

**Belmont Forum e-Infrastructures and Data Management
Open Data Publishing Policy Outputs and Recommendations
Report**



Belmont Forum e-Infrastructures and Data Management	1
Open Data Publishing Policy Outputs and Recommendations Report	1
Executive Summary	3
Objectives	3
Background and Method	4
Outputs and Recommendations	5
Open Data Policy Development	5
Community Engagement Issues	6
Cross-references with Other e-I&DM Action Themes	6
Transformation and Delivery	7
Action Plan and Timeline	7
Appendix A: Contributors List	9

Executive Summary

Using evidence derived from a workshop convened in June 2017, this report provides the [Belmont Forum](#) Principals a set of recommendations and accompanying actions for:

- Delivering the Belmont Forum Open Data Policy
- Supporting the sharing and re-use of research data generated by Belmont Forum Collaborative Research Actions (CRAs) and ultimately
- Taking the first steps towards a stronger, more functional global environmental change research e- infrastructure through the development of a deeper mutual understanding among funders, publishers and other stakeholders in this space.

As well as the recommendations themselves, this report takes the first steps to outline an initial five-point action plan and time-frame for achieving the objectives. The actions include:

- The formation of a Publishing Liaison Board¹ to advise on the development of the data policy/ies.
- Investigation into the development of a publishing-oriented research programme within Belmont Forum.
- Raising the status of data-related research activities within the research communities through Belmont Forum staff outreach, participation in relevant scientific meetings and interaction with relevant scholarly societies.
- Establishment of a plan to harness the power of persistent identifiers across the research cycle.

Finally, the inclusion of the list of contributors provides an initial community with whom to engage and transform these initial findings into comprehensive strategies, tools and services. However, engagement efforts should also extend to other existing groups and initiatives, such as FORCE11, Research Data Alliance, DataOne, etc.

Objectives

The objectives for the workshop were developed through consultation among Belmont Forum personnel, publishing professionals, and other key stakeholders over the months leading up to the event. They included carving out opportunities to:

- Test the [Belmont Forum Open Data Principles](#) with publishers. Publishers are key stakeholders who would potentially deliver and support compliance.
- Establish potential lines of future communication to ensure that journal data policy/ies are compatible with funder policy/ies in intention, language and underlying rationale.
- Enable journals to develop and implement clear, rigorous data policies.
- Explore other potential opportunities to work together to strengthen the scholarly communication ecosystem more widely.
- Establish communication channels and explore potential collaboration opportunities with publishers.

¹ Need to check that the PLB remit is appropriate with respect to decision makers such as scientists and their institutions.

- Provide publishers with a clear indication regarding ‘what funders really want’ in order to encourage commitment to and investment in appropriate services.
- Accelerate the practice and status of sharing and re-using research data in the course of regular scholarship.

Background and Method

Following a series of discussions with key stakeholders², Robert Gurney, Co-lead of the Belmont Forum e-Infrastructures & Data Management Project (e-I&DM), commissioned Fiona Murphy to convene a Data Publishing Policy Workshop to facilitate information exchange and debate between members of the scholarly publishing community and members of the Belmont Forum e-I&DM Team.

The Institute of Physics (UK) kindly agreed to sponsor the workshop by providing the venue, materials, catering, paying for Fiona Murphy’s time and providing in-house organisational and logistical support. The materials and outputs for the workshop are to be posted on the Belmont Forum and e-I&DM websites, with the IoP participating in the communications activities during Q3 of 2017.

As well as contributing via discussions and lightning talks on the day itself, participants were also invited to contribute to the first draft of the Outputs and Recommendations Report. In addition, there were a number of domain experts who were unable to attend the actual event but who agreed to review and comment upon the Report. These individuals, together with their affiliations, are listed in Appendix A.

The workshop took place on Friday 23 June 2017. Thirty delegates attended, representing major scholarly publishers³, as well as Crossref, OASPA, Digital Science, CODATA, Digital Science, Wellcome Trust, World Data System, University of Cambridge and CERN and the Belmont Forum e-I&DM Project⁴.

