

## Evaluating the Belmont Forum's Progress Toward Making Data Open/FAIR v.1

### Background and Scope

The Belmont Challenge is “to deliver knowledge needed for action to avoid and adapt to detrimental environmental change including extreme hazardous events.” Toward that end, the Belmont Forum (BF) adopted the Open Data Policy & Principles in 2015 and subsequently funded the e-Infrastructures & Data Management (e-I&DM) Project to provide recommendations and essential resources, such as the Data and Digital Outputs Management Plan (DDOMP), e-infrastructure exemplars, and data management training, to help implement the Open Data Policy & Principles.

e-I&DM's specific charge is “the coordination of four interrelated programs that leverage existing knowledge and resources to illuminate **achievable, reproducible systems for effective, sustainable data management practices**” (e-I&DM Implementation Plan 2017). As part of this focus, the e-I&DM team suggests possible evaluation metrics to assist the Belmont Forum in evaluating and demonstrating progress toward open data in its research program. Metrics can also be useful to identify the need for programmatic ‘course corrections’ and to encourage data reuse and other full path data lifecycle and research practices that BF seeks to acculturate, along with multinationalism and transdisciplinarity, in its research funding programs.

The purpose of this document is to provide a suggested approach to evaluating the Belmont Forum's progress towards full implementation of its Open Data Policy & Principles. The paper proposes basic, easy-to-collect (and/or easy-to-generate) metrics (i.e., absolute numbers and percentages), the stage(s) of the Collaborative Research Action (CRA) process at which the metrics could be collected, and a suggested responsible party to collect/generate each metric. It is important to note that metrics to evaluate “openness” or “FAIR-ness” of BF data or “success” of the program in achieving the Belmont Challenge are beyond the scope of this document (however, suggested reading is this preprint article: “[Evaluating FAIR-Compliance Through an Objective, Automated Community-Governed Framework](#)”<sup>1</sup>).

As we move forward, core concepts such as open access, open science, open data and FAIR data are important to these metrics. Open access is narrowly defined as the practice of providing online cost-free access to the results of scientific work, in the form of publications in peer-reviewed journals, which are freely accessible to the general public. Open science is broader in that it refers to the availability of research data, methods, results and publications, within the limits of research agreements. This does not equate to “free science,” as the intellectual property is protected before making knowledge publicly available so as to help translate research results into innovation.

Furthermore, although the concepts of open data and FAIR Data have similarities, note that open data ≠ FAIR data. The key difference is that for data to be considered “open” it should be available to anyone to access, use and share, without licences, copyright or patents. It is expected that open data at most should be subject to attribution/share-alike licenses (e.g., [CCBY](#)). In comparison, FAIR data are data that are Findable, Accessible, Interoperable and Reusable as defined by [Force11](#), [GO FAIR](#) and these [Guiding Principles](#). The term “Accessible” means accessible by appropriate people, at an appropriate time, in an appropriate way. This means that data can be FAIR even when it is private, so long as it is accessible by a defined group of people, or when it is accessible by everyone (open data).

### Key Questions and Metrics to Answer Them

In considering the value-added benefits of the Belmont Forum's unique funding partnerships, it is important to ask: Is our research funding program effective in promoting access to transdisciplinary research data as intended by the Open Data Policy & Principles? To answer this, it is helpful to parse this question into four key stages of the CRA process: Pre-Proposal, Full Proposal, End-Term Valorization, and Publication/Outcomes. It should be noted that evaluation metrics during the Project Implementation stage were considered, but were deemed uninformative based on observations and feedback from researchers at mid-term valorization workshops<sup>2</sup>. Table 1 below outlines questions and

<sup>1</sup> Wilkinson, Mark D., et al., BioRxiv, doi: <https://doi.org/10.1101/418376>.

proposed metrics to answer them at these key stages of each CRA and a suggested responsible party for metrics collection. These metrics are intended to be collected for individual CRAs but metrics can also be aggregated for all CRAs to date.

