Draft White Paper

Active Data Management Plans: A metadata-driven model for Data Management Plans

Data Management Plans (DMPs) are an effective tool for documenting the management of research data throughout the full research data lifecycle[[1]](#footnote-0). Recently, prominent communities and groups interested in furthering standards for research data have raised concerns about the current ability of DMPs to meet researcher needs around documenting data management practices, and moreover, the ability to reuse DMP information that is required by funders, administrators, institutional review and ethics boards, depositors, curators etc., thus adding to the list of expectations of researchers. Access to authoritative, machine-exchangeable documentation about research data management is crucial for compliance around open science, as well as for administrators seeking to understand institutional research needs around data management infrastructure and resources.

It is our position that a metadata-driven model for DMPs would enhance the value of the information they contain and would facilitate the exchange of DMP information with other systems and platforms managing aspects of data throughout the research lifecycle. A metadata model would structure DMP information in understandable and predictable ways, enabling the reuse and repurposing of the content by researchers, other systems, and a variety of stakeholders in the research process. An agreement on the concepts and definitions of such a metadata model would advance a common understanding of a shared vocabulary around data management practices. While DMPs are currently viewed as static documents, the move to ‘active’ or ‘living’ DMPs would be best achieved through a metadata standard, documenting research data management activities and facilitating the flexible update of information by researchers.

There are several communities interested in developing DMPs further to meet evolving researcher needs:

* Active Data Management Plans Interest Group, Research Data Alliance (RDA)
  + See [Note on Data Management Plans](https://www.rd-alliance.org/sites/default/files/Note_on_Data_Management_Plans-v4.docx) (ADMP-IG, 2015)
  + See [E-mail thread in response to Note on Data Management Plans](https://www.rd-alliance.org/group/active-data-management-plans-ig/post/short-note-dmps.html) (ADMP-IG, 2015)
* JISC Ethics Review DMP Profile Elements (JISC-CASRAI)
  + See [Dictionary](http://dictionary.casrai.org/DMP_Ethics_Review)
* DDI Alliance
  + See [Active DMP Working Group](https://ddi-alliance.atlassian.net/wiki/display/DDI4/Active+Data+Management+Plans+Team)
  + See Position Paper: [Data Management Planning and the Data Documentation Initiative (DDI)](http://blogs.ucl.ac.uk/dmp-ss/files/2013/05/Data-Management-Planning-and-the-Data-Documentation-Initiative_1_0.pdf) (T. Castillo & A. Gregory, 2013)
  + See DDI-L Specification WG
* Digital Curation Centre (DCC)
  + See DMPOnline
  + See [DMPOnline Roadmap](https://dmponline.dcc.ac.uk/roadmap) (2015)
  + See [DMP Lifecycle model and API development](http://www.dmao.info/blog/2015/07/03/dmponline-api-and-dmaonline.html)
* Portage DMP Working Group, Canadian Association of Research Libraries (CARL)
  + See CARL-Portage [Data Management Plan generic template](https://dmp.library.ualberta.ca/)
* [Data Management Plans as a Research Tool (DART)](http://dmpresearch.library.oregonstate.edu/)

Currently, no common, internationally accepted model exists for Data Management Plans or information about data management planning activities. Yet, requirements exists to access, update continuously, exchange, and reuse this information throughout the research data lifecycle. We have identified work that has begun in this area in the the initiatives outlined above. Most notably, the recent product of a JISC and CASRAI working group on terms and definitions for use in DMPs to address the ethical treatment of research data had laid some groundwork for structured information in this category of research administrative information.

**The DDI Alliance’s call for action**

The DDI Alliance is an organisation maintaining and developing international metadata standards and vocabularies for the description of research data in the social and behavioural sciences. Recently, the Alliance formed a working group tasked with addressing the inclusion of DMP information in future iterations of DDI specifications. The DDI Active Data Management Working Group (DDI-ADMWG) is comprised of researchers, technical scientists, administrators, and librarians interested in expanding DDI specifications to incorporate DMP information and metadata in a structured manner.

DDI-Lifecycle, the DDI specification that allows for description of research data and activities surrounding the management of data throughout the entire data lifecycle, is well suited for interoperability with any metadata standard or model that specifically addresses DMP information. As researchers begin to document data management and planning activities in DMPs, information should be reusable in a variety of complementary systems, tools, processes, and activities related to research data management. Information collected in DMPs is also sought by granting and funding agencies, institutional administrators and research offices, platforms and repositories for data sharing, access, and preservation. Common sense suggests the information in a DMP should be readily adapted and reused for other purposes. The best approach would be an underlying metadata model that describes this information in a well-documented, understood, shared, and reused way. The DDI-ADMWG supports all efforts within the international data community to establish a metadata-driven model for DMPs.

