Data Management Planning and the Data Documentation Initiative (DDI)

Version 1.0

Contents

I. Overview

Data management planning has become an increasing focus among the bodies which fund research, not only among the various research councils and other funders within the UK, but also elsewhere in Europe, and notably among the major research funders in the US. At the same time, the Data Documentation Initiative (DDI) metadata standards have become tools that are increasingly being used to facilitate data management solutions within organizations. Although based on a model of the data production lifecycle, DDI does not explicitly capture information regarding data management planning in its current versions. While there are some point s of similarity, the DDI fundamentally captures the metadata which data producers and archives have today – for the DDI, as for researchers more generally, data management planning is a relatively new area.

There is, however, strong interest among the members of the DDI community in being able to record the details of data management planning in a standard fashion, allowing the information to be exchanged freely between applications and organizations that use it.

This document presents the challenges faced by DDI in incorporating a new set of information – the description of data management plans – into the existing set of DDI information, suggesting some current points of overlap, and setting the stage for a future discussion regarding how data management planning might fit into the overall model of the data production lifecycle, and how it can address the wide variety of models used internationally to describe data management plans.

II. The Data Documentation Initiative (DDI) Standards

The Data Documentation Initiative (DDI) standards are the products of the membership-based DDI Alliance, currently housed within the Interuniversity Consortium for Policy and Social Research (ICPSR) based within the University of Michigan. The members of the DDI Alliance include data archives and libraries, statistical offices, research institutes and data producers.

There are two major work products from this group: DDI Codebook (formerly known as versions 1.0 to 2.1) and DDI Lifecycle (formerly known as DDI 3.0 and 3.1). These are complementary XML formats for encoding metadata for quantitative data files.

The DDI Codebook product is intended to be the description of a single study, and is used heavily by data archives throughout the world. It is essentially an XML data dictionary, supplemented by some additional information about the study itself: it describes variables, categories and codes, and other information regarding the data files for the study. DDI Codebook provides an after-the-fact description for data that has already been produced, for use within archives and in other settings where researchers need good metadata and documentation for the discovery and secondary use of data.

The DDI Lifecycle standard is intended to support more complex uses than DDI Codebook. It has a high-level model of the data lifecycle, and has many different modules which describe various phases, such as study concept, data collection, data processing, as well as archiving and dissemination. It provides XML formats for describing questionnaires, tabulation of data, and many other aspects of the data

throughout its production and use. DDI Lifecycle is a super-set of DDI Codebook, and it also supports the documenting of longitudinal and repeat-cross-sectional studies, which span time and cover many waves of data collection.

Often, when we think of metadata standards, we think of the very popular Dublin Core. DDI is not the same type of standard: it is much, much more detailed. Where Dublin Core has dozens of elements, DDI has hundreds. (In fact, DDI Lifecycle uses the native Dublin Core elements for citations.) The intended use of these standards is not the same: Dublin Core is essentially a citation standard; DDI is intended to allow for automatic processing of data sets and questionnaires by providing very detailed metadata. DDI Lifecycle, particularly, is intended to support both human and machine-actionable applications when working with data.

To give some simple examples of "machine-actionable" metadata, both versions of DDI allow for automatic transformation between different statistical packages like SAS, SPSS, and Stata. DDI Lifecycle allows for the automatic creation of online and print questionnaires from the metadata it holds regarding data collection.

As one can imagine, DDI has many pieces of metadata which stem from the data management planning process, although to date DDI has not addressed data management planning explicitly.

III. Data Management Plans and the DDI: Common Information

Some of the material contained in data management plans fits into sections of the DDI standards. There are two main production lines of the DDI, a simpler "codebook" version of the standard (DDI Codebook), and a newer "lifecycle" version of the standard. We will consider these two versions of the standard separately.

DDI Codebook (previously known as DDI versions 1.0 – 2.5) is designed to describe existing data sets and the studies that produced them in an after-the-fact fashion. The focus is to capture information first about the study itself, and then about the variables, describing how they are organized in the data file, and how they are represented (codes, categories, numeric types, string types, etc.). It is possible to capture information about the questions that populated the variables, and also to connect concepts to the variables. Some of the information in data management plans – especially that information which describes the project – has a place within this structure. Description of organizations, funding, project title, goals and purpose of the project, duration, and related administrative information can all be encoded using existing study-level DDI elements. However, much of the rest of the information concerning the data management plan itself does not have a good place to fit within the DDI Codebook structure.

The major difficulty here is that the DDI Codebook metadata format is designed to describe data sets that have already been collected. The most natural "fit" with DDI Codebook would be to put 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 in its natural format.

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 in future versions of the standard, this information is not present today. The situation is very similar to that of DDI Codebook – the study-level administrative information fits well into the existing structure, but the information that is more specific about data management planning is not. One area where DDI Lifecycle provides more detail is in the area of the intended methods for data capture – there are several DDI lifecycle elements which can be used to describe or link to survey instruments, to describe software, to describe methodologies, and to provide other information about data capture.

Again, however, a best practice would be also to include a link from the DDI Lifecycle description of a study to the full data management plan, even if there is some overlap in content. Ideally, however, a coherent structure would be provided for describing all the details of data management plans, and would be attached to the DDI structure at an appropriate place in the lifecycle, during the study planning and conception phase. The challenges to doing this are addressed below.

Alternately, DMP information could be explicitly referenced from within the DDI at an appropriate point in the structure – this approach could be combined with having a "native" DDI structure for holding the information.

IV. DDI's International Focus and the DMP Landscape

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; regardless, DDI must be useful to an international audience – it has implementers across the globe. Thus, a standard designed to describe a data management plan in the UK does not fit well into the structure of such a document as required by the National Science Foundation in the US, for example. The major challenge facing DDI, therefore, is how to describe data management plans in a generic way that is still useful in providing technology solutions, regardless of the national framework.

