RDA Data Foundation and Terminology
DFT 4: Use Cases

Contributors: Peter Wittenburg, Gary Berg-Cross, Hans Pfeiffenberger, Reagan Moore

Technical Editors: Gary Berg-Cross, Peter Wittenburg
Copy Editor Karen Green (UNC)

December 2014

Version 1.6

The general outline of documents from DFT WG is as follows:

- Data Models 1: Overview
- Data Models 2: Analysis & Synthesis
- Data Models 3: Term Snapshot
- **Data Models 4: Use Cases**
- DFT 5: Term Tool Description

1. Introduction

Major products of the DFT effort include a summary, analysis & synthesis of a large body of definition work on data management terms models & associated terms. The intent has been to develop a vocabulary and an associated model to promote common understanding of data organizations and data sharing. As work has progressed some use cases have been developed to illustrate requirements and validate relationships across models. Use cases are often used in engineering efforts to capture the requirements of a system and act as a means of discussing these with stakeholders. Such scenarios of use provide a context for otherwise abstract concepts and can thus be used to check the fitness of portions of the synthesis and its terms in specific relationships details by scenarios.

Scenarios may also provide test situations for subsequent adoption pilots that explore the usefulness of the DFT model for specific data sharing project activities portrayed in the use case scenarios.

The Use Cases

Several Use Case scenarios were developed by the community and discussed at Plenaries as examples of relevant work using pertinent concepts such as Research Data Object. In addition the model overview document describes some application scenarios.

Use Cases and Associated Vocabulary

In the sub-sections below the use case scenarios graphics are presented along with additional textual propositions that assert what we should capture in our definitions or issues. On the whole these use cases focus on repository operations such as registration of PID as supported by PID systems. Because of

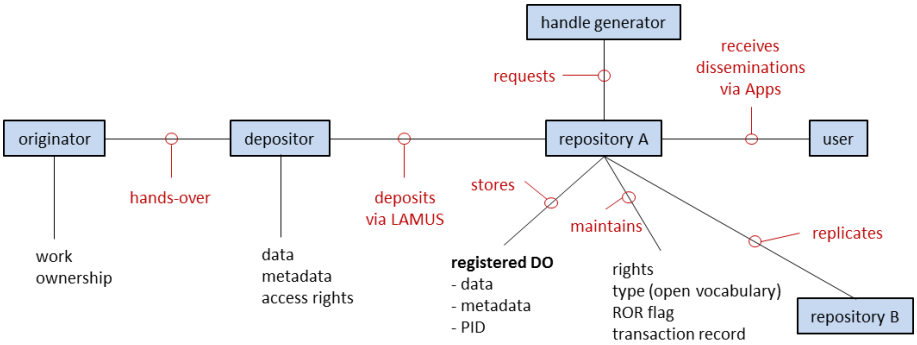
early RDA WG focus PIDs are singled out from other types of metadata and curations which is not detailed beyond noting some simple relationships. This is evident in the first Figure later in this section. These relations can be expected to detailed is subsequent work.

1. Scenarios in the realm of EUDAT.

1.1 Community Data Organization

Duplication is an important activity with the CLARIN project with its Language Data Infrastructure a portion of which featuring data being replicated from one repository to another is shown below. CLARIN as well other community research infrastructures such as ENES (Climate Modeling), VPH (Virtual Physiology of Humans), EPOS (European Plate Observing System), LifeWatch (biodiversity), EUON (European Ontology Network), BBMRI (biobanking and biomolecular resources) and DRIHM (hydrology) use the EUDAT data infrastructure for data storing, replication, management/curation and computational purposes. EUDAT is currently offering 5 services to communities: (1) B2DROP which is a Dropbox like synchronization service, (2) B2SHARE which is a deposit service for long-tail data, (3) B2SAFE which is replication service for large data sets, (4) B2STAGE which is a service that allows to do computations on uploaded data sets and (5) B2FIND to do metadata searches on all data objects stored in the EUDAT domain. Except for B2DROP which is a temporary store all objects that are uploaded to the EUDAT domain or that are created within the EUDAT domain, PIDs are assigned including state information as well as metadata is being created to support discovery and curation functions.

In the following diagram a typical data organization within the CLARIN research infrastructure community (language data & technology) is being shown.



