Research Data Management

Wade Bishop, Associate Professor
@WadeBishopUTK
https://orcid.org/0000-0002-5022-2707
School of Information Sciences
University of Tennessee

Data Deluge

- To manage the scientific “data deluge” more work is needed to ensure that data will be findable, accessible, and interoperable for reuse by both humans and machines (Hey & Trefethen, 2003).
- Data reuse improves and advances science permitting others to verify results; enabling the repetition of experiments; and to conduct new research (Pryor, 2012).
- Big data presents large-scale challenges as researchers try to navigate massive quantities of data, work across disciplinary boundaries, and keep pace with the requirements of DMPs and preservation needs (Jaguszewski & Williams, 2013).
- In response to DMP requirements, academic institutions, libraries, publishers, and scientific and professional associations from all disciplines have made strides to make data more findable, accessible, interoperable, and re-usable.

Hey, A.J.G. & Trefethen, A.E. (2003). The data deluge: An e-Science perspective. In, Berman, F, Fox, G C and Hey, A J G (eds.) Grid Computing - Making the Global Infrastructure a Reality. Wiley and Sons, pp. 809-824.
Jagszewski, J. M., & Williams, K. (2013). New roles for new times: Transforming liaison roles in research libraries, (August), 1–17. Retrieved from http://www.arl.org/storage/documents/publications/nrnt-liaison-roles-revised.pdf
Pryor, G. (2012). Why manage research data? In G. Pryor (Ed.), *Managing research data* (pp. 1-16). London, UK: Facet Publishing.

Importance of Data Management for Science

- Data may only be collected once in real-time
- Data have enduring value (in perpetuity)
- You can verify results! (reproducibility)
- Data may be at a global scale
- New research via re-use

Why Manage Data: Researcher Perspective

- **Keep yourself organized** – be able to find your files (data inputs, analytic scripts, outputs at various stages of the analytic process, etc.)
- **Track your science processes for reproducibility** – be able to match up your outputs with exact inputs and transformations that produced them
- **Better control versions of data** – easily identify versions that can be periodically purged

- **Quality control your data more efficiently**
- **To avoid data loss** (e.g. making backups)
- Format your data for **re-use** (by yourself or others)
- **Be prepared**: Document your data for your own recollection, accountability, and re-use (by yourself or others)
- Gain **credibility and recognition** for your science efforts through data sharing!

https://dataoneorg.github.io/Education/

Data Management Facilitates Sharing and Re-use...

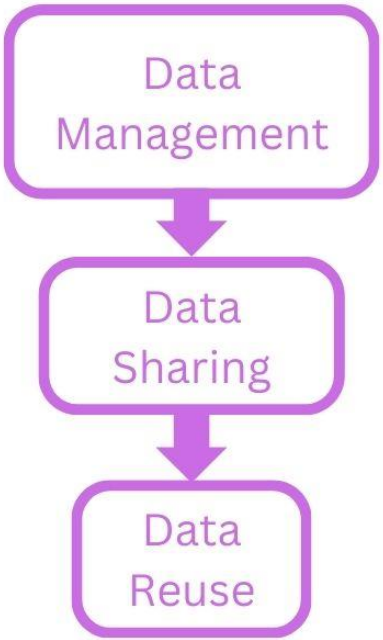


Image by Jaxx Fox

Publicly accessible data is important: why?

Here are a few reasons (from the UK Data Archive):

- Increases the impact and visibility of research
- Promotes innovation and potential new data uses
- Leads to new collaborations between data users and creators
- Maximizes transparency and accountability
- Enables scrutiny of research findings
- Encourages improvement and validation of research methods
- Reduces cost of duplicating data collection
- Provides important resources for education and training

Value of data management

- Natural disaster
- Facilities infrastructure failure
- Storage failure
- Server hardware/software failure
- Application software failure
- External dependencies
- Format obsolescence
- Legal encumbrance
- Human error
- Malicious attack by human or automated agents
- Loss of staffing competencies
- Loss of institutional commitment
- Loss of financial stability
- Changes in user expectations and requirements

Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B., & Stafford, S. G.. (2019). Non-geospatial metadata for the ecological sciences. *Ecological Applications,* 7. 330-342.

The life of data

- All decision-making gains from a better understanding and implementation of the data lifecycle.
- Historically, the research lifecycle concluded with dissemination of findings in a scholarly text but to an “ever-growing extent the published report is accompanied by supplementary data” (Pryor, 2012, p. 6).
- Traditionally, researchers were incentivized in the production of scholarly publications, but the data were a less disseminated byproduct of the research enterprise.
- “Open” movements demand data dissemination to verify results, enable repetition of experiments, and execute new research using the generated data (Higgins, 2012).