There are several ways in which the DDI Alliance might structure their standards, given this challenge. Ideally, data management planning would itself have a model of the information it requires, which could be agreed internationally. While such a standard may emerge, it is likely to require a good deal of focused time and effort. The DDI itself could play a role in developing such a standard, possibly. However, that model does not exist today.

A "super-set" of data management planning information 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.

V. Challenges to Modelling Data Management Information

To be useful in the near term, DDI has some mechanisms which might allow for the inclusion of more complete DMP information into the lifecycle, while still producing useful DDI metadata. DDI provides a standard extension mechanism, which allows users of DDI to declare their additions to the standard in a known fashion, using XML schema extensions. This is a technique which could be used to incorporate additional metadata into otherwise-standard DDI messages, and it is one which is used in many DDI implementations today. However, it has the down-side of putting DMP information into a structure which, although described in a standard fashion, may not be accessible by tools which support the DDI standard.

A less-sophisticated technique is to reference the data management plan itself as an external document, as described above. This provides a link to the DMP information, but does not necessarily allow for machine-actionable functionality – the link may only be to a human-readable document, where a processable format such as XML might be more desirable. The human-readable format may be enough for some purposes, but denies the kind of re-use of information for which the DDI itself is designed.

It would, of course, be possible to establish a best practice where the reference to an external DMP would always be to a standard structured form of the information, which would overcome this difficulty.

Ultimately, DDI standards should provide a mechanism for capturing and transmitting standard information about data management planning between applications. There is support for this view within the DDI community, notably from the ICH in the UK, and from CISER at Cornell University in the US. Further, the DDI lifecycle model should include data management planning as an integral part of the data lifecycle, given its increasing importance. Most important is having an agreed model, and an agreed set of information for incorporation into the DDI – the generation of an XML format is a straightforward task once a model has been agreed.

In looking at the various structures for data management plans, the one developed by the research councils in the UK provides the most comprehensive example, compared to others found elsewhere in the world, and is probably the best starting point for developing a standard model for incorporation into the DDI.

VI. The ICH-DMP Project

One example of a forward-looking implementation around data management planning and metadata reuse is the project recently conducted by the Institute for Child health in the UK. This project provides an example of how the challenges faced by the creation and reuse of DMP metadata can be successfully met.

Fundamentally, this system has two different types of metadata: metadata describing the structure or template of the needed data management plan, and another set of metadata which are the contents of the plan itself. This project used this metadata to produce a portal, allowing the structures and contents of a range of data management plans – those used within the UK – to be visible for re-use. It also

provides tools to create XML for the plans, and to validate them against the template. This approach is an interesting one, and the basic idea is illustrated below.

It is important to understand the problem: you have a range of possible topics within a data management plan, based on the requirements of funders, etc. This project uses the template description to capture this information, in reference to a super-set model of all possible topics (see the annexes for a listing of the super-set model used in this project). Thus, we get a template for a specific DMP as appropriate for a particular funder.
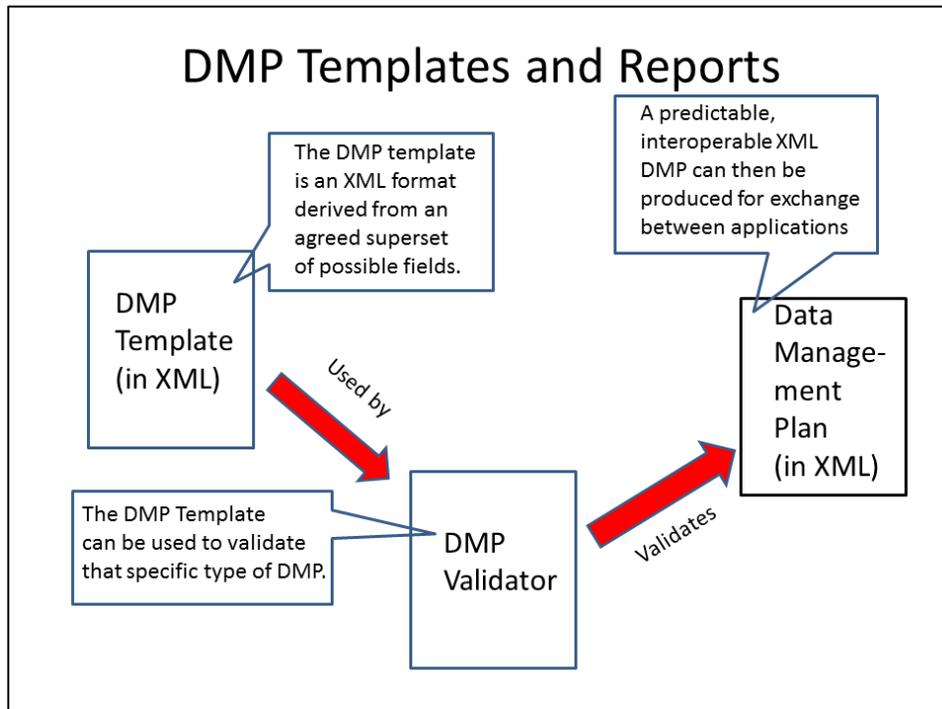
The benefit of this approach is that all the different possible DMP formats can be expressed in a standard XML, while still having the simplest possible XML structure for any given flavor of DMP. Software can be built to support the super-set model, if needed (as in the case of the ICH DMP project) or can be developed to support the flavor used only by a specific funder.