<b>Table 1. Metrics to Evaluate Progress Toward Open Data (CRA)</b>		
<b>Questions</b>	<b>Metric(s)</b>	<b>Responsible Party</b>
<b>Pre-Proposal Stage</b>		
How many pre-proposals will produce or plan to reuse datasets or digital outputs of long-term value?	Absolute number of proposals and as percentage of total	pre-proposal evaluation team or as reported via bfgo.org
How many pre-proposals have identified a team member(s) primarily responsible for data management?	Absolute number of proposals and as percentage of total	pre-proposal evaluation team or as reported via bfgo.org
How many pre-proposals have accounted for the anticipated costs of full-path data management?	Absolute number of proposals and as percentage of total	pre-proposal evaluation team or as reported via bfgo.org
<b>Full Proposal Stage</b>		
How many proposals will produce datasets or digital outputs of long-term value?	Absolute number of proposals and as percentage of total	proposal evaluation team or as reported via bfgo.org
How many proposals will reuse datasets or digital outputs of long-term value?	Absolute number of proposals and as percentage of total	proposal evaluation team or as reported via bfgo.org
For proposals that will produce or reuse long-term value datasets/outputs, how many include data management plans that conform to the DDOMP Annex?	Absolute number and percentage of total	proposal evaluation team or as reported via bfgo.org
How many data management plans include, if necessary, appropriate restrictions on access to or reuse of project data and digital outputs?	Absolute number and percentage of total	proposal evaluation team or as reported via bfgo.org
What percentage of the total proposal cost does data management represent (including salaries for members tasked with data management)?	Percentage of total budget	proposal evaluation team or as reported via bfgo.org
<b>Award Stage</b>		
How many awarded projects' DMPs provide responses to the 5 types of additional information requested in the DDOMP? 1) Agreed 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); 2) Policies for broad access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements; 3) Policies and provisions for mining, reuse, re-distribution, and the production of derivatives; 4) Contact information for the person(s) responsible for updating the Data and Digital Outputs Management Plan as needed to comply with these guidelines; and 5) A list of anticipated trustworthy, long-term repositories or data centers that will be used to ensure preservation of access to data and digital outputs following completion of the project.	Absolute number and percentage of total	Secretariat or as reported via bfgo.org

<b>Table 1. Metrics to Evaluate Progress Toward Open Data (cont'd)</b>		
<b>Questions</b>	<b>Metric(s)</b>	<b>Responsible Party</b>
<p>How many projects have fulfilled the following measures at this stage?</p> <ol style="list-style-type: none"> <li>1) Clearly defined policies are in place regarding broad access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.</li> <li>2) Clearly defined policies are in place regarding mining, reuse, re-distribution, licenses and the production of derivatives.</li> <li>3) Data clearly expressed in universal standardized units are in place.</li> <li>4) Contact information is readily available for the person responsible for updating the DMP.</li> </ol>	Absolute number and percentage of total	Secretariat
<b>Outcomes/Publication (approximately 2 years after end of project term)</b>		
How many project teams have published their results with a Data Accessibility Statement (whether or not required by the publisher)?	Absolute number and percentage of total	self-report or Secretariat (via survey to PIs or reported via bfgo.org)
How many projects have deposited their data and digital objects in a trustworthy repository? (requires compliance with repository requirements, e.g., metadata, PID, etc.)	Absolute number and percentage of total	self-report or Secretariat (via survey to PIs or reported via bfgo.org)
How many projects' data and digital outputs have been/are being reused (beyond its original context and original creators)?	Absolute number and percentage of total	self-report or Secretariat (via survey to PIs or reported via bfgo.org)
How many projects have had their data and/or digital objects cited?	Absolute number and percentage of total	self-report or Secretariat (via survey to PIs or reported via bfgo.org)
<b>Optional Metrics</b>		
How many BF member agencies have a formal Data Policy?	Absolute number and percentage of total	Secretariat via online research or survey of member agency representatives
To what degree do proposers agree that the DDOMP clearly defines the expectations and requirements of the Belmont Forum's Open Data Policies and Principles?	Likert scale value 0 - 5	self-reported or as reported via bfgo.org or via Secretariat