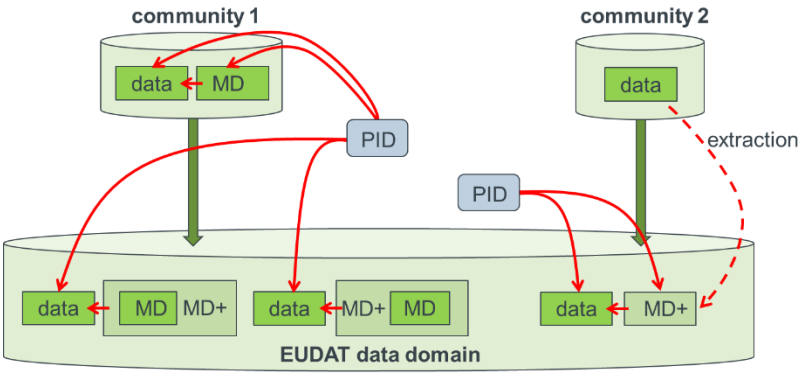
As shown, pre-processed data is deposited in a trusted (certified according to Data Seal of Approval) repository. During the upload Digital Objects (DO) are registered and given a PID along with data properties that are stored along with the PID record and the metadata description. The core data model within CLARIN is thus fully compliant with the DFT core model and during repository certification adherence to this model is being checked. The CLARIN community has also been accepted as WDS network partner due to its clear requirements for proper data organizations.

The data organizations of other communities differ slightly as can be seen from the model overview document or in some cases as in VPH traditional ways are being used, i.e. some metadata is covered in a relational database and files are stored in file systems. Thus, several communities have adopted the basic RDA DFT data model, i.e. their data objects are assigned with PIDs including state information and are described with metadata.

Summary: we can conclude that several of the communities engaged with EUDAT already adopted the basic data model as described by DFT results.

1.2 EUDAT Replication Service

For the replication step (B2SAFE) in the EUDAT domain we can distinguish two scenarios:
(1) If a bit sequence of a DO is being replicated from an original repository following proper data organization principles to another repository, the original PID record is then extended to cover the path (URL) to access the new, duplicated instances1. A schematic diagram which highlights PIDs and metadata involved in the replication act is shown below. The point made is that the PID assigned to a DO is being stored in the PID registry to maintain stable references, but it is extended by the new paths, so that the user or a software module can chose which of the instances of the bit sequence is going to be accessed. During the replication process checksum information as a property of the bit sequence and being stored in the PID record is used to verify the correctness of the replication step, i.e. it is verified whether the resulting bit sequence is exactly the same as the original one. Only in case of a successful replication step the PID record is being updated by the new paths information. At the replication site the metadata description is being updated to include some system information resulting in enhanced metadata or a MD+. However, the original metadata information being referred to by the original PID describing many of the properties of the DO are not being changed of course.
(2) In the case that a community follows a traditional way of data organization such as indicated for VPH for example, EUDAT will request a new PID record and create a new metadata description. When scripts can be built that extract some metadata from available information the metadata description will be richer, in the other case it will simply contain properties that can be generated during the replication step (size, type, date, etc.). The PID record will be created during replication and thus point to the bit sequence, the metadata description and contain additional properties such as checksum.



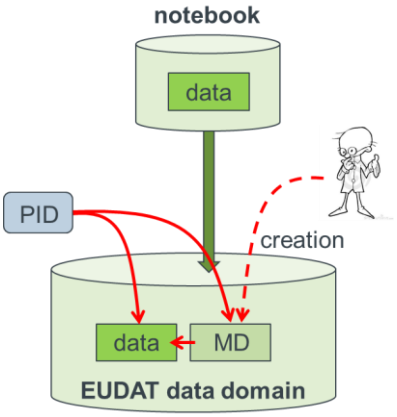
It has been part of the EUDAT experience that "full" replication, i.e. using existing identifiers and metadata descriptions is very expensive due to the large heterogeneity of data organizations. In the first case it is impossible to let machines automatically find all necessary information as indicated above. And in the second case scripts need to be developed and maintained to do the appropriate extraction. To reduce the enormous effort and thus come to scalable solutions EUDAT now offers a default B2SAFE services which simply ignores all existing PID and MD information (as indicated in case two without MD extraction). If this information needs to be replicated as well and if all relationship information needs to

1 Additional IDs are being generated at the hosting sites for management reasons, but they should not be exhibited as THE PIDs to access a given bit sequence. They do evidence that fact that a DO may have more than one ID.

be preserved collaboration projects need to be carried out that involve experts from both sides to resolve issues. While disappointing this is not surprising since agreement on things like metadata involves issues of meaning that are challenging to automate.

Summary: In the EUDAT replication services the basic DFT data model is being applied already now which is due to a fruitful bi-directional interaction between the DFT and the EUDAT work.