If this approach were to be used internationally by the DDI Alliance, the following steps could be taken, based on the ICH-DMP example, given that the DDI needs to produce XML schemas as a generic validation tool:

(1) An analysis of all current DMP structures would need to be conducted, to produce a superset model for DMPs internationally;

(2) A standard XML for describing DMP templates in reference to the super-set model would be created;

(3) Predictable rules for transforming the template XML into XML schema (xsd) would be developed and documented. (This replaces the functions of the ICH-DMP validator in the picture above.)

This approach is not one which the DDI Alliance uses today, but there are some threads within the discussions around what the next-generation DDI will look like which would seem to fit well with this type of technology approach. Among these is the idea that the DDI would no longer be a single huge XML document type, but would incorporate many different XML messages, among which might be messages for DMP templates and for data management plans themselves. Another stated intent in the next generation of DDI is the move toward model-orientation, in which XML (and other) technology implementations would be predictably derived from a model. The ICH DMP approach is exactly this: a model of DMP metadata is transformed into two different types of XML message, according to predictable rules.

Thus, we find in the ICH DMP project a good model for how we might approach the challenging variety of metadata found within data management plans in an international context through the DDI.

An alternative to directly incorporating DMP metadata into the DDI would be to establish a standard, processable format similar to the ones developed to support the ICH-DMP project, covering both template and DMP metadata. This could be explicitly aligned with the structure of the DDI standards, so that the external DMP metadata could be referenced from the DDI, or transformations could pass it back and forth between the two formats as needed.

VII. Data Management Planning and the Data Lifecycle

Data management planning as an activity is becoming an increasingly important part of the data lifecycle, and it is appropriate that a standard such as DDI Lifecycle should address the needs of organizations and researchers as regards this information. Historically, DDI has reflected the needs or those who are producing, archiving, and using data files. As a result, the metadata contained within DDI addresses a specific set of topics which do not include the planning stages of data management. However, there is an argument to be made that such information belongs in a standard such as the DDI, which aims to provide useful metadata across the entire data lifecycle.

Data is increasingly being seen as an organizational asset – something to be shared and re-used by larger numbers of researchers. We see evidence of this in the increased focus on data citation, and on "enhanced publications" which link research papers with the data on which they are based. Also, among funders, there are attempts now to quantify the results of their investment in research. These broader themes reinforce the idea that data – no less than any other type of organizational resource – should be managed, and that management should be planned. Planning around data management is a critical activity for the producers of data, and potentially a subject of interest for those using and archiving the data afterward.

Further, the structures and contents of data management plans are a reusable asset, which – if made visible through projects such as the ICH DMP – could lead to the emergence of agreed best practice internationally. Today, it is not easy to see what other researchers are doing for data management, nor what other funders are requiring in data management plans. This could change, if the ICH DMP project's example is followed in places other than the UK.

The exchange and reuse of DMP information forms part of a comprehensive view of the entire data lifecycle, and one which appears to be growing in importance as data is increasingly understood to be an asset to be reused, and not simply the property of a single research project.

VIII. Engaging with the DDI Alliance

The DDI Alliance is a member-based consortia, where any member organization may propose features to be included in future versions of the standard, or for new work products to be created. Once proposed, a working group is typically formed, based on a call for participation issued by the DDI Alliance management. Outside experts may be included in the working groups if this seems desirable.

Currently, the DDI Alliance is discussing how the next major version of DDI Lifecycle will be designed, and what its scope will be.

If the research councils and other funders in the UK wished to engage with the DDI Alliance to provide better support for data management planning, then there are a few reasonable approaches which could be taken.

First, it would make sense that the desired outcomes be identified: would the research councils in the UK want to have DMP features included natively in the DDI standard, or do they wish to maintain a separate, aligned standard based on their existing model? As described above, either of these possibilities could be chosen, although the second one comes with the burden of maintaining and publishing a standard model and implementations of it (in XML, presumably), which may not be something the research councils and other funders in the UK wish to undertake.

Once a decision has been made in this regard, it would make sense to approach the members of the DDI Alliance who are most interested in data management planning. Notable among these is CISER, at the Cornell in the US. The current head of CISER, William Block, has been very active within the DDI Alliance and would be a good point of contact inside the DDI community. CISER is very interested in the topic of data management planning. It would be good then to informally discuss the needed functionality within the DDI with CISER staff, and other interested parties within the DDI community, so that a single proposal could be crafted with the agreement of all interested parties.

There are members of the DDI Alliance within the UK, and CISER is also a member of the Alliance. The research councils of other funders in the UK might also consider joining the DDI Alliance: currently, this costs 2500 USD annually. In any event, a member would be needed to put the proposal forward to the DDI Alliance, with the current director, Mary Vardigan, as the contact point. Once this was done, the normal process of creating a working group – whether to natively include DMP functionality, or to align with an externally maintained structure – could be followed.

At the current time, this can be done with the simple agreement of the DDI Alliance management. However, under the new bylaws being adopted, the creation of a working group will require a vote among the membership. In either case, it is likely that coverage of data management planning would be voted in as a useful addition to the DDI standards. Because of the current state of the DDI – notably the

discussions around the design of the next major version of DDI Lifecycle – the first half of 2013 is a particularly good time to raise the topic of data management planning as part of the data lifecycle.

IX. Conclusions

While the inclusion of data management planning information into the DDI makes sense, and is of interest to some organizations which use DDI today, the current versions of the standard do not capture this information in a satisfactory manner. What is needed is a standard model of data management planning information, which could be the basis of a future version of DDI which does provide for this information, at an appropriate point in the data lifecycle. The model of data management planning information developed in the UK is the most comprehensive and detailed structure of its kind, and should form the basis for any work to include such information in the DDI.