Digital Curation Centre's Curation Lifecycle Model

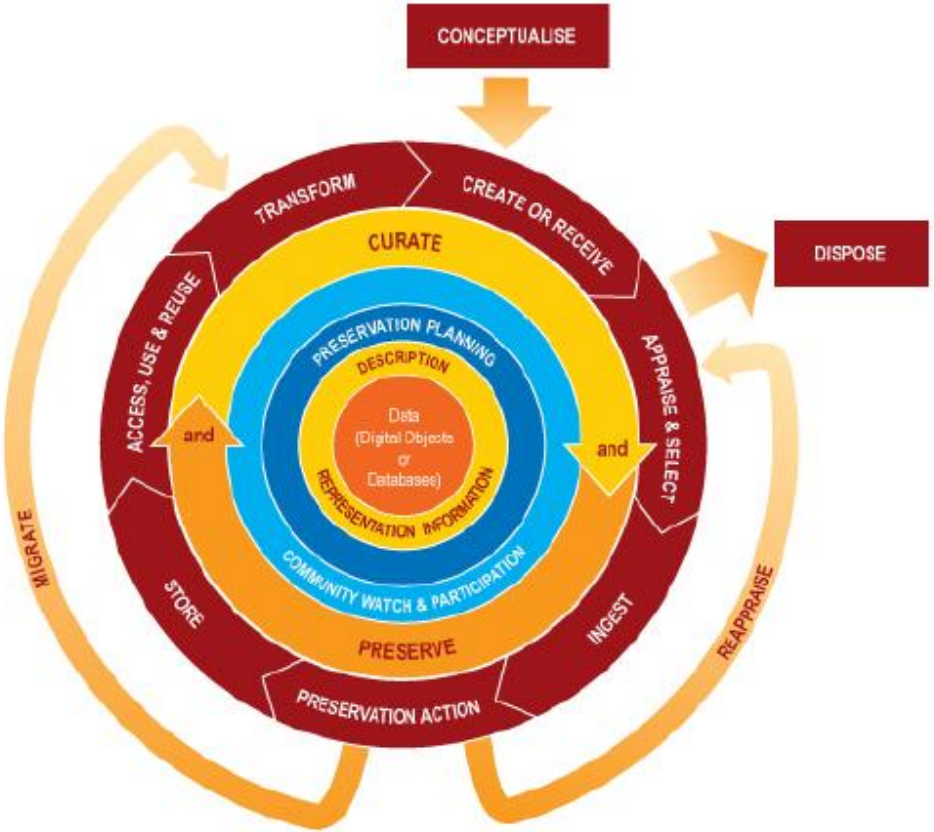


Image by DataONE

Research Data Management (RDM)

- Education
  - Raise awareness of RDM’s importance
  - Encourage RDM skill-building
  - Disclose RDM tools and resources
- Expertise
  - Decision support for RDM problems
  - Customized solutions to RDM problems
- Curation
  - Technical infrastructure to support RDM
  - Related services that support RDM

Not just the library!

- Various aspects of RDM are often distributed across different support services and academic departments.
- Researchers need support in planning, organizing, security, documenting, and sharing datasets for deposit, preserving them on a short and a long-term basis.
- They may seek advice on copyright, licensing and intellectual property issues.
- To address all these issues, libraries must engage in high levels of interaction with researchers, while there is also a need for cooperating with other support service providers.

The shift is library services

- Libraries are shifting their focus to “what users do (research, teaching, and learning) rather than on what librarians do (collections, reference, library instruction)” (Jaguszewski & Williams, 2013, p. 4).
- Wittenberg et al. 2018 designed trainings to address these learning outcomes:
  1. Librarians and library staff will be able to recognize data manage-ment questions and confidently respond.
  2. Librarians and library staff will actively build and participate in a research data management community of practice.
  3. Librarians and library staff will understand the data management needs of researchers in their domain and incorporate that awareness into reference and instruction work.

But the course title is Science Liaison Librarianship

- Academic liaison librarians offer a domain expertise in data identification, selection, organization, preservation, and access which should be leveraged by the RDM consultants in order to provide a comprehensive service
- At UC-Berkeley, when RDM receives a request for a consultation or group training, they alert the library liaison who provides support for the researcher's department.
- Liaisons have the knowledge and ability to identify connections with the research lifecycle and data, which compliments the RDM consultants whose expertise lies in IT.

Data Management Plan (DMP)

- DMP is a structured, formal document describing roles and responsibilities for maintaining and managing data during and after the conclusion of a research project (Bishop & Hank 2020).
  - Formal document
  - Outlines what you will do with your data *during* and *after* you complete your research
  - Ensures your data is safe for the **present** and the **future**

- DMPs are intended to document data lineage, and allow portability, transferability, and future use of data.