Note that two questions in Table 1 are listed as “Optional Metrics”. While responses to these questions would not strictly indicate progress toward open data, they would inform the Secretariat of progress made by member agencies in adopting and implementing open data policies and procedures, as well as provide guidance on the effectiveness of the DDOMP. Furthermore, the number of agencies having formal Data Policies is valuable information that could help in maintaining and updating the Belmont Forum's Policy Comparison Tool.

### **Metrics Interpretation**

More important than an absolute number for any given metric is a trend demonstrating consistent forward progress toward full implementation of the Open Data Policy & Principles in all CRAs and throughout the full path of the data lifecycle for all funded projects. For this, the percentage of projects complying with or fulfilling their DMP should trend ideally towards 100% over time. This trend will be most likely be discernible in the Pre- and Full Proposal stages, with percentages at the End-Term and

Outcomes stages likely to be less straightforward due to the variable timeframes associated with publication and the logistical challenges of tracking, measuring and reporting data reuse.

**Considerations and Questions for Further Discussion**

Considerations

1. The proposed evaluation metrics and any given funded project’s program compliance may be related, but the evaluation metrics listed above are intended only to reflect general programmatic trends and the effectiveness of the DDOMP, in making Belmont Forum-funded research more “open”.
2. If data are not reused within the evaluation stages, that in and of itself does not necessarily imply or signal non-alignment with the Open Data Policy or that data are ‘closed’ or un-FAIR.
3. The Secretariat should collaborate with Belmont Forum trainers to iteratively improve and update training, DDOMP requirements and resources as feedback is received from researchers, trainers, evaluators and member agencies.
4. It may be important to track when and what kinds of training, if any, was made available to researchers, as this may have an effect on metrics (as well as compliance).
5. It may be useful to consider these metrics in relation to proposal evaluation/compliance metrics, e.g., if data reuse is a desired activity that Belmont Forum wants to promote.
6. It may eventually be most effective for the Secretariat to ‘automate’ evaluation metrics collection via self-reporting by researchers/PIs via the bfgo.org site.
7. Data management/Open Data Policy implementation is likely more complicated when CRA funding involves ERA-Net, H2020, or other funding programs and their respective data requirements, and/or when individual BF members’ data management requirements differ from those of the DDOMP Annex.
8. An additional consideration (or perhaps to be included as a metric of its own), is the degree to which BF agencies are providing the financial support necessary for individual project teams to achieve the Open Data Policy and Principles.

Questions for Further Discussion

9. How will/should the metrics be interpreted (i.e., is there a point that represents a ‘good’ vs. ‘unacceptable’ result in terms of achieving the goal[s]). Who decides? Are exceptions allowed and, if so, under what circumstances? What steps could/should be taken for any necessary course correction?
10. How do you ‘test’ or measure progress toward FAIR data for transdisciplinary data? Do the same metrics used by GO FAIR and others apply to non-traditional outputs (this question may be more productive/appropriate for discussion at the Secretariat metrics workshop)?
11. [FAIR Data advocates](#) have argued that 5% of total funding should be allocated for data stewardship. Is this an amount/percentage that all Plenary members are willing and able to support? Should this become a formal (or informal) benchmark or guideline for proposal evaluation purposes?
12. Should cost for data management training be included in the recommended 5% for Data Management portion of the budget?
13. How should the additional [data management questions](#) that arose from the survey of SUGI projects (June 2018) and earlier e-I&DM survey of Belmont Forum member data management policies be addressed?

**Document Revision History:**

Date (yyyymmdd)	Version & Changes	By
20190221	v1. Document submitted to Belmont Forum Secretariat	T. Lee, e-I&DM Proj. Coord.; M. Charles-Ayinde, NSF AAAS Fellow 2019