Further, the ICH DMP project gives us a good model of how a model-oriented approach could allow for widespread interoperability of DMP information, despite the wide variety of structures required by different funding agencies across the globe.

Annex: ICH DMP Superset Model and DDI Mappability

This annex lists out the superset model developed during the project, describing the possible fields used for data management plans within the UK. It also includes a column which gives an assessment of whether any given field can be mapped against DDI (even if only the Lifecycle version of that standard). This assessment is not absolute – determining whether a specific field fits into the DDI or not requires a degree of interpretation. However, if the mapping to DDI Lifecycle was conducted, the result would be a fragmented set of information spread out across at least 3 modules (StudyUnit, DataCollection, Archive) and would be similarly fragmented in a mapping to DDI Codebook. In neither case is such a mapping useful to software developers.

In many cases, there are specific fields within the ICH DMP model which can be generically included in broader, descriptive text fields in DDI. In these cases, the mapping is described as "possible". This type of mapping may or may not be something which is useful in a particular application; in those cases where the information is intended to be machine-actionable, this approach will typically not be very useful. For simple presentation to an end user, it may be sufficient.

Group-ing	DCC Qu. #	DCC Question	Type	Requ'd for Core DMP?	Default Guidance	Maps to DDI?
1	1	Introduction and Context	Heading	Yes	This section records administrative details which tie the plan to a particular project	
1	1.1	Basic Project Information	Heading	Yes		
1	1.1.1	Project Name	Text	Yes		If only one SU is associated with this DMP, can be the SU Title and Abbrev
1	1.1.2	Funding body/bodies	Text	No		Yes
1	1.1.3	Budget	Text	No		Yes
1	1.1.4	Duration	Text	No		Yes
1	1.1.5	Lead Partner Organization	Text	No		Yes
1	1.1.6	Other Partner Org.	Text	No		Yes
1	1.2	Short Description of the project's fundamental aims and purpose	Text	Yes	Information summarised from the main body of your research proposal will help potential re-users understand the purposes your data has been collected or created for, and they are unlikely	Yes

					to have access to your proposal. Briefly summarise what you set out to discover and how that is likely to affect the kind of data you collect or create and how.	
1	1.3	Related Policies	Heading	Yes	Some of the information you give in the remainder of the DMP will be determined by the content of other policies; these policies may also have additional requirements that are not covered here. In case of doubt it is helpful for data managers to know what other policies were in force when the DMP was written.	Possible (in Description)
1	1.3.1	Funding Body requirements relating to the creation of a data management plan	Text	Yes	Guidance: - DCC comparison of Research Funders' DMP Requirements (http://www.dcc.ac.uk/site s/default/files/documents/ publications/UK%20resea rch%20funder%20expect ations%20for%20data%2 0plan%20coverage.pdf)	No
1	1.3.2	Institution or research group guidelines	Text	Yes	For multi-partner projects, you may also wish to mention any formal consortium agreement agreed, e.g. on data sharing, publication, IPR.	As Other Material
1	1.3.3	Other policy-related dependencies	Text	Yes	Examples of other relevant policies may include institutional ethics, regulation, information governance, and guidance and requirements from the data centre to which the data will be submitted.	No
1	1.4	Basic Data Management plan information	Heading	Yes		
1	1.4.1	Date of Creation of this plan	Text	Yes	Recording date information is important for version control and placing the DMP in context	No

1	1.4.2	Aims and purpose of this plan	Text	Yes	Here you may wish to address the following: protecting IPR, protection of sensitive data, adding value, ensuring longer term access, etc.	No
1	1.4.3	Target Audience for this Plan	Text	Yes	Your target audience may be the researchers/data creators, the principal investigator, future data reusers, data librarians, and representatives of your funders.	No
2	2	Data Types, Formats, Standards and Capture Methods	Heading	Yes	It is of critical importance that research datasets are adequately documented. The information in this section will help you and any subsequent user understand why and how the data were created, what they represent, and whether they are likely to be compatible with other datasets.	No
2	2.1	Give a short description of the data being generated or reused in this research	Text	Yes	When describing the type of content to be created, you may wish to refer to the RIN data types as a way of classifying what you will create: Scientific experiments; Models or simulations; Observations; Raw data; Derived data; Canonical or reference data. (See "To Share or not to Share: Publication and Quality Assurance of Research Data Outputs", Research Information Network, 2008) (http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-notshare-research-data-outputs)You should also consider the implications of data volumes: do you have sufficient storage? Will the scale of the data pose challenges when sharing or transferring data between sites?	Yes (in Data Collection)

2	2.2	Existing Data	Heading	Yes		
2	2.2.1	Have you reviewed existing data, in your own institution and from third parties, to confirm that new data creation is necessary?	Boolean	Yes	Performing this check helps also helps to ensure the uniqueness of the research.	No
2	2.2.2	What existing datasets could you use or build upon?	Text	Yes	If none, enter "n/a"	No
2	2.2.3	Describe any access issues pertaining to the pertinent, existing data	Text	Yes	If relevant, include financial costs of accessing or using the data	Possibly
2	2.3	New Data	Heading	Yes		
2	2.3.1	Why do you need to capture/create new data?	Text	Yes	Reasons to capture/create new data will include: non-existence of suitable existing data; extending existing data to cover new areas; performing comparison over time.	No
2	2.3.2	Describe the process by which you will capture/create new data	Text	Yes	Here you should explain the capture process. If you're doing observations, how will they be recorded? (e.g. in a dated and numbered field notebook.) Also note what kind of equipment you will use and the software required. If you plan to use proprietary software, could you export to an Open format so the data can be reused more widely? You may also wish to cover: content selection; instrumentation; technologies and approaches chosen; file naming conventions; versioning; meeting user needs. Your answer should be sensitive to the location in	Yes (Lifecycle only)