- In 2011, National Science Foundation (NSF) and other U.S. research funding agencies began requiring researchers to submit and execute a Data Management Plan (DMP) to fully reap the benefits from research investments.

- With the push for more public-facing scientific research and accountability, many funding agencies (86% of UK Research Councils and 63% of U.S. funding bodies) require DMPs within the initial funding application (Smale et al., 2018).

Why Data Management Plans?

- Presently, the Horizon 2020, EU research and innovation program, requires data from all publicly funded research be accessible to anyone, free of charge, in addition to ensuring Open Access to all peer-reviewed scientific publications relating to its results (Koumoulos, 2019).
- Several academic journals now also require researchers to make public the data and digital outputs associated with a publication (The Royal Society, 2017; PLOS, n.d.).
- DMP use has been imposed onto the research community rather than through grassroot efforts of researchers'' themselves.
- However, research has shown that DMPs are uneven in terms of their quality, with the most missed elements being clear roles and responsibilities for data management, metadata standards for describing research data, and policies for protecting intellectual property rights (Samuel et al., 2015).

Components of a General DMP

1. Information about data & data format
2. Metadata content and format
3. Policies for access, sharing and re-use
4. Long-term storage and data management
5. Roles and responsibilities
6. Budget

Tools for Creating Data Management Plans

Data Stewardship Wizard
https://ds-wizard.org/.
DMP Tool
https://dmptool.org/
DMP Online
https://dmponline.dcc.ac.uk/

Data Flow & Use Considerations

Technical Obsolescence: Occurs when hardware and software are no longer updated, maintained or used, including when newer versions or services are created or released to replace an older or retired version, or when a technology is discarded outright, typically due to a decline in use.

Interoperable: The capacity to effectively exchange complete, consistent and compatible information between computing systems regardless of differences in hardware, software and communication protocols.

Data Lineage: Describes the history of research data, from its origin to how it is used, managed and stored. May also be referred to as data provenance, reflective of the fundamental concept of provenance in archival sciences, which refers to the origin, custody and ownership of archival records over time.

Data Flow & Use Considerations

Existing data: If existing data are used, what are their origins? Will your data be combined with existing data? What is the relationship between your data and existing data?

How data will be managed in short-term: Version control, Backing up, Security & protection, Who will be responsible? What metadata are needed, Any details that make data meaningful?

How metadata will be created and/or captured: Lab notebooks? GPS units? automated metadata?

What format will be used for the metadata: Standards from the community, discipline? Justification for format chosen needed

Storage Considerations

- Are there backups of the backups?
  - Necessary for high-value data
  - Usually different copies of backups are kept in different locations
- How long do you keep your backups?
  - Depends upon specific situation, and should be determined in concert with stakeholders and resource managers
  - Understand relevant guidelines, policies and rules for retention of data
  - What are the long term storage and access solutions that are relevant for the project? What to do when funding ends or key staff depart?
- Who will be around after a project ends?
  - Changes in the status of the project, funding, or key staff are important reasons to have a full understanding of related options and requirements for storage and access

Roles and Responsibilities

- Who will be responsible for implementing the data management plan?
- How will adherence to this data management plan be monitored for compliance?
- What process is in place for transferring responsibility for the data in perpetuity?
- Who will have responsibility over time for decisions about the data once the original personnel are no longer available?

Stakeholder considerations

- Who owns the copyright?
- Institutional policies
- Funding agency policies
- Embargos for political/commercial reasons
- How should data be cited when used?
- Persistent citation?

Choosing an open license

- Why use an open license?
  - Facilitate data sharing and discovery
  - Increase visibility of your data
  - Advance knowledge
- Creative Commons
  - CC0 (not a license, but a waiver)
  - CC BY (Attribution)
  - CC BY-ND (Attribution-NoDerivs)
  - CC BY-NC (Attribution-NonCommercial)
  - CC BY-SA (Attribution-ShareAlike)

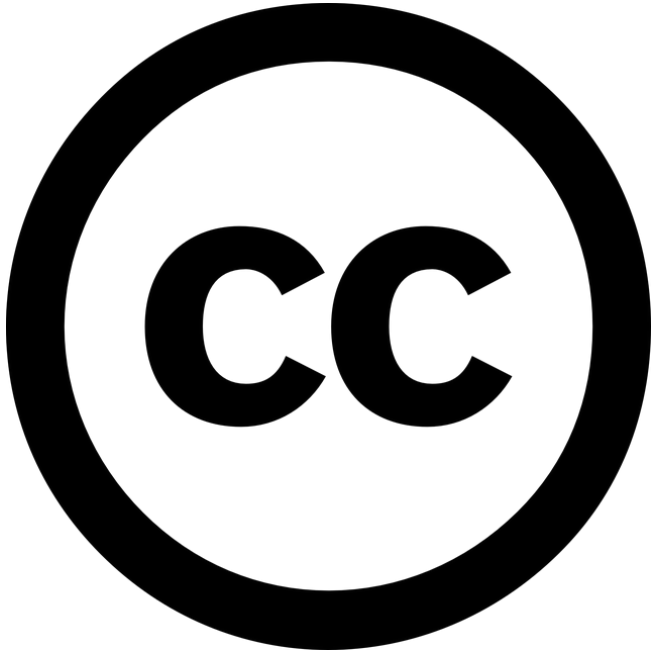


Image by CopyrightFreePictures from Pixabay

Cost considerations

What are the anticipated costs?