1.3 EUDAT Deposit Service

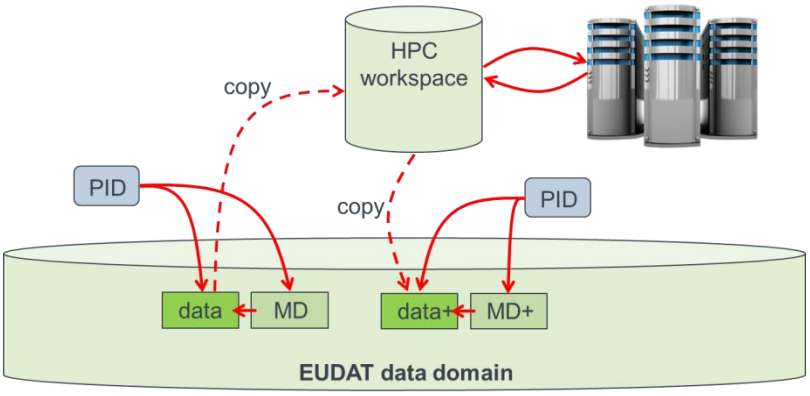


It can be added here that the B2SHARE service for typical long-tail data follows almost the same principles as the case 2 with one exception: at the upload step the user needs to provide useful metadata about the objects to be stored in the EUDAT domain. During the upload process a PID including typical information types such as checksum and pointers to the locations of the bit sequence and the metadata description is being registered and a metadata description is being generated. Implied in all of this is the existence of suitable registries and repositories for storing the needed information.

Summary: Also for its deposit services EUDAT adopted the basic DFT data model.

1.4 EUDAT Computation Service

The B2STAGE service from EUDAT copies selected digital objects into the HPC workspace which will lead to some computational activity and in general to the creation of derived data objects as shown in the diagram below. These are then copied from the HPC workspace into the EUDAT data domain. B2STAGE also makes use of the DFT data model in so far as PIDs are assigned to the new digital objects and also metadata is being created by making use of the "old" metadata description and by adding information about the computation that had been carried out (this would include provenance information).



Overall Summary: We can observe that RDA discussions had a huge influence on the way EUDAT has defined its internal data organizations and how the services were designed. Also due to the close interaction with a variety (about 15) of research infrastructures from various disciplines it had an enormous impact changing community their ideas about data organizations. Of course there was also a high impact of the concrete infrastructure work from the EUDAT experience and that of closely

collaborating communities involving discussions within DFT. In many cases the data organizations are highly complex, since a lot of relationships need to be maintained. In general these are stored in metadata descriptions making it so important to standardize policy for all steps needed to maintain and update metadata.

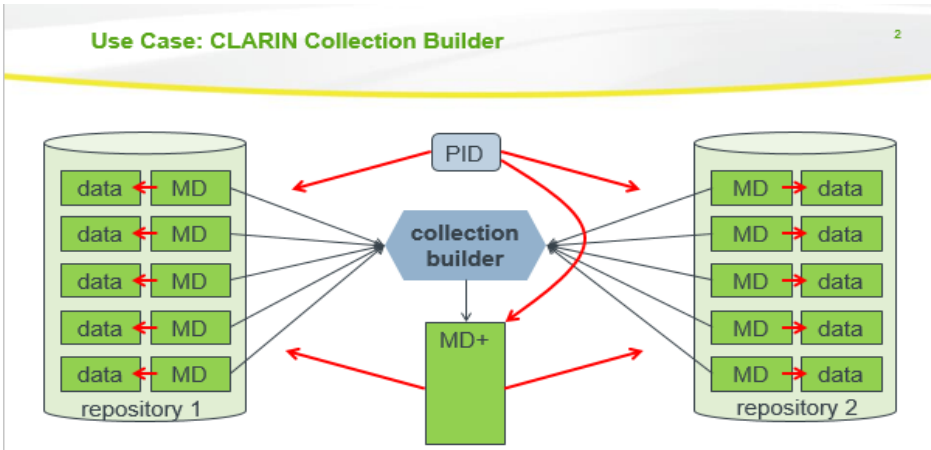
It should be noted here, for example, that the replication work in EUDAT is guided by practical policies which include a lot of detailed steps. We did not further describe details here, but refer to use case 6 which gives a more detailed view on what kind of steps practical policies must include.

2. CLARIN related work

2.1 Collection Building

Collection building is an important activity within the CLARIN project and it also utilizes the data-metadata connection as discussed previously for replication along with a PID.

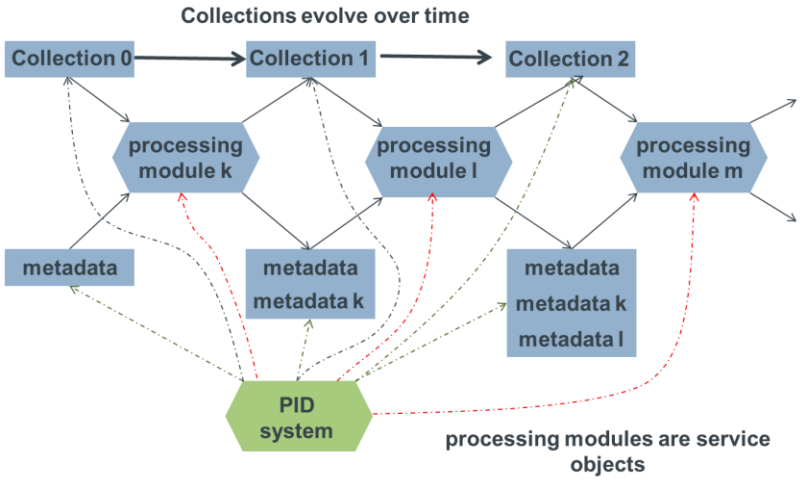
As shown in the diagram below data and associated metadata is stored in repositories at different locations. New (virtual) collections (meaningful aggregations) are built from this repository data. These collections are assigned a PID and have additional metadata describing the properties of the collection. The collection metadata allows one to know what data it was aggregated from by means of PIDs pointing to its components. As noted a collection has accessible information via a PID record that contains useful state information and is citable (has citation information). Collection building results in complex relationships between the various digital objects. Some are included in the original metadata descriptions and others are being added as a result of the collection building process.



