the data (in particular the qualitative data) speak to different communities.
 - The partners have different approaches/experiences when it comes to storage, anonymity and the long term reusability or sharing of data. It will require considerable energy to find a common ground.

3. In regards to the outcomes included in your DMP, what challenges do you expect to achieve them?

- Other data standards
 - There is still a big diversity in data storage repositories available, they are neither uniformly structured nor using and supporting equal metadata or data standards, some guidance to use a common platform in this respect would be helpful
- Long-term data
 - Long-term data conservation and curation will be challenging with the multinational partners working together in the projects, we hope to solve this by establishing an own long-term existing Knowledge-base on our projects main topics.

4. What kind of support from the Belmont Forum/funding agencies would assist you with your DMP?

- Data protection
 - Templates for data protection in line with the new EU directive - Guidelines for IT departments of coordinating institutes on these issues
- Best practice
 - Supply examples of good practices. In particular with respect with working across the divide between academic data management and working with various communities of practice.
- Common platforms
 - 1) communication of experiences from other projects 2) recommendations for data repositories 3) recommendation of DMP tools (online or offline)

5. Would your project's team benefit from specific member(s) trained to solve data problems?

- In the scope of EU new regulations on data management, it would be a good idea that project participants could count on help and support of someone trained
- We have 1) personal integrated with DB and GIS experiences, 2) an IT team at hand with one partner 3) one team partner responsible for dissemination WP Some external training or seminar offered by the funding agencies could be helpful

Backup

3. In regards to the outcomes included in your DMP, what challenges do you expect to achieve them?

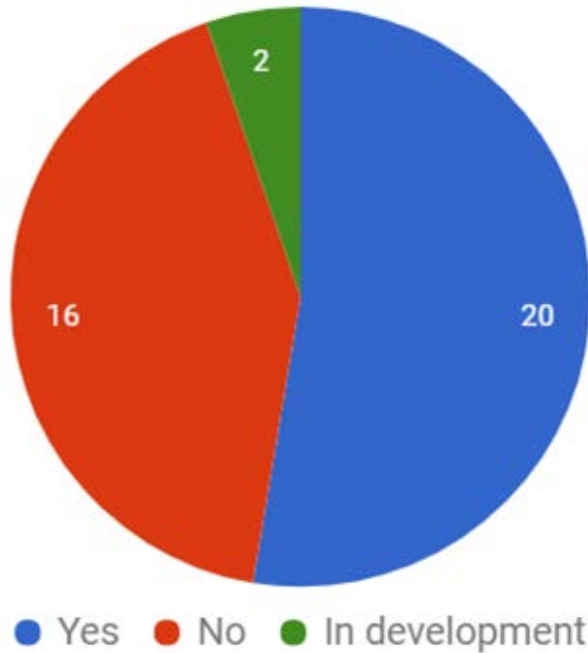
- How do ask shareholders
- Tool box for data collection
- Help with dissemination strategy

More questions for feedback

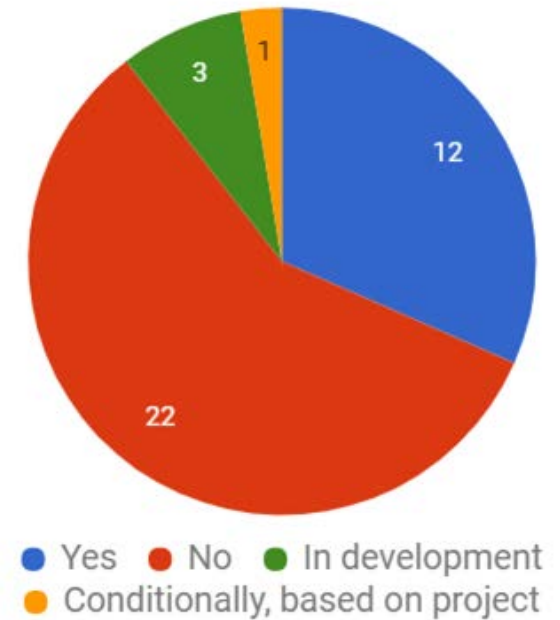
- What interest do they have to actively shape capacity building/policy in the climate change research community?
- Are they interested in writing a post-review potentially exemplifying "best practice" in tackling research problems?
- What interest is there to work on cross-project platforms/standards for common data dissemination?
 - Call text: "These platforms should also serve as a hub for new research results emerging from projects addressing the themes 2 and 3."

Belmont Forum Member Data Policies

Funders with written policies for open data



Funders requiring a Data Management Plan (at any stage)



Data Management Plan Requirements

SUGI

A Data and Digital Outputs Management Plan (DDOMP) describes the data management life cycle for the data to be collected, processed and/or generated by a research project funded within this call.

As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DDOMP should include information on:

Data Management Plan Requirements

SUGI

- the handling of research data during and after the end of the project;
- the types of data, samples, physical collections, software, curriculum materials, and other materials to be collected, processed and/or generated in the course of the project;
- the standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);

Data Management Plan Requirements

SUGI

- policies for broad access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- policies and provisions for re-use, re-distribution, and the production of derivative datasets; and
- plans for archiving data, samples, and other research products, and for preservation of access to them via an institutionally-supported repository.

Curricula overview

Core

- programming, environmental data, visualisation, management, interdisciplinary data

Optional

- software development, object orientated, data science, data organisation

Principal Investigators

- DMPs and repositories, team skills and development

Core Modules



- i. Programming for data intensive research
- ii. Environmental data: expectations and limitations
- iii. Visualising environmental data
- iv. Data management
- v. Interdisciplinary data exchange



Jon Blower – CTO & data visualisation expert at the IEA

Optional Modules



- Software development
- Object-orientated programming
- Data science topics
 - Databases
 - Machine Learning
- Data organisation
 - Workflow
 - Code sharing facilities

CODATA-RDA Research Data Science Summer School 2017
Course co-director Hugh Shanahan & delegate Shaily Gandhi

Principal Investigators



DMPs and
data
repositories

Skills for data
intensive
research

- Prioritising data management
 - Funder requirements
 - DMPs as living documents
 - Resourcing
- Team roles
 - Software engineering is a specialism
 - How to recognise contributions of 'non-publishers'

Example course

- Open online course, launch June 2018
- 20 hours of content
- Science PhD students and early career researchers
- Face-to-face promotional workshops June/July
- NERC funded



Data Management

1. CONTEXT
2. PRACTICALITIES
3. NERC SPECIFICS

Data Application

4. VARIABILITY
5. VISUALISATION

End-Users

6. POLICY
7. BUSINESS
8. MEDIA & PUBLIC

 datatree.org.uk  [@_datatree](https://twitter.com/_datatree)