					which data capture will take place. Guidance: - JISC digital media guidance on filenaming (http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/choosing-a-filename/) - University of Edinburgh Records Management file naming guidance(http://www.recordsmanagement.ed.ac.uk/InfoStaff/RMstaff/RMprojects/PP/FileNameRules/Rules.htm)	
2	2.3.3	Which file formats will you use, and why?	Text	Yes	Here you should outline and justify your choice of format, e.g. Microsoft Excel for recording measurements or SPSS for analysis, as these are in widespread use, the University has the relevant software licences or they're accepted standards in your field, etc. Decisions relating to file formats may also be made with recourse to staff expertise, a preference for Open formats, accepted standards, or widespread usage with a given community. Guidance: - UKDA Guidance on recommended data formats (http://www.dataarchive.ac.uk/sharing/acceptable.asp)	No
2	2.3.4	What criteria will you use for Quality Assurance/Management?	Text	Yes	Quality management mechanisms may include: documentation, calibration, validation, monitoring, transcription metadata, peer-review.	No

2	2.4	Relationship between old and new data	Heading	Yes		No
2	2.4.1	What is the relationship between the new dataset(s) and existing data?	Text	Yes	This is concerned less with existing data that may be used in the Research Activity, but rather with the disciplinary context. A typical answer might identify a body of data with which it would be helpful to harmonise newly generated data, or from which methodologies might be drawn, e.g. ISO standard materials testing data, time/motion studies data.	Possibly (Lifecycle only)
2	2.4.2	How will you manage integration between the data being gathered in the project and pre-existing data sources?	Text	Yes	Here you may wish to cover issues such as technical integration, provenance, trust and data quality.	Possibly (Lifecycke only)
2	2.4.3	What added value will the new data provide to existing datasets?	Text	Yes	Value which new data can bring to old may include: greater detail, wider coverage, verification of existing data, etc	No
2	2.5	Data Documentation and Metadata	Heading	Yes	Metadata is the information that makes your new data usable. NISO defines three main categories of metadata: Descriptive metadata is the information used to search and locate an object such as title, author, subjects, keywords, publisher; structural metadata gives a description of how the components of the object are organized; and administrative metadata refers to the technical information including file type. Two sub-types of administrative metadata are rights management	No

					metadata and preservation metadata. (Source: Wikipedia) Annotation briefing-paper (http://www.dcc.ac.uk/resources/briefingpapers/introduction-curation/annotation)	
2	2.5.1	Are the datasets which you will be capturing/creating self-explanatory, or understandable in isolation?	Boolean	Yes	You may wish to consider this from the perspective of a typical reader of a journal for your discipline.	No
2	2.5.2	If you answered No to DCC 2.5.1, what contextual details are needed to make the data you capture or collect meaningful?	Text	Yes	Think about what kind of documentation is needed for others to understand your data. This may include: a description of the data capture methods, explanation of data analysis, details of who has worked on the project and performed each task, etc. Guidance: - JISC Digital Media Introduction to Metadata (http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/an-introduction-tometadata/) - UKDA Guidance on Data Documentation and Metadata (http://www.dataarchive.ac.uk/sharing/metadata.asp)	No

2	2.5.3	How will you create or capture these metadata?	Text	Yes	You may wish to address the balance between automatic and manually created metadata. Creating documentation takes time so consider whether anything you're already creating can be used e.g. publications, websites, progress reports, etc. Also note where information about the data will be recorded e.g. in a database with links to each item, in a 'readme' text file, in file headers / under properties in Word or PDF. Guidance: - DCC Briefing Paper on Annotation (http://www.dcc.ac.uk/res ources/briefingpapers/intr oduction- curation/annotation)	No
2	2.5.4	What form will the metadata take?	Text	Yes	Where appropriate, give details of the standards used. Using standards such as Dublin Core and TEI can make your data interoperable, so consider what others in your field have used or follow data centre recommendations. Using controlled vocabularies for description will also help improve consistency. Guidance: - DCC Briefing Paper on Metadata Standards (http://www.dcc.ac.uk/res ources/briefing- papers/standards- watchpapers/what-are- metadata-standards)	Possibly
2	2.5.5	Why have you chosen particular standards and approaches for metadata and contextual documentation ?	Text	Yes	Decisions relating to metadata standards may be made with recourse to: staff expertise, a preference for Open standards, or widespread usage with a given community. Guidance: - DCC Briefing Paper on	No

					Metadata Standards (http://www.dcc.ac.uk/res ources/briefing-papers/standards-watchpapers/what-are-metadata-standards)	
3	3.1	Ethical and Privacy Issues	Heading	Yes		
3	3.1.1	Are there ethical and privacy issues that may prohibit sharing some or all of the dataset(s)?	Boolean	Yes	Guidance: - UKDA Guidance on Consent, Confidentiality and Ethics (http://www.dataarchive.a c.uk/sharing/confidential. asp)	Possibly (in Lifecycle Archive module)
3	3.1.2	If you answered Yes to DCC 3.1.1, How will these be resolved?	Text	Yes	Ways to resolve these may include: anonymisation of data; referral to departmental or institutional ethics committees; or formal consent agreements. The consent agreements you make with research participants and Data Protection legislation affect how you store data, who can see/use it and how long it is kept. You should show that you're aware of this and have planned accordingly.	Possibly (Lifecycle only)
3	3.1.3	Is the data that you will becapturing/cr eating "personal data" in terms of the Data Protection Act (1998) or equivalent legislation if outside the UK?	Text	Yes	Guidance: - DCC Briefing Paper on Data Protection (http://www.dcc.ac.uk/res ources/briefing-papers/introduction-curation/dataprotection)	No
3	3.1.4	What action will you take to comply with your obligation under the Data Protection Act	Text	Yes		No