**What is a metadata-driven model for DMPs?**

A metadata driven model for DMPs is a commonly agreed upon set of concepts, definitions, and relationships that can be expressed in a basic web-enabled information technology, such as, a database schema, XML, or RDF. The implementation of such a metadata model should support machine-to-machine as well as machine-to-human exchanges of information.

**Why a standardize metadata model for DMPs?**

* To build a common understanding of and shared use of a vocabulary for data management planning activities and practices.
* To reduce barriers to interdisciplinary research and understanding.
* To facilitate the exchange and transfer of structured information between systems, including access to and reuse of information, the flexible updating of information (e.g. the transfer between ethics and compliance systems, funding and grant agency application systems, repositories, deposit platforms, review and publication systems etc.)

**DDI and interoperability with DMPs**

A review of the DDI specifications and their appropriateness for describing DMP information[[2]](#footnote-1) provides guidance on how best to approach a DDI metadata-driven model for DMPs. While the current DDI standard does not address DMPs and their content adequately, there is a need for development in this area. Given the variety of DMPs used globally, we find it difficult to envision a common standard for all DMPs. Nevertheless a collaborative and flexible approach may assist in establishing standards to reduce duplication and redundancy in this area.

**Opportunities and considerations for DDI and a metadata-driven model for DMPs:**

**Initial DDI development and directions**

1. Data Management Plans would have a model of the information it requires, which could be agreed upon internationally. A mapping of DDI-Lifecycle and DDI-Codebook to a model of DMP information would be established to facilitate interoperability and exchange of information.

**Steps / Actions:**

* 1. Identify standards and common elements found within DMP Profiles that can be used as the basis for a DMP model that could be brought into and referenced in DDI;
  2. Ensure appropriate interoperability with identified external standards, such as JISC - CASRAI’s Ethics Review Glossary;
  3. Develop a standard metadata model for the structure of a DMP information (that can map to DDI-Codebook and DDI-Lifecycle as best as possible);
  4. Identify a list of elements / activities that are not currently addressed or are not mappable to DDI for consideration and incorporation by the DDI-Model specification and development teams. This could include reference to a DMP itself and related activities.
  5. Encourage DMP developers and vendors to adopt a standard metadata-model for DMP information that is agreed upon internationally.
  6. Develop and promote DDI tools and crosswalks that can be readily implemented in DMP tools for greater exchange of information across the research lifecycle.

**Other potential opportunities for DDI**

1. Establish a reference to DMP information within DDI specifications having identified an appropriate place in the structure. This approach could be combined with having a “native” DDI structure for holding the information.
2. “A *super-set* of data management planning information, collected across various examples of DMPs, could also be created and used as the basis for a DDI model. This approach does not solve the problem of developing an international standard, but would at least mean that DDI would address the needs of specific implementations across the globe.” (Tito & Gregory, 2013)
   1. Develop a “super-set standard” that other standards can be identified and brought together under a new DDI-model specification, e.g., NSF DMP Profile, DCC DMP Profile, or Portage DMP Profile elements.

**Example metadata mapping: JISC-CASRAI Ethics Review Glossary and DDI[[3]](#footnote-2)**

Through an example of metadata mapping provided below (see Notes), there are a number of administrative activities missing from current iterations of DDI specifications.

* How do we add or integrate new DMP related information such as Ethics Review into the DDI-Model specification?
* Do DMPs and related activities constitute a DMP Model view and be explicitly referenced in DDI? How would this be determined in terms of DDI-Model objects and attributes?

**Challenges with the different supported DDI specifications:**

There are several major challenges outlined by Tito and Gregory, 2013 for supporting and integrating DMP information into current iterations of DDI specifications.

1. One major difficulty is that the DDI-Codebook is designed to describe datasets after the data have been collected. Whereas Active DMPs capture information throughout the data management planning and early stages of collection. The most natural fit with DDI-Codebook would be in the administrative details at the study-level, describing the overall project, and then to reference the data management plan itself, using a citation and link, so that the full set of information can be expressed and referenced appropriately.
2. DDI-Lifecycle is designed to describe the entire process of data collection, starting with the concepts and universe information about the study and then proceeding through data collection, data processing, dissemination, and archiving. While there is interest in addressing data management planning and incorporating this information in future versions of the standard, this information is not present today and would need to be addressed by the DDI Development Team.
3. One of the difficulties faced by DDI when it comes to DMP information is that there is no accepted international practice regarding the structure and organization of such information. However, DDI must be useful to an international audience – it has implementers and vendors across the globe. For example, a standard design to describe a data management plan in the UK may not fit the structure for such a document required by the National Science Foundation in the US.