2.2 Workflow Work

Collections can also be the source of new or extended collections via collection processing which is shown in the figure below which was created for the CLARIN language processing flows (tokenizing, part-of-speech tagging, syntax parsing, etc.). A PID system can provide identification for the series of related collections (including their components) and for the processing modules allowing altogether reproducibility (assuming that the code still can be executed). As is indicated schematically after each

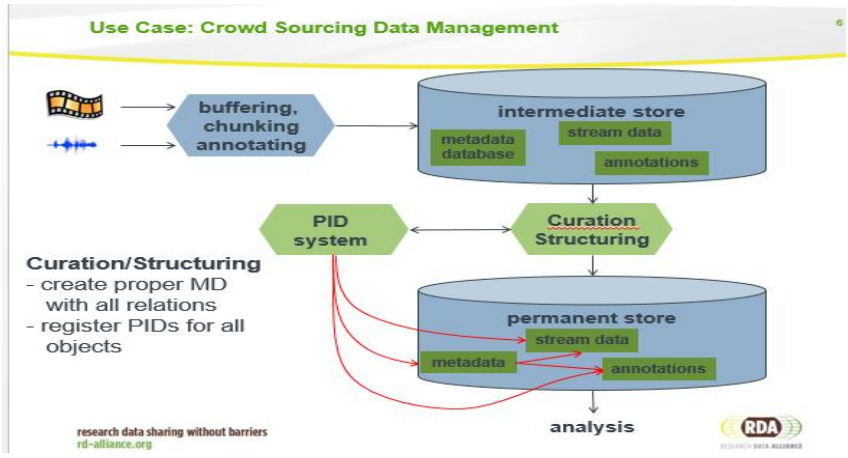
processing step the metadata will be enriched by provenance information which will allow tracing history.



Along the path of collection building the type of data may change since data derived from other data by analytical processes may be generated. Also here complex relationships need to be maintained within the PID records and metadata descriptions. In the diagram only those are indicated that emerge during the workflow processing steps.

Summary: Needless to say that such collection building and processing amounts in utterly complex relationship structures which need to be maintained. Since the CLARIN research infrastructure strictly requires one to follow the model as being described by DFT, we can call CLARIN an early DFT adaptor.

2.3 Managing Crowd Sourcing Data



Crowd sourced data is being generated at unpredictable moments by a huge variety of persons and thus will be stored temporarily in specialized databases after some pre-processing (buffered, annotated, reduction, etc.) as shown in the figure above. For a research infrastructure like the CLARIN initiative in the domain of language resources and technology, annotation information includes aspects such as detecting and tagging voiced segments, detecting who is speaking in a segment as well of some rough classification of segments. To come to referable digital objects subsequent processing needs to curate

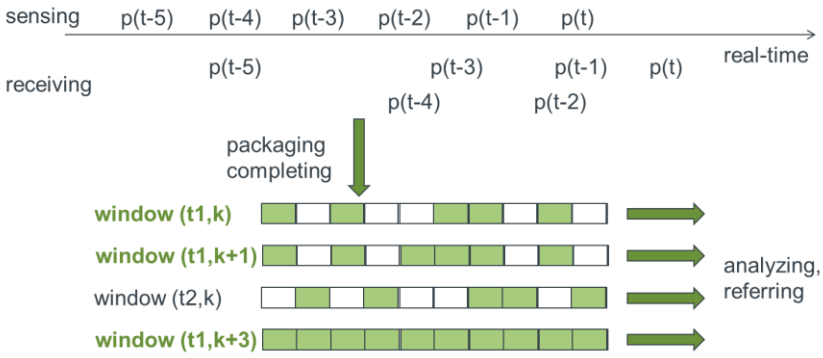
and structure the collected crowd sourcing data including the annotations by assigning PIDs and proper metadata at specific time steps. It should be noted that once these digital objects are released the scientific analysis and annotation work starts, i.e. the early annotation processes as indicated in the diagram can be viewed as including automatic pre-processing steps that have the task of reducing the amount of data by cutting non-attractive segments and roughly classify segments to simplify searches for example.

Also in this case of course complex relationships between the different digital objects need to be stored and maintained which are to a large extent contained or should be contained in metadata. The figure above shows a design which has not yet been fully implemented.