		(1998) or equivalent legislation if outside the UK?			
3	3.2	Intellectual Property Rights	Heading	Yes	It is important to strike an appropriate balance between concern for legal implications and getting research done. Inactivity due to legal overwhelm is better avoided!	
3	3.2.1	Will the dataset(s) be covered by copyright or the Database Right? If so give details in DCC 3.2.2, below.	Boolean	Yes	Guidance: - DCC Legal Watch Paper on the Database Right (http://www.dcc.ac.uk/res ources/briefing-papers/legal-watch-papers/iprdatabases)	Possibly
3	3.2.2	If you answered Yes to DCC 3.2.1, Who owns the copyright and other Intellectual Property?	Text	Yes	For multi-partner projects, this may be worth covering in a consortium agreement. Ideally, this should address the risk of movement of staff between institutions mid-project.	Possibly
3	3.2.3	If you answered Yes to DCC 3.2.1, How will the dataset be licensed?	Text	Yes	Any restrictions on use should be justified, and a timeframe for data release outlined to assure the funder of wider public benefit where possible. For example will there be: delays in releasing data while you seek a patent? Planned embargo periods / right of first use to secure publications? Prevention of data sharing due to terms of commercial partnership agreements? Guidance: - DCC Legal Watch Paper on Creative Commons (http://www.dcc.ac.uk/res ources/briefing-papers/legal-watch-papers/creativecommons-licensing) - DCC Legal Watch Paper on Science	Possibly

					Commons (http://www.dcc.ac.uk/resources/briefing-papers/legal-watch-papers/sciencecommons)	
3	3.2.4	For multi-partner projects, what is the dispute resolution process / mechanism for mediation?	Text	Yes	You may wish to cover this in a consortium agreement, in which case you can just answer "As per the consortium agreement."	No
4	4	Access, Data Sharing and Reuse	Heading	Yes	There are often conflicting pressures on researchers to share or withhold their data. Early consideration of the issues can help to resolve these conflicts.	
4	4.1	Access, Data Sharing	Heading	Yes		
4	4.1.1	Are you under obligation or do you have plans to share all or part of the data you create/capture?	Boolean	Yes	Your funding body may insist on data sharing, and - if you are in the UK - your project may be subject to Freedom of Information (FoI) legislation. (Note that FoI legislation differs in Scotland from England and Wales.) Guidance: - UKDA Guidance on Data Sharing (http://www.dataarchive.ac.uk/sharing/whyshare.asp)	No

4	4.1.2	If you answered No to DCC 4.1.1, why will you not share your data?	Text	Yes	You may not plan to share data due to: ethical reasons; non-disclosure agreements; or quality-related issues. (You may also choose to share only part of your dataset(s): if so, give details here.) Guidance: - DCC Legal Watch Paper on Sharing Medical Data (http://www.dcc.ac.uk/resources/briefing-papers/legal-watch-papers/sharingmedical-data)	No
4	4.1.3	If you answered Yes to DCC 4.1.1, How will you make the data available?	Text	Yes	Here you will want to explain how the data will be shared e.g. will they be deposited in a data centre, will you forward copies on request to interested parties, etc. Also consider how potential users will find out about your data, e.g. will you publish details of your research, present at conferences, blog about your findings, promote your research outputs on a website? etc.	No
4	4.1.4	If you answered Yes to DCC 4.1.1, When will you make the data available?	Text	Yes		Possibly (Embargo)
4	4.1.5	If you answered Yes to DCC 4.1.1, What is the process for gaining access to the data?	Text	Yes	Ways of accessing data include: downloading from a data centre; requesting direct from the researcher; downloading from a Web page.	Yes
4	4.1.6	If you answered Yes to DCC 4.1.1, Will access be chargeable?	Boolean	Yes		No
4	4.1.7	If you answered Yes to DCC 4.1.6, Please give	Text	Yes		Yes

		details.				
4	4.2	Exploitation	Heading	Yes	Exploitation of data may comprise using the data in support of academic publications, or for some other kind of gain (e.g. commercial).	Possibly
4	4.2.1	Does the original data collector/ creator/ principal investigator retain the right to use the data before opening it up to wider use	Boolean	Yes		No
4	4.2.2	If you answered Yes to DCC 4.2.1, Please give details.	Text	Yes	All the funders that we've examined permit embargoes, but expect them to be reasonable and expect justification (e.g. for the time limits set).	Yes (Embargo)
4	4.2.3	Are there any embargo periods for political/comm ercial/patent reasons?	Boolean	Yes		No
4	4.2.4	If you answered Yes to DCC 4.2.3, Please give details.	Text	Yes		Yes (Embargo)
4	4.3	Reuse	Heading	Yes		
4	4.3.1	Which groups or organisations are likely to be interested in the data that you will create/capture ?	Text	Yes	There is a push for publicly funded data to be of wide benefit, so it may help to show that you envisage your data being of use beyond your group, or even beyond your discipline	No
4	4.3.2	How do you anticipate your new data being reused?	Text	Yes	Explain how the data will be developed with future users in mind, i.e. are your choices of formats, technologies and metadata appropriate to these audiences?	No