**DDI-ADMWG next steps**

The DDI-ADMWG plans to communicate its call for action on a metadata-driven model for DMPs and at the same time promote better interoperability of common DMP information within DDI. This can include producing a conceptual metadata mapping of common DMP elements (such as the DCC-generic template) to current DDI-Lifecycle elements.

The primary goals of this group and its activities are to identify areas of development within the DDI to facilitate exchange of DMP information, act as a partner in the area of metadata standards for DMPs, and to provide recommendations about the flow of information in a systematic way to further develop and adopt DMPs within the scholarly research community. We are excited about the future of DMPs and what they can provide in terms of improving data stewardship across a variety of scholarly disciplines.

**DDI-ADMWG call for action on DDI development**

Seven external experts will join the DDI development team at the Leibniz-Zentrum für Informatik in Dagstuhl, Germany on October 18 – 23 , 2015 to evaluate progress of the model-driven DDI. This event is titled: [DDI Moving Forward: Facilitating Interoperability and Collaboration with Other Metadata Standards](http://www.dagstuhl.de/en/program/calendar/evhp/?semnr=15433). This would be an excellent opportunity for the DDI development team to explore approaches within DDI-Model to interoperate with DMPs. To assist such an exercise, the DDI-ADMWG has prepared a few use cases based on DMP applications for the DDI development team to examine.

**Possible use cases for DDI-DMP Interoperability**

1. A researcher creates a DMP using a web-based tool and wishes to reuse this information to deposit her data in a repository, such as Dataverse.
   1. Research enters administrative and project-level information
   2. Information is documented using a DMP standard metadata model
   3. The DMP tool supports mapping to DDI-study level elements
   4. The DMP tool connects researchers to deposit their data in a data repository, such as Dataverse, and DDI-study level elements are populated in a new Dataverse.
2. An administrator at X University would like to review the DMPs that have been prepared as part of grant proposal at the institution for a given academic department.
   1. The DMP tool allows the administrator to review a DMP and to transfer information from her administrative information system to the DMP that incorporates institutionally supported research data management services.;
   2. The administrator also extracts information from the DMP that is accumulated with other DMPs for use in estimating demands on institutional resources, such as, storage space.

3) Research Service Office (RSO) uses DMP information to document compliance requirements for sensitive data.

1. An RSO officer submits a request from her administrative information system to capture information under the DMP Ethics section regarding the treatment of any sensitive data.
2. The information exchanged flags sensitive data and provides a match to an external Ethics Review system, which is used to identify the status of this particular ethics application.
3. The results of the DMP and Ethics Review system are integrated into a compliance monitoring tool to assist the RSO officer in tracking the safe treatment of the data.

4) An official at a funding agency would like to confirm that the research data from a project has been deposited with a data repository identified in a DMP.

1. The official’s administrative information system queries the project’s DMP through its grant number and requests information about the deposit of research data with a repository.
2. The DMP tool provides the deposit information and an DOI for the dataset, which had been incorporated into the DMP as a receipt by the data repository after the data had been deposited.
3. The official’s administrative information system confirms through the DOI that the data are publicly available and provides a summary of the dataset’s attributes.