The proceedings included several presentations, initially from the Belmont Forum and the Institute of Physics. There were eight lightning talks by delegates on relevant topics, the slides for which have been posted on the Belmont Forum site here (<http://www.bfe-inf.org/info/open-data-publishing-policy-workshop>). There were two breakout “World Café” sessions that focused on 1) the constituent requirements for a functional data policy; and 2) how to operationalize a data policy. There was a final, whole group, session that pulled together the day’s work with a view to responding to and building upon the workshop’s objectives, and to begin defining recommendations and next steps.

² These included the ALPSP/DPC Seminar: ‘Standing on the Digits of Giants’ (March 2016), various conversations at International Data Week (September 2016), and the STM Association Seminar: ‘Publishing in a larger world, or a smaller one?’ (December 2016)

³ Publishers included: IoP Publishing, The Royal Society, Wiley, Elsevier, Springer Nature, Copernicus, elife, PLOS, Hindawi, F1000

⁴ The Belmont Forum e-I&DM Project was represented by Maria Uhle, Robert Gurney, Barron Orr, Mustapha Mokrane, Mark Thorley, Carrie Seltzer, Bob Samors and Katie Kinsley.

Outputs and Recommendations

Open Data Policy Development

- The [Belmont Forum Open Data Policy and Principles](#), presented in the main session and used as the foundation for the World Café breakout sessions, were agreed to be a sound basis upon which to build more actionable policies.
- Simplicity and clarity are crucial, so ideally a small number of template policies could be agreed upon by the publishers and funders [this could be conducted in conjunction with the **Research Data Alliance Data Policy Interest Group**⁵]
- Definitions or standard terms to express the required levels of openness, curation, persistence and other variables such as how to handle software⁶, how to decide which data should be included, etc., are necessary.
- Data Accessibility Statement templates, instructions and standards should be included as part of the policy development.
- Include Repositories in the discussions – both as a subset of publishers (many datasets are worth ‘publishing’ but are not directly associated with a journal article) and in the terms of reference, e.g., what is a trusted repository?
- Use the outputs from the World Café sessions as well as other projects and policy work, such as the Jisc Research Data Policy, [Force11 Data Citation Principles](#) and [Data Citations implementation pilot subgroups](#), and the Springer Nature data policies⁷, to develop the initial template policy/ies.
- Use the UK Open Data Concordat⁸ and Open Access policy experiences as learning opportunities.
- It was agreed that working together through the Belmont Forum would allow coordination between publishers as well as with funders, and this is a great benefit, and a much greater benefit than working with funders individually.
- It was agreed that in parallel with the policy development efforts, interested parties should work towards developing a pilot project proposal focused on addressing some of the most important technological/process/policy barriers, including but not limited to the need for a persistent

⁵ <https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation>

⁶ Software has subtly different requirements around it to data, specifically around the nature of the licences, and the software dependency graph. Journals such as the Journal of Open Research Software, have been developing great practice around how to review software.

⁷ Link to a preprint (peer reviewed and accepted in International Journal of Digital Curation) that explains the project <https://doi.org/10.1101/122929>

⁸ <http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf/>. On the approach to how the concordat was developed, it defined a set of guiding principles rather than specific policy requirements, for the most part. This is a different approach to e.g. the Springer Nature policies and the RDA group in that implementation as well as policy features and requirements was focused on. The "guiding principles" approach is however useful where policy is in an early stage of development/evolution and different stakeholders need to implement policies in different ways e.g. for certain communities or stakeholders (institutions, publishers, societies etc). A lot was also learned in implementing the RCUK OA policy - in terms of the need to bring the community along with the policy change, rather than meeting resistance as perceived to be implementing a policy from 'on high'.

“grant identifier” that would contribute to linking published manuscripts to associated data to the funding that led to the underlying research.