Summary: In designs of such advanced systems anticipating future applications, the discussions within DFT had an impact on how to set up the backend system with the purpose of making crowd sourced data a part of the open domain of accessible data.

3. Gappy and Dynamic Data

At any one moment data being collected in real time from sensors or crowds may be incomplete, but data received may already have been processed for fast decision taking (for example calculating early alarms in case of volcano eruptions) and stored in a repository assigned with a PID and metadata for referencing purposes. Such gappy data may have gaps that may be filled in later at unpredictable times as data arrives after varying delays as shown in the figure below as part of a dynamic data process. As implied in the diagram transmitted data packages include timestamp information which allows proper assembly of the whole stream once received at the collecting station. At a certain moment in time all packages will have arrived and the received data can be viewed as being in final form.



There are two problems now that need to be sorted out to fit with DFT's data model:

- Such sensor data streams are of quasi infinite duration. How to chunk such information to create digital objects and assign a PID including, for example, checksum information to verify identity and integration?
- Such data streams change their nature while they are already being processed and decisions are being taken. How to refer back to from documented decisions to a specific state before completion?

Interested communities such as EPOS, COOPEUS, etc. using such sensors for various purposes are working on these questions and in RDA the Data Citation working group is discussing these aspects to come to recommendations.

Summary: The DFT core model had a strong influence on the discussions within the above mentioned communities/projects in so far as special workshops were devoted to the topic. A wide agreement could be seen that PIDs and metadata descriptions need to be created along with the chunked streams becoming then Digital Objects. However, a couple of unsolved issues need to be solved out due to the dynamic nature of the data objects and thus of associated IDs. The discussions are not finalized and are not easy to take, since large existing software packages which are being used would have to be changed or replaced.

4. Relationships in Research Data Objects (RDOs)

This use case features grow out of the "Earth System Science Data" (ESSD) Journal experience and has – compared to the previous use cases is very complex because of the addition of a new element linking. This features "research data objects"2 that are complex, compound and/or networked objects with many relations. Such compound objects are featured in the OAI-ORE model which is included in the DFT Model Overview paper. Compound (data) information objects are aggregations of distinct information units whose combined/related form some human recognize as a logical and/or thematic whole. One manifestation of these are digitized books that have an aggregation of chapters, where each chapter is an aggregation of scanned pages; or a music album that is the aggregation of several audio track elements. Another example is an image object that is the aggregation of image parts. A continental image made of images shot from a satellite would be one example. Imaging and e-journal projects often differentiate between their well-managed (and described) "master" files and the derived versions (thumbnails, JPEG files, PDFs) made available through the Web. These are examples of a network of related DOs from an original or master form to derived forms.

	A	B	C	D	E	F	G
1		**Terrestrial CO_2 sink (positive values represent a flux from the atmosphere to the land)**					
2		All values in petagrams of carbon per year (PgC/yr), for the globe. For values in carbon dioxide (CO_2), multi					
3		1PgC = 1 petagram of carbon = 1 billion tonnes C = 1 gigatonne C = 3.67 billion tonnes of CO_2					
4		**Cite as:**					
5		CLM4CN	Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence,				
6		HYLAND	Levy, P. E., M. G. R. Cannell, et al. (2004). "Modelling the impact of future changes in clim				
7		LPJ-GUESS	Smith, B., I. C. Prentice, et al. (2001). "Representation of vegetation dynamics in the mod				
8		LPJ	Sitch, S., B. Smith, et al. (2003). "Evaluation of ecosystem dynamics, plant geography and				
9		O-CN	Zaehle, S., P. Ciais, et al. (2011). "Carbon benefits of anthropogenic reactive nitrogen offs				
10		ORCHIDEE	Krinner, G., N. Viovy, et al. (2005). "A dynamic global vegetation model for studies of the				
11		SDGVM	Woodward, F. I. and M. R. Lomas (2004). "Vegetation dynamics - simulating responses to				
12		JULES	Clark, D. B., L. M. Mercado, et al. (2011). "The Joint UK Land Environment Simulator (JULE				
13		VEGAS	Zeng, N., A. Mariotti, et al. (2005). "Terrestrial mechanisms of interannual CO2 variability.				
14							
15		Terrestrial CO2 sink as a residual		**Models**			
16	Year	of the global carbon budget		CLM4CN	HYLAND	LPJ-GUESS	LPJ
17	1959	0,42		0,79	2,02	0,42	-0,83
18	1960	1,14		0,75	1,53	1,16	0,81
19	1961	1,20		0,30	1,71	-0,07	-0,55
20	1962	1,76		0,79	2,37	1,25	0,57
21	1963	1,72		-1,20	1,81	0,26	-0,37

2 The term "research data object" is being used here in this context and has an important role in this use case. It was considered as being one of the core terms in DFT but there are still ongoing disputes of what "research data objects" exactly are. Fedora objects might be considered examples of an RDA but are nothing more than a special form of collections as described by one of the core terms and as discussed in the analysis document.