- Time for data preparation & documentation
- Hardware/software for data preparation & documentation
- Dedicated Personnel
- Archive/repository costs

How costs will be paid?

Key Concepts and Definitions

- **Digital repositories:** An entity providing long-term data storage, typically affiliated with a research institution or university
- **Persistent Identifier:** A unique reference to a researcher, data set, or other digital object that distinguishes it from all others and typically includes a stable hyperlink to the object may be discovered (e.g., DOI).
- **Data citation:** A process to that enables the discovery of data and digital objects beyond a digital creator through the creation of a persistent identifier and the ability to reference a data set or other digital object in research publications.

Institutional repositories (IRs)

- Institutional repositories are a type of digital repository located and affiliated with one specific institution, typically a university or research institution
- Other digital repositories may be disciplinary in nature, rather than institutional, as they serve many communities
- These data-purposed digital repositories have a focused scope and audience, whereas institutional repositories are typically created with the intent of preserving access to scholarly publications from the institution’s researchers

Figshare (https://figshare.com/)

- Make your data more discoverable and open to all your readers
- Secure hosting and visualization in the browser of all file types
- Authors can easily upload files with no concerns about file size or format
- All data is citable and has a DOI

Dryad Repository (http://datadryad.org)

- Dryad uses DSpace, a free and open source repository software that preserve and access all types of digital content (e.g., text, data, and so forth)
- Dryad allows scientists to submit data, which may be linked to related publications using that data, and assigns Digital Object Identifiers (DOIs) to enable data citation

Persistent ID

- a globally unique identifier
- once assigned, it refers to a single object and is not re-assigned (persistent)
  - Digital Object Identifier (DOI) is one form, widely used for data
  - ORCHID iD for researchers
- foundational to advancing FAIR principles
- advances data sharing
- underpin citation of datasets
- Help to connect people and outputs
- Help to connect outputs to funding and organizations

What is ORCID?