Community Engagement Issues

- The policy needs to be made as simple and easy as possible for researchers and institutions to comply – consider including reference manager software providers (e.g. EndNote or Zotero) in the discussions, funding the development of tools, and providing templates and other support.
- There is currently a lower-than-ideal level of leadership in the research community with respect to prioritising data sharing and re-use practices – so there is a consequent need to develop domain champions⁹ and foster a clear understanding of what funders require regarding data management/sharing/publication/re-use among the research communities that depend upon them and the benefits of this.¹⁰
- Interact at research conferences, e.g., joint sessions with publishers and funders. Possible meetings include: the American Association for the Advancement of Science (AAAS), American Geophysical Union (AGU) and Society for Neuroscience (SfN). As well as communicating the Belmont Forum’s ideals to researchers, the Forum should also listen to the needs and opinions of the researchers so they feel like their concerns are taken into account.
- Work with domain expert associations such as Institute of Physics, American Institute of Physics, Institute of Electrical and Electronics Engineers, British Ecological Society and American Geophysical Union.
- More cross-functional communication and collaboration among funders, publishers and other stakeholders such as Crossref, Open Access Scholarly Publishers Association, FORCE11, Research Data Alliance, DataCite and ORCID is desirable.

Cross-references with Other e-I&DM Action Themes

- Monitor one or more specific CRA funded projects as controls (AT1: coordination).
- Include Data Management (DM) and Data Publishing (DP) modules in early career researcher and PI training (AT4: capacity and skills building).
- Extend DM and DP considerations to the full research cycle so that researchers are engaging with this issue as early as possible. Actions could include encouraging deposit of data in repositories as part of Data Management Plans (AT2: data management plans).
- Link the concepts and requirements of publishing and data management plans as far as possible [engage with **Research Data Alliance Active Data Management Plans Interest Group** and **Exposing Data Management Plans Working Group**] (AT2: data management plans).
- The ideal of a pilot project taking on the most critical barriers to the harmonization of open data approaches among publishers and funders (e.g., developing a “grant identifier” fits the intention of an anticipated Belmont Forum e-I&DM call (AT3: exemplars).¹¹

⁹ Activities could include direct funding support or providing a platform for these researchers to promote their work and careers.

¹⁰ Including work on reward structures for sharing data.

¹¹ Amongst the various stakeholders, publishers and funders were also identified as the top priority in the first RDA data policy IG meeting. Report from the April meeting in Barcelona:

https://docs.google.com/document/d/1GiJl7kJA3MgDvJyC9zw-zHlbg3n2azhN16W_2Kn1uwM/edit?usp=sharing

- Explore infrastructure funding partnerships with publishers and/or other parties (either AT3: exemplars, or could be defined as spanning the whole e-I&DM scope).¹²
- Draw up a template policy, do a test call (to establish test cases and collect evidence) and run a number of projects in collaboration with publishers (AT1-4).

Transformation and Delivery

- Rethink the funding of data infrastructures – explore options that have long-term sustainable and strategic roadmaps rather than relying on piecemeal, indirect funding flow via Article Processing Charges and other ‘soft’ mechanisms such as short-term grant awards.¹³
- Take advantage of opportunities that result in automation and interoperability (e.g., persistent identifiers such as ORCID, as well as FORCE11 and other initiatives) that will allow end-to-end tracking of research outputs connected to funding.

Action Plan and Timeline

Action Item	Timeframe	Who	Success Criteria
1 Develop initial recommendation plan to present to Belmont Forum Principals	Early September 2017	All Contributors (coordinated by Fiona Murphy, Robert Gurney, Barron Orr)	Plan submitted in for Belmont Forum Plenary
2 Form a Publishing Liaison Board (PLB) to oversee the development of the first template data policy/ies	PLB in place as soon as possible	Subset of the delegates/experts associated with this report, include work with RDA Policy IG	Template policy/ies compiled, scrutinised by RDA Policy IG by mid-2018
3 PLB or other appropriate group to develop a possible research programme within Belmont Forum	Goal to issue a call in 2018, co-ordinated with AT3 Exemplars call	A suitable group to be identified, formed and provided with a framework for devising the programme. This is likely to require further consultation with publishers	A rigorous programme call to be in place by the suggested deadline
4 Re-prioritising the value of research data – its management, sharing and	A programme of actions in place by March 2018. Could include - but not be limited to -	A cross-functional group of funders, publishers, learned	Put feedback loops in place to gauge reach and researcher responses.