More recent network DO examples are research/ scholarly publications that relate/aggregate text and supporting materials including the data cited, software tools, and reference material.

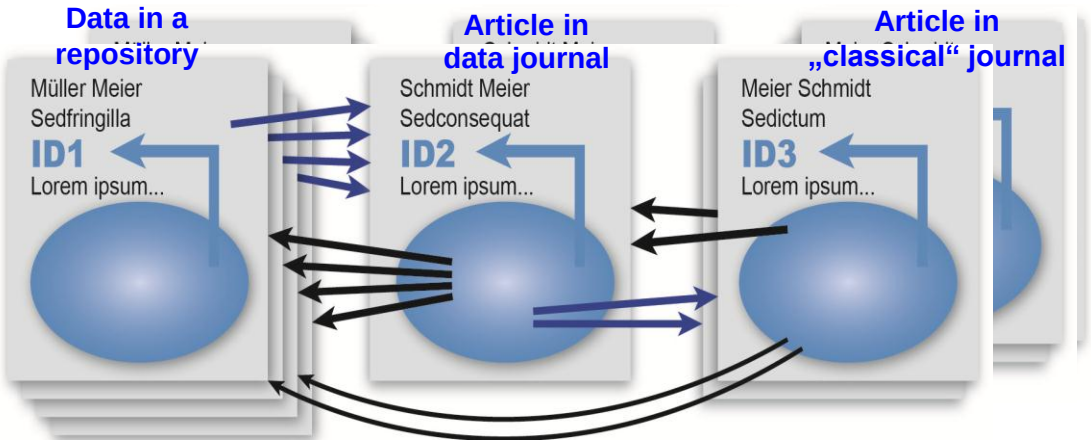
The Earth System Science Data (ESSD) journal maintains links between data/data sets and journal articles. Articles published by such journals can be identified by a single ID (e.g. DOI), but each article can point to several data files aka a data set. Data can be in a simple and relatively undocumented form (e.g. Excel) such as shown above in the graphic.

But data also can come in complex forms such as networks of sets managed by one or more repositories with linked metadata. As shown in the figure below articles in the digital journal, like ESSD, can be linked to-from classical publications or from a citation. Data is linked to-from an article in a data journal. Articles can include updates which improve on submitted drafts to publications to later addendums based on additional data or changes to the cited data etc.

These reflect a number of different relationships such as between:

- Collections and sub-collections
- Collections and objects
- Objects and components (complex hierarchical objects) and
- Other related resources internal or external to a particular complex object from the Digital Asset Management System

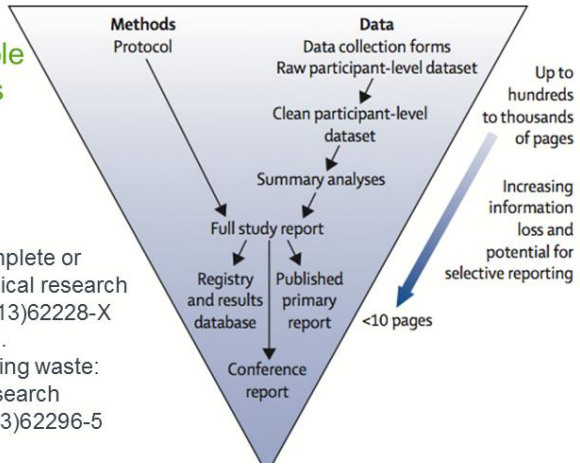
Linking Text and Data



One implication is that some versioning of the complex research data object is required and this versioning must preserve the links to the other portions of the research data object.