**Notes**

Ethics Review Sub-elements mapping to DDI-Lifecycle

|  |  |  |
| --- | --- | --- |
| Ethics Review Element | DDI-Lifecycle & DDI-Codebook Elements | Notes |
| 1. [Principal Investigator](http://dictionary.casrai.org/Principal_Investigator) | <creator> |  |
| 1. [Principal Investigator/Family Name](http://dictionary.casrai.org/Principal_Investigator/Family_Name) | <IndividualName> / <LastFamily> | <IndividualName> / <FullName> |
| 1. [Principal Investigator/First name](http://dictionary.casrai.org/Principal_Investigator/First_name) | <IndividualName> / <FirstGiven> |  |
| 1. [Principal Investigator/ID Type](http://dictionary.casrai.org/Principal_Investigator/ID_Type) | <Individual> / <Individual Type> | or <ResearcherIdentification> / <TypeofID> |
| 1. [Principal Investigator/ID](http://dictionary.casrai.org/Principal_Investigator/ID) | <Individual> / <Individual ID> | or <ResearcherIdentification> / <URI> |
| 1. [Project](http://dictionary.casrai.org/Project) | ? |  |
| 1. [Project/Institutional Project ID](http://dictionary.casrai.org/Project/Institutional_Project_ID) | ? |  |
| 1. [Project/Name](http://dictionary.casrai.org/Project/Name) | ? |  |
| 1. [Research Dataset Contributor](http://dictionary.casrai.org/Research_Dataset_Contributor) |  |  |
| 1. [Research Dataset Contributor/Role](http://dictionary.casrai.org/Research_Dataset_Contributor/Role) |  |  |
| 1. [Research Dataset Contributor/Family Name](http://dictionary.casrai.org/Research_Dataset_Contributor/Family_Name) |  |  |
| 1. [Research Dataset Contributor/First Name](http://dictionary.casrai.org/Research_Dataset_Contributor/First_Name) |  |  |
| 1. [Research Dataset Contributor/ID Type](http://dictionary.casrai.org/Research_Dataset_Contributor/ID_Type) |  |  |
| 1. [Research Dataset Contributor/ID](http://dictionary.casrai.org/Research_Dataset_Contributor/ID) |  |  |
| 1. [Research Dataset Contributor/Institution ID Type](http://dictionary.casrai.org/Research_Dataset_Contributor/Institution_ID_Type) |  |  |
| 1. [Research Dataset Contributor/Institution ID](http://dictionary.casrai.org/Research_Dataset_Contributor/Institution_ID) |  |  |
| 1. [Research Dataset](http://dictionary.casrai.org/Research_Dataset) |  |  |
| 1. [Research Dataset/Storage Type](http://dictionary.casrai.org/Research_Dataset/Storage_Type) |  |  |
| 1. [Research Dataset/Storage Access](http://dictionary.casrai.org/Research_Dataset/Storage_Access) |  |  |
| 1. [Research Dataset/Content Description](http://dictionary.casrai.org/Research_Dataset/Content_Description) |  |  |
| 1. [Research Dataset/Destruction Policy](http://dictionary.casrai.org/Research_Dataset/Destruction_Policy) |  |  |
| 1. [Research Dataset/Embargo Start Date](http://dictionary.casrai.org/Research_Dataset/Embargo_Start_Date) |  |  |
| 1. [Research Dataset/Embargo End Date](http://dictionary.casrai.org/Research_Dataset/Embargo_End_Date) |  |  |
| 1. [Research Dataset/Preservation Start Date](http://dictionary.casrai.org/Research_Dataset/Preservation_Start_Date) |  |  |
| 1. [Research Dataset/Preservation End Date](http://dictionary.casrai.org/Research_Dataset/Preservation_End_Date) |  |  |
| 1. [Research Dataset/Capture Modes](http://dictionary.casrai.org/Research_Dataset/Capture_Modes) |  |  |
| 1. [Research Dataset/Ethics Compliance](http://dictionary.casrai.org/Research_Dataset/Ethics_Compliance) |  |  |
| 1. [Research Dataset/IPR Ownership](http://dictionary.casrai.org/Research_Dataset/IPR_Ownership) |  |  |
| 1. [Research Dataset/License](http://dictionary.casrai.org/Research_Dataset/License) |  |  |
| 1. [Research Dataset/Licensing of Existing Data](http://dictionary.casrai.org/Research_Dataset/Licensing_of_Existing_Data) |  |  |
| 1. [Research Dataset/Managed Access Procedures](http://dictionary.casrai.org/Research_Dataset/Managed_Access_Procedures) |  |  |
| 1. [Research Dataset/Managed Access Procedures](http://dictionary.casrai.org/Research_Dataset/Managed_Access_Procedures) |  |  |
| 1. [Research Dataset/Storage Security](http://dictionary.casrai.org/Research_Dataset/Storage_Security) |  |  |
| 1. [Research Dataset/Sharing Method](http://dictionary.casrai.org/Research_Dataset/Sharing_Method) |  |  |
| 1. [Research Dataset/Structural Approach](http://dictionary.casrai.org/Research_Dataset/Structural_Approach) |  |  |

1. See Research data lifecycle diagram [↑](#footnote-ref-0)
2. Tito & Gregory, 2013. [↑](#footnote-ref-1)
3. Ethics Review Sub-elements mapping to DDI-Lifecycle. See ‘Notes’ section. [↑](#footnote-ref-2)