Open **R**esearcher and **C**ontributor **Id**entifier

ORCID iD is a 16-digit unique identifier

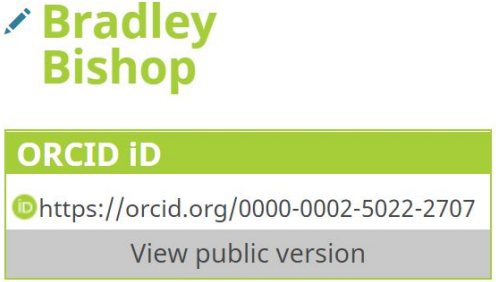
Free to register and use your iD

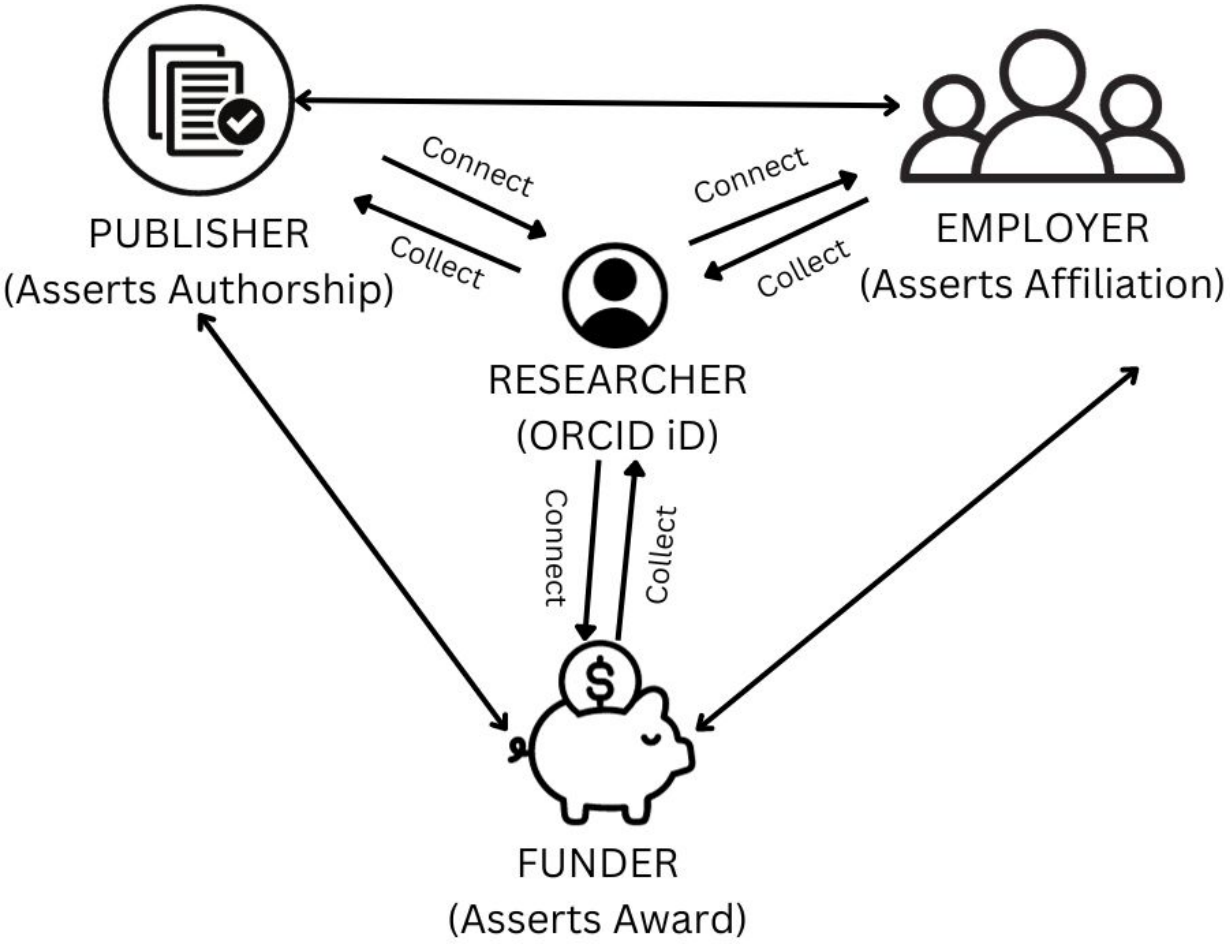
orcid.org/register

Orchid ID captures biographical details

ORCID is used around the world by

- Publishers
- Research organizations
- Funders




PUBLISHER
(Asserts Authorship)
Connect
Collect
Connect
Collect
EMPLOYER
(Asserts Affiliation)
RESEARCHER
(ORCID iD)
Connect
Collect
FUNDER
(Asserts Award)

ORCID iD https://orcid.org/register

- distinguishes you and ensures your research outputs and activities are correctly attributed to you
- reliably and easily connects you with your contributions and affiliations
- improves recognition and discoverability for you and your research outputs
- is interoperable
- is persistent

Shared names, different versions, transliteration

Problems Arose

- Despite these external pressures to create and follow DMPs, the compliance with these requirements has lagged. For example, one study evaluated 119 DMPs and found that 51% did not identify the individual(s) responsible for data management, which is consistent with prior research findings (Van Loon et al., 2017).
- Also, the DMP benefits purported have been assumed, without direct, systematic study (i.e., proof).
- In fact without evaluation or any follow-up, DMP outcomes cannot be measured to inform best practices and support these assumed benefits.
- Most researchers do not follow their plans.

Potential Solutions

- Ideally, the DMP system used would:

1) allow for iterative interaction for the purpose of review and updating;
2) focus on the data collection, processing and analysis methodologies/methods relevant to the specific field of research;
3) be easily integrated into researcher workflows; and,
4) be scaffolded through DMP training (face-to-face or online) and other associated resources and materials that form part of a DMP system, potentially consisting of:
    - A training manual/online modules
    - An institutional storage options chart
    - A referral map of all research support services across the institution
    - A DMP self-assessment rubric

Data Management Plan Scorecard

Bradley Wade Bishop[1], Judit Ungvari[2,3], Hannah Gunderman[4], Heather Moulaison-Sandy[5]

1: University of Tennessee, United States of America; 2: Belmont Forum, Uruguay; 3: Florida Museum of Natural History, United States of America; 4: Carnegie Mellon University, United States of America; 5: University of Missouri, United States of America

DMPs Rubrics: Prior Work

- Despite their importance to the research enterprise, no method currently exists to *evaluate* DMPs (i.e., determining the value of the DMPs and allowing for across-the-board comparison).
- Prior work to assess DMPs has been rubrics limited in overall usability, e.g. the Document Assessment and Review Tool (DART) project, which made a rubric to assess NSF DMPs; its exhaustive list of elements to assess and the amount of time the rubric took to use were impediments to adoption (Roland et al., 2015).
    - Other rubrics have been developed (e.g., Poole & Garwood, 2020; Van Loon et al., 2017), but included all elements as binary.
    - Prior rubrics focused on completeness of DMPs and were not intended to evaluate quality (Carlson, 2017).