5	5	Short-Term Storage and Data Management	Heading	Yes	You should note what support is provided, e.g. "we will use the University's networked service, which is backed up daily by computing support." Or, if you will mange your own storage and backup, explain how you will do that, noting any agreements you have in place e.g. mirroring data on a second server at the project partner's University. Additionally, more and more researchers keep data on portable devices (laptops, USB sticks, etc). It is crucial that short-term storage policies address and make provision against unintended loss of portable equipment.	No
5	5.1	Storage Media and Data Transfer	Heading	Yes	This section relates primarily to in-project storage, as opposed to longer-term storage/preservation.	
5	5.1.1	Where (physically) will you store the data during the project's lifetime?	Text	Yes	Storing data on laptops alone is very risky: backed-up network drives are far preferable. Guidance: - UKDA Guidance on Data Storage (http://www.dataarchive.ac.uk/sharing/datastorage.asp)	Possibly (Archive Module in Lidecycle)
5	5.1.2	What media will you use for primary storage during the project's lifetime?	Text	Yes		No

5	5.1.3	How will you transfer/transmit the data, if this is required?	Text	Yes	You may need to consider the data transfer speeds supported by your primary storage device, and if possible seek guidance from your institution's computing service on whether the available bandwidth on the local network, and your institution's network infrastructure, will be sufficient to meet your project's needs for short term collaborative working and any Web-based data publication. (You may also want to address encryption if this is appropriate/necessary, and whether it is appropriate to transfer your data across unsecured network connections.	No
5	5.2	Back-up	Heading	Yes		
5	5.2.1	How will you back-up the data during the project's lifetime?	Text	Yes	Remember to consider all of the costs of backup, e.g. logging storage locations, version control, and of recovering data from the backup. These time/staff costs will far exceed the price of the storage device. If these are set against the risks of the device failing, becoming lost, destroyed or unusable, a centralized backup service is more likely to be justifiable. This service may be provided by your institution; you may also choose to incorporate off-site storage for additional protection, or arrange your own backup regime. Guidance: - UKDA Guidance on Data Backup (http://www.dataarchive.ac.uk/sharing/backup.asp)	No

5	5.2.2	How regularly will back-ups be made?	Text	Yes	This may be something you choose to leave to your institutional or departmental support, but it's worth recording the information here	No
5	5.2.3	Who is responsible for backup?	Text	Yes		No
5	5.3	Security	Heading	Yes	Security decisions may be made with a view to your data's financial value and/or its sensitivity.	
5	5.3.1	How will you manage access restrictions and data security during the project's lifetime?	Text	Yes	This may be managed via various levels of password protection. Guidance: - DCC Briefing Paper on Information Security Management (http://www.dcc.ac.uk/resources/briefing-papers/standards-watchpapers/information-security-management-iso-27000-iso-27k-s) - UKDA Guidance on Data Security (http://www.dataarchive.ac.uk/sharing/security.asp)	Possibly
5	5.3.2	How will you implement permissions, restrictions and/or embargoes?	Text	Yes	You may wish to give details of any policies in place governing making copies of data.	Yes (Embargo)
5	5.3.3	Give details of any other security issues.	Text	Yes		No
6	6	Deposit and Long-Term Preservation	Heading	No	Section 6 is about long-term preservation. Many researchers will not perform these tasks themselves, so data centre staff or other long-term stewards may be best placed to answer these questions. Guidance: - DCC Briefing Paper on Digital Repositories (http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/digitalrepositorie	

					s)	
6	6.1	What is the long-term strategy for maintaining, curating and archiving the data?	Text	No	Here you will want to demonstrate consultation between data creators and the relevant repositories / data centres to secure an appropriate place of deposit. Give details on the rationale for choosing this particular place of deposit. (N.B. Funders may require data to be offered to a particular data centre on completion of the project.) If there isn't anywhere you can deposit, explain how you will address sustainability e.g. by choosing open standards, or note how your institution can support you to store and manage the data in the longer term. Remember that you can consult institutional archivist(s) and records managers in formulating long-term retention plans. Guidance: - DCC Briefing Paper on Digital Preservation (http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/digitalpreservation) - JISC Briefing Paper on Digital Preservation (http://www.jisc.ac.uk/media/documents/publications/digitalpreservationbp.pdf)	Yes (Lifecycle archive Module)
6	6.2	Long-Term Specifics	Heading	No	This section addresses three key issues: Selection, Retention, and Transformation.	

6	6.2.1	Will or should data be kept beyond the life of the project?	Boolean	No		No
6	6.2.2	If you answered Yes to DCC 6.2.1, How long will or should data be kept beyond the life of the project?	Text	No	Your funding body or institution may specify time-spans for retention. If not, general guidance is given in the RCUK Code of Good Research Conduct which says that "data should normally be preserved and accessible for ten years, but for projects of clinical or major social, environmental or heritage importance, for 20 years or longer."	Possibly
6	6.2.3	If you answered Yes to DCC 6.2.1, What data centre/ repository/ archive have you identified as the long-term place of deposit?	Text	No	Your funder may have a preferred place of deposit.	Possibly
6	6.2.4	What data will be preserved for the long-term?	Text	No	You may wish to preserve all, none, or a selection of data over the long-term. You should also indicate here whether you will preserve raw data, derived data, samples, etc.	No
6	6.2.5	On what basis will data be selected for long-term preservation?	Text	No	You may wish to include timeframes here as well. Guidance: - DCC Briefing Paper on Appraisal and Selection (http://www.dcc.ac.uk/res ources/briefing-papers/introductioncuratio n/appraisal-and-selection)	No
6	6.2.6	If the dataset includes sensitive data, how will you manage this over the longer term?	Text	No	This should include a justification of decisions and should cover deletion of data if appropriate.	No