¹² Useful resource here: <https://www.rd-alliance.org/final-report-income-streams-data-repositories.html>

¹³ Could possibly use access to the BF funding agencies to get their take on how data (and other) e-infrastructures should be supported. Can build on work done by OECD, CODATA etc on sustainability. In the same way that we have BF data policy principles, is there a need to develop a set of BF e-infrastructure principles, which will articulate what the requirement of research funders is to support?

	re-use – amongst researchers	communications during calls for proposals, sessions at research meetings, stick and carrot approaches (such as withholding or making available certain funding streams) other outreach activities, a specific Belmont Forum call (or subset thereof) focused on re-using existing data rather than collecting new data. Would need to ensure that CRA call peer reviewers are educated in the value of data re-use.	societies, institutions to be consulted. The RDA may be able to help with this.	Ultimately, the success of this action will be measured by the quality and quantity of proposals received by the Belmont Forum that include sharing and re-use of data
5	Harness the power of persistent identifiers¹⁴ across the whole cycle of research. This includes using ORCiDs, potentially developing a Grant ID research project¹⁵, working with Scholix, attaining an in-depth understanding of PIDs’ potential in supporting e-research capacity	Develop action plan by May 2018 ¹⁶	Work with Crossref, ORCID, RDA and other experts. This also feeds into action item 4.	Need to set measurable targets for uptake (ORCID), run and test the Grant ID potential, require metrics of the emerging services to inform future funders’ and institutional decision-making ¹⁷

¹⁴ Starting with the Open Funder Registry- which already exists and is being used as an identifier for Funders (not grants) <https://www.crossref.org/services/funder-registry/>

¹⁵ Possibly joining the grant identifier project being initiated by Crossref and DataCite as an extension of the Open Funder Registry.

¹⁶ Need to clarify Belmont Forum’s best leveraging opportunities. Likely to include evidence gathering around needs and use cases for persistent identifiers.

¹⁷ Also consider data identifiers, which probably default to DOIs at this point, but would be useful to consider hash representations of the data too. This is also about what metrics/reports the funders would be able to rely on to understand data usage, grant efficacy, PI performances, program impact, etc.

Appendix A: Contributors List

Name	Institution	Participated in the Workshop	Presented at the Workshop	Reviewed the Draft Report
Kathryn Sharples	Wiley	X		
Emma Ganley	PLOS	X	X	X
Jamie Hutchins	IoP Publishing	X		
Maria Uhle	National Science Foundation	X	X	
Paul Hardaker	IOP	X	X	
Fiona Murphy	MMC Ltd	X		
Phill Jones	Digital Science	X		
Robert Gurney	Belmont Forum	X	X	
Bob Samors	Belmont Forum	X		
Katie Kinsley	Belmont Forum	X		
Carrie Seltzer	National Science Foundation	X		x
Mustapha Mokrane	WDS	X	X	
Barron Orr	Belmont Forum	X		
Hans Pfeiffenberger	Copernicus	X	X	
Wouter Haak	Elsevier	X	X	X
Suenje Dallmeier-Tiessen	CERN	X	X	X
Artemis Lavasa	CERN	X		X
Geoffrey Bilder	Crossref	X	X	
Melissa Harrison	elife	X	X	X
Holly Murray	F1000	X		X
Cameron Neylon				X
Meredith Morovati	Dryad			X
Thomas Faust	Hindawi	X		
Marc Gillett	IOP Publishing	X		
ian mulvany	SAGE Publications			X
Christopher Wileman	IOP Publishing	X		
Claire Redhead	OASPA	X		X
Iain Hrynaskiewicz	Springer Nature	X	X	X
Lauren Cadwallader	University of Cambridge	X		X
David Carr	Wellcome Trust	X	X	X
Kevin Ashley	Digital Curation Centre			X
Phil Hurst	The Royal Society	x		x
Mark Thorley	NERC	x		x