Purpose

- The Belmont Forum DMP scorecard addresses the need for a tool to easily evaluate (i.e., *score*) DMPs, going beyond assessment to evaluate the quality of the DMPs with examples in the rubric to improve quality of the DMP content.
- This paper presents the development of a DMP scorecard (i.e., DMP evaluation tool) for a large, international funding agency, the Belmont Forum.
- A project funded by the Belmont Forum must include collaborators from at least three countries and therefore these projects must consider a number of different policies and strategies at different levels, including ones from funders, nations, and larger governmental bodies (i.e., EU).

Data and Digital Outputs Management Plan

- The DDOMP includes Data and Digital Outputs, but is not limited to:
    - Quantitative and qualitative digital data created during research activities;
    - All metadata describing the data and digital outputs;
    - The associated code, software, workflows, and provenance information; and
    - Stakeholder-oriented digital outputs such as maps, videos, white papers, and so forth.

- DDOMP is required of all Belmont Forum-funded projects.

DMP Scorecard

- The Belmont Forum provides training materials for applicants and awardees (https://bfe-inf.github.io/toolkit/index.html).
- DDOMPs may be refined and expanded as changes occur in any large project. This allows these DMPs to serve as dynamic documentation throughout a project.
- The DMP scorecard consists of 14 applicable criteria each on a three point scale (0-2). For example, for data sharing beyond the project team—if a data repository is named, 2 points are earned; however, if language is included that data will be made available, but no data repositories are identified, then that receives a score of 1.

Scoring

- Once all of the applicable criteria have been reviewed, the scores can be totalled and divided by the number of applicable criteria to produce a average score.
- Those scoring higher than an average of 1 provide more detail and those less than 1 should be revised as some required elements are missing and/or incomplete.
- For example, if the average over 14 applicable criteria is 0.8, that DDOMP needs more specific responses to address required criteria in the plan.

Current Use

- The scorecard was used by four trained reviewers on all 21 multi-national projects' DMPs and received an average score of 0.8. Fourteen of the projects' average scores were below a 1, which indicated several criteria of a quality DDOMP were completely ignored.
- Following a full-day of training attended by at least one team member from each project and encouraging awardees to complete other modules (https://bfe-inf.github.io/toolkit/index.html) each team was given six weeks to update their DMPs.
- All revised DMPs received passing scores with an average of 1.4.

Conclusion

- Given guidelines must be broadly applicable, much of the DMP guidance and assessment remains hard to operationalize for both researchers and evaluators.
- This scorecard and its subsequent use may lead to improved DMPs simply because it gives some expectations to researchers. The assumption that DMPs are of high quality without evaluation requires further investigation.
- The understanding that better data management leads to scientific advancement should appreciate DMPs value enough to evaluate these integral investments.

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)
1.1 Plan lists the types of data and other digital outputs of long-term value. *(e.g. text, databases, images, 3D models, software, audio files, code, video files, reports, surveys, patient records, samples, and so forth)*	Datasets and other digital outputs of long-term value are identified, including data type and encoding. *"Environmental data will be delivered as NetCDF (Network Common Data Format) files. Raster files will use the raster2pgsql PostGIS module. Maps and other geographic data will use shapefiles."*	Datasets and other digital outputs of long-term value are identified, but lack detail for users beyond the project to understand. *"A combination of geo-referenced data at various spatial, temporal, and taxonomic scales (e.g., populations, regions, nations, circumpolar, biomes, habitats) will comprise our data of long-term value."*	No information about data types is included.
1.2 Plan describes how the data and other digital outputs will be collected, captured, or created. *(e.g., new observations, results from models, reuse of other data, or other)*	Clearly defines how data will be collected, captured or created, including methods, instruments, software, or infrastructure where relevant. *"Socio-economic data will include household food security, nutrition, and demographic data. Spatial data produced will include ground-truthed land use/land cover data ~3 km from 50 farms, land use scenario maps for 12 villages and 4 regions. All ecological/social data will be recorded on physical datasheets and entered directly into Excel or STATA."*	Missing some details regarding how some of the data will be produced; makes assumptions about reviewer knowledge of methods or practices. *"Models will produce a broad range of output simulation data."* *"Data collection includes gathering in-the-field various phytoplankton, zooplankton, fish and flooded forest biodiversity."*	No information about data collection, capture or creation.
1.3 Plan lists (quantifies) how much data is anticipated.	Datasets and other digital outputs volume estimated. *"The project is expected to produce up to 9 GB of new data."*	Datasets and other digital outputs amount is vaguely estimated or partially described. *"A new database will be made."*	Amount of expected data is not mentioned.