6	6.2.7	Will transformations be necessary to prepare data for preservation and/or data sharing?	Boolean	No	Examples of transformation may include data cleaning/anonymisation where appropriate, or migration to another file format.	No
6	6.2.8	If you answered Yes to DCC 6.2.7, what transformations will be necessary to prepare data for preservation / future re-use?	Text	No	Examples of transformation may include data cleaning/anonymisation where appropriate, or migration to another file format.	No
6	6.3	Metadata and Documentation for Long-Term Preservation	Heading	No	If you are a researcher submitting your data to a data centre or repository, the earlier you consider their metadata and documentation requirements the less painful it will be to provide the essential details, the better the chances of your data being found and re-used, and therefore the higher the chance of it having a lasting impact. Here you will want to show that you are aware of data centre standards for deposit, and have reflected these in your data development plans. You may wish to include (e.g.) references, reports, research papers, fonts, the original bid proposal, etc. You may also wish to include contextual/ related/ representation information	

6	6.3.1	What metadata/ documentation will be submitted alongside the datasets or created on deposit/ transformation in order to make the data reusable?	Text	No	Digital files are fundamentally strings of binary digits (bits). In order to process them, one must know the format they are in and what software is needed to read that format. Even after the file has been successfully opened, extra information may be needed in order to fully understand the contents. In the terms of the Open Archival Information System (OAIS) Reference Model, the information required to transform a stream of bits into something intelligible is called representation information. Guidance: - DCC Glossary Definition of Representation Information (http://www.dcc.ac.uk/digi tal-curation/glossary)	Possibly
6	6.3.2	How will this metadata/doc umentation be created, and by whom?	Text	No	The AHDS Catalogue Form is used to produce a full catalogue record for online catalogues. Guidance: - AHDS Catalogue Form (http://www.ahds.ac.uk/de positing/catalogueform.ht m)	Possibly
6	6.3.3	Will you include links to published materials and/or outcomes?	Boolean	No		No
6	6.3.4	If you answered Yes to DCC 6.3.3, please give details.	Text	No		Possibly

6	6.3.5	How will you address the issue of persistent citation?	Text	No	You may wish to refer to Digital Object Identifiers (DOIs), Persistent URLs, etc. Guidance: - DCC Briefing Paper on Persistent Identifiers (http://www.dcc.ac.uk/resources/briefing-papers/introductioncuration/persistent-identifiers) - The Digital Object Identifier System (http://www.doi.org/)	Possibly
6	6.4	Longer-Term Stewardship	Heading	No		
6	6.4.1	Who will have responsibility over time for decisions about the data once the original personnel have gone?	Text	No	This is likely to be either an institutional library or repository, or some other data custodian (e.g. a data centre.)	Yes (Archive Module)
6	6.4.2	In the event of the long-term place of deposit closing, what is the formal process for transferring responsibility for the data?	Text	No	This should be completed by a representative of the original place of deposit	Possibly
7	7	Resourcing	Heading	Yes	It is important that data management is treated as a first-class research activity, with appropriate funds and effort allocated to it.	
7	7.1	Outline the staff/organisational roles and responsibilities for implementing this data management plan.	Text	Yes	This could include: data management time allocations; project management of technical aspects; training requirements; storage and backup; contributions of non-project staff, etc. Individuals should be named where possible. Continue in an Annex if necessary.	Yes (Organization Scheme with Roles specified – Lifecycle only)

7	7.2	How will data management activities be funded during the project's lifetime?	Text	Yes	This should cover (e.g.) payments to service providers within institutions, payments to external data centres for hosting data, income derived from licensing data, etc). It is also important to remember to build costs of inproject data management into the project budget. (N.B. Some funders state explicitly that they will meet the cost of preparing data for deposit, so remember to include this in your time and budget allocation too!)	No
7	7.3		Text	Yes	This should cover (e.g.) payments to service providers within institutions, payments to external data centres for hosting data, income derived from licensing data, etc). It is also important to remember to build costs of inproject data management into the project budget.	No
8	8	Adherence and Review	Heading	Yes	A data management plan can only be effective if everyone agrees to adhere to it. Communication is also important: the human aspects of data management are widely held to be more difficult than the technical aspects. All aspects of the environment will change over time, so it is similarly important that you keep your plans up-to-date via regularly scheduled review	
8	8.1	Adherence	Heading	Yes		
8	8.1.1	How will adherence to this data management plan be	Text	Yes		No

		checked or demonstrated?				
8	8.1.2	Who will check this adherence?	Text	Yes		Yes (Organizatio nScheme with Roles – Lifecycle only)
8	8.2	Review	Heading	Yes	A data management plan should be a living document, and it is important - especially for longer term work - that it is reviewed on an ongoing basis.	
8	8.2.1	When will this data management plan be reviewed?	Text	Yes		Possibly (as Lifecycle Event – Lifecycle only)
8	8.2.2	Who will carry out reviews?	Text	Yes		Yes (Organizatio nScheme with Roles – Lifecycle only)
8	8.2.3	Does this version of the DMP supersede an earlier plan?	Boolean	Yes		No
8	8.2.4	If you answered Yes to DCC 8.2.3, you may wish to enter information about the relationship between versions here.	Text	Yes	You may want to consider including previous versions of the plan as Annexes to this version.	No
9	9	Statement of Agreement	Heading	No	You may wish to formalise your Data Management Plan with a statement of agreement, including signatories if relevant.	
9	9.1	Statement of Agreement	Text	No		No
10	10	Annexes	Heading	Yes		
10	10.1	Contact details and expertise of nominated	Text	Yes		Yes (Organizatio nScheme with Roles –

		data managers/named individuals				Lifecycle only)
10	10.2	Glossary of terms	Text	No		Yes - as Other Material
10	10.3	Other annexes as required	Text	No	You may wish to attach other annexes to your data management plan. You can use this space to list these annexes.	Yes - as Other Material