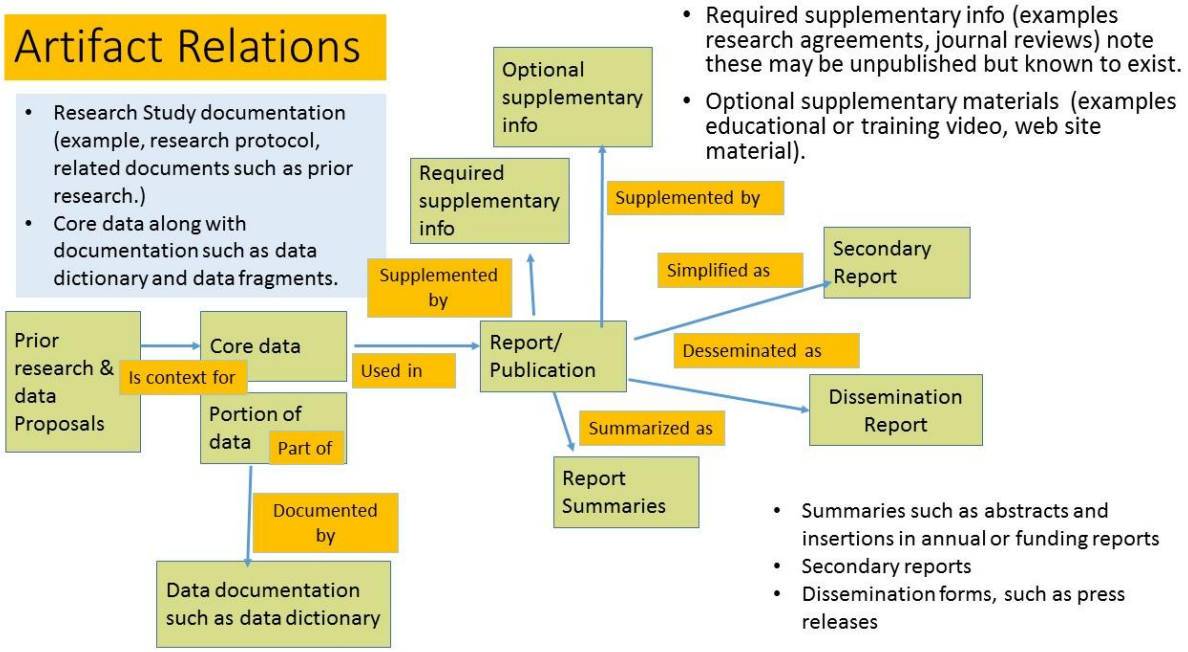
A very complex model using a research example from a January 2014 Lancet publication is shown below with 9 reliable and stable, bidirectional relationships between research elements such as protocols used, raw and processed data, summary analysis etc.:

- The Lancet, Jan 2014
- „9. Reliable and stable bidirectional linkages between all these elements“
- „9. …“: Paul Glasziou et al. Reducing waste from incomplete or unusable reports of biomedical research DOI:10.1016/ S0140-6736(13)62228-X
- Picture: An-Wen Chan et al. Increasing value and reducing waste: addressing inaccessible research DOI:10.1016/S0140-6736(13)62296-5



One may think of the ensemble of networked information in what OAIS terms an archival package which would include Packaging Information (PI) that binds the network of digital object and its associated metadata into an identifiable unit or package (i.e., an OAIS Archival Information Package). The graphic below is a Draft Model of Research Object and includes examples of relations such as might make up an archived research object package.

Some aspects of these activities can also be seen in a Use Case story generated by Benjamin Armintor [1] to illustrate metadata services within a Fedora object model. The Use Case statement is available online [2] and illustrates the selection of particular data objects, in this case a Fedora object that: “manages an image and a numerical dataset for the object. The object also has something analogous to an external data stream representing the current state of metadata description of the object; this is maintained in an external system.”



In this case the numerical data is about spectrometer measurements. A researcher “goes through the collection by date and timestamps, selecting those objects generated by the same experiment, and assigns to all of them an investigation identifier, a project identifier, and her name as creator, thus updating the pre-existing metadata record. After enriching the whole collection in that way, the researcher searches for absorption spectra with certain characteristics (wavelength and curvature of peaks), and tags all objects in the result set as “good measurement”. A result set may contain dozens to thousands of objects. “

This use case illustrates some of the typical metadata used to document and preserve a newly created research collections with the addition of new metadata such as the collector’s name & ID, tags characterizing the data, project information etc. All of this in preparation of the new objects being archived. Part of this seems essential for good collection documentation.

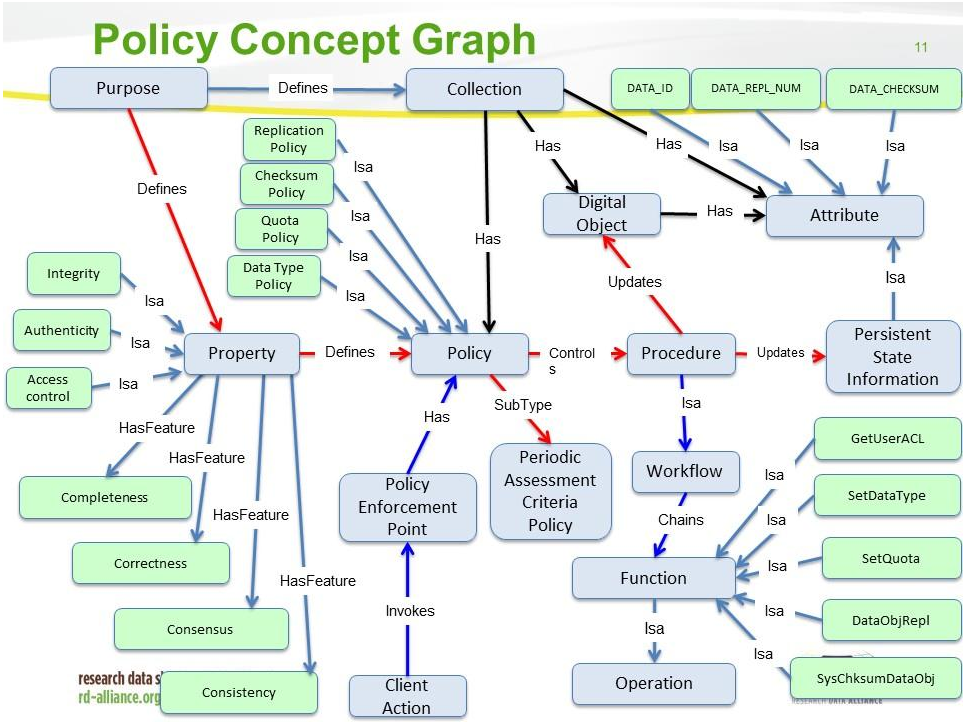
Summary: It is not fully obvious in how far the model presented here is fully compliant with the DFT model. (1) As indicated the term "Research Data Object" was not considered to be taken up in the core set of terms although it is an important one in this use case due to the varying opinions. (2) This use case makes a difference between collections and compounds while the DFT model just views compounds as a specific type of collection. However, the term "compounds" here may fit with the term "aggregation" being used in the DFT term set. (3) It mentions the use of DOIs (as a specific implementation of PIDs) to identify the collection being aggregated in a Fedora object, but does not explicitly state whether the objects (data, metadata) contained in the Fedora object are also data objects in the DFT sense. So again here the term "aggregation" might be more appropriate to describe the use case.

5. Practical Policy use case

The RDA Practical Policy (PP) WG has ambitiously covered a wide spectrum of data management activities. As noted in other DFT products there are policy issues inherent in some of the models covered. For example, the phenomena of gappy data produces what was called mutable data objects as initial data is used for analysis and later updated as data gaps are filled. It is a policy issue whether the same ID is used for the initial and updated version of data and as indicated the Citation Data group is working on harmonizing these aspects. Similar issues may apply to activities like data synchronization and replication and in the EUDAT cases as mentioned above practical policies are guiding the processing. Many policy issues are involved with proper development and management of metadata and relations between metadata and data and metadata components such as checksums, names and data typing.

Some notion of the Policy space and it relation to DFT concepts like data objects can be seen in the collaborative graphic created by Reagan Moore (PP) and Gary Berg-Cross3. As shown there replication and data type policy for data objects and collections are examples of policies which are defined with properties for integrity, authenticity and access control.

3 It should be noted here that the diagram presents work in progress and was done early on before the snapshot terms were decided on, or the PP work finished. Thus some terms and relationships are placeholders and not yet harmonized with the DFT snapshot which is something to be done while PP group is advancing.



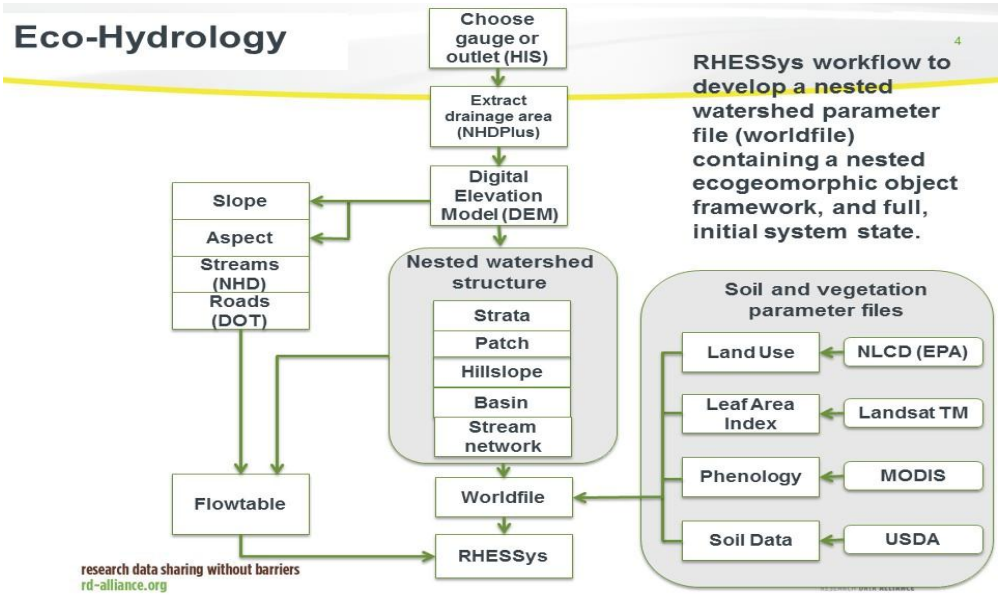
As modeled policy defines procedures for such things as data replication and calculating checksums, concepts discussed in other use cases. There are policies for handling DO attributes such as data replication number and data checksum. At a lower level supporting functions are listed. Appropriate terms for procedures, some functions and some operations are thus suggested for DFT vocabulary and these involve some data concepts like collections and metadata.

To further illustrate operations a Practical Policy use case was drawn from a situation of a hydrologist working on regional-hydro (RHE)- eco simulations requiring data discovery, access, analysis and