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)
2.1. Plan specifically addresses metadata standards or formats for the data and other digital objects. (e.g., Dublin Core, Ecological Metadata Language (EML), W3C RDF (Resource Description Framework), and so on)	The metadata standard or format is named for the datasets and other digital outputs. *“Data representation will be guided by FAIR principles regarding metadata (e.g. following ISO 19115:1 for geographic information).”* *“Data from the project will have descriptive metadata according the DataCite Metadata scheme 4.0. ”*	The metadata standard or format is not named for the datasets or digital outputs, but metadata is mentioned. *"Deliverables from the project will have key-words and appropriate metadata to maximize discoverability."* *“Deposited data will be accompanied by a range of metadata, such as geographical coordinates, dates, taxonomic information and information on publications and authors.”*	The metadata standards or formats is not named.
2.2 Plan describes when data and other digital outputs will be made available outside and within the project team.	The plan defines when the data and other digital outputs will be made available outside and within the project team. *"Data generated from our research will be shared incrementally throughout the time period of the project, and the full dataset will be available no later than one year after the project ends or at publication of our research."*	The plan provides only a vague timeframe for when the data and other digital outputs will be made available outside and within the project team. *“It will be the policy of the project to publish relevant findings expeditiously in the peer-reviewed literature.”*	The plan does not estimate when the data and other digital outputs will be made available outside and within the project team.
2.3 Plan describes how data and other digital outputs will be made available beyond the project team.	The data repository is named. *"The data that accompany publications or other published deliverables will be made available by using an online repository for ecological data (e.g. DRYAD, http://datadryad.org/)."*	The data repository is not named, but some language exists to indicate that the data will be made available. *”The non-sensitive data collected as part of this project will be made available upon reasonable written request for academic, research, educational and other not-for-profit professional purposes (e.g. conservation, environmental policy) according to the DMP.”*	Does not provide information on how the data and other digital outputs will be made available.

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)
3. Plan describes which member(s) of the team will be responsible for developing, implementing, overseeing, and updating the DDOMP.	The plan indicates by name or role the member(s) responsible for developing, implementing, overseeing, and updating the DDOMP. *"The data manager (WP4.Task 3 leader) and NAME OMITTED partner will ensure the data management of the project."* *"A full-time data manager responsible for overseeing, implementing and updating this data management plan will be hired upon award of funding."*	The plan indicates that developing, implementing, overseeing, and updating the DDOMP will be done, but in vague qualifiers, or by multiple team members working independently, or done passively. *"Overseeing and updating the data management approach during the project period will be the responsibility of each team's leader(s) of data collection work in the different WPs. Each leader will be responsible for developing and implementing their respective data management plans."*	The plan does not indicate the member(s) or a team role responsible for developing, implementing, overseeing, and updating the DDOMP.
4. The plan recognizes describes the security measures to prevent unauthorized access to the data and other digital outputs. *Data security refers to the process of protecting data from unauthorized access and data corruption throughout its lifecycle. Data security includes data encryption, tokenization, and key management practices that protect data across all applications and platforms.*	The plan describes security measures for the data and other digital outputs to prevent unauthorized access with specific examples. *"The data security plan includes: i) data input in REDcap or other secure data management platform; ii) storage within an encrypted drive with access controlled by research team members; and iii) a process for managing human research ethics that complies with national funding body guidelines."*	The plan describes security measures for the data and other digital outputs to prevent unauthorized access, but in vague terms. *"Data will be stored on the secure research servers of the team's respective universities during the project."*	The plan does not describe security measures for the data and other digital outputs to prevent unauthorized access.

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)
5.1 Plan indicates how long the data and other digital outputs will be retained	The plan indicates the specific timeframe for how long the data and other digital outputs will be retained. *"These data will be stored for at least 10 years after the project has ended using the institutional infrastructure of (named partner)."* *"At publication of results, all data will be stored in a trustworthy and/or certified repository; after 5 years, data will be archived for permanent storage in the repository.".*	The plan indicates how long the data and other digital outputs will be retained, with a vague time, including *in perpetuity*. *"This archive will be kept for at least several years (with the intention of making these available for a longer time)."*	The plan does not indicate how long the data and other digital outputs will be retained.
5.2 Plan indicates who will be responsible for managing data *after the project ends* to ensure their long-term accessibility.	The plan indicates who or what entity will be responsible for managing after the project ends. *"The Data Manager, responsible for developing, implementing and updating the Data Management Plan, will ensure data are archived for permanent storage in a certified repository after publication of our research."*	The plan indicates management of the data and digital outputs after the project ends, but in vague qualifiers. *"Project leads from each country will be responsible for ensuring long-term access to data from their country."*	The plan does not indicate management of the data and digital outputs after the project ends.

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)
6.1 (if applicable) Plan describes how sensitive data and other digital outputs will be made available beyond the project team. (Note: human subjects research can be made open if informed consent contains the language)	The plan explains how sensitive data and other digital outputs will be made available outside the project team and outlines the specific considerations. *"Sensitive data (human identifiers, endangered species, etc.) will be anonymized before publication. A letter explaining the purpose, approach and dissemination strategy of the research will be prepared, translated into the relevant languages and shared with all participants in any event."* *"In order to comply to the EU General Data Protection Regulation (GDPR), we will not provide data which can be linked to individuals or certain locations."*	The plan explains how sensitive data and other digital outputs will be made available beyond the project team, but in a vague way. *"These data will be made available beyond the project team members only upon written request by qualified researchers."*	The plan does not explain how sensitive data and other digital outputs will be made available beyond the project team.
6.2 (if applicable) Plan describes any limitations on the ability to share data and other digital outputs *(e.g. proprietary nature, indigenous rights, or others)*	The plan describes any limitations on the ability to share data and other digital outputs with detailed reasoning to limit access. *"In recognition of the important socio-cultural and ecological information being collected about and with local communities, we will also honor the principles of community-owned research, drawing on the OCAP (Ownership, Control, Access and Possession) principles developed by First Nations communities in Canada (http://fnigc.ca/ocap.html). Storing hard copies of data at the Soils, Food, and Healthy Communities office will ensure that local community members can control and access the data. Researchers interested in using the hard copy data will need to sign an agreement of*	The plan describes any limitations on the ability to share data and other digital outputs with vague terms and conditions. *"Our data will be accessible after the end of the project with a specific license on how to use them in both scientific and other contexts; appropriate licensing will be selected to avoid or minimize misuse or over-interpretation of the results by non-informed people."*	The plan does not describe any limitations on the ability to share data and other digital outputs.

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)
7.1 Plan describes the intellectual property rights to the data and other digital outputs	The plan indicates the intellectual property rights to the data and other digital outputs in detail. *"The Intellectual Property Rights (IPR) relating to all project data, analyses, and outputs will be negotiated and specified in the Consortium Agreement and the Data Management Plan. The Steering Committee will maintain an IPR Directory throughout the lifetime of the project, which will inventory all items of knowledge relating to the work of the project and make explicit for each item its owner, nature, accessibility status and dissemination and protection measures."*	The plan indicates the intellectual property rights to the data and other digital outputs in detail, but with vague terms. *"Public data are managed according to the highest standards available and will be available for reuse upon request in compliance with the data's reuse license."*	The plan does not indicate the intellectual property rights to the data and other digital outputs.
7.2 Plan describes licensing of the data and other digital outputs	The plan describes licensing of the data and other digital outputs by naming a license(s). *"We will use CC-BY license (h ttps://creativecommons.org/licenses/by/4.0/)."*	The plan describes licensing of the data and other digital outputs, but in vague terms. Provides a general overview of how data may or may not be reused, or the applicability of the policy can be inferred from general/broad/blanket statements about data being made open or being kept private, or policies can be inferred based on the sharing location. *"Upon publication of our research, data will be accessible under open access and a specific license to avoid any misuse."*	The plan does not describe licensing of the data and other digital outputs.

Criteria	Complete Response (score = 2)	Incomplete Response (score = 1)	No Response (score = 0)
8. Plan describes the supporting documentation and metadata that will be created to make data and digital outputs publicly accessible.	The plan includes supporting documentation and metadata on how to make data and digital outputs publicly accessible with specifics. *"i) On the project website we will upload project's public results, e.g. conference presentations, workshops, scientific papers, deliverables / milestones.* *ii) For the produced project datasets, descriptive metadata will be available according the DataCite Metadata scheme 4.0. All data sets of the project will be stored in X data repository.* *iii) We plan open access data publication (e.g. Copernicus publication), and iv) open access scientific papers."*	The plan includes supporting documentation and other information (e.g., metadata) on how to make data and digital outputs publicly accessible in vague detail. *"In our published results, we will provide supporting documentation (e.g., meta-data) that explain how the data were collected, provide an overview of dataset contents and variables, and specify the point of contact for accessing the data."*	The plan does include supporting documentation and other information (e.g., metadata).
9. Plan specifies the costs or estimated costs associated with long-term data management or an assigned data manager role.	The plan includes information on institutional, national, or other fiscal support for long-term data management and names the funder. *"As a publicly funded authority, Organization X guarantees the long-term availability of the data through already existing institutional storage infrastructure, which also means we can account for costs of long-term storage of the data."*	The plan includes information on institutional, national, or other fiscal support for long-term data management in vague detail. *"The costs of maintaining data to be collected are already taken care of outside the project through the respective affiliate institutions of the research team."*	The plan does not address the costs associated with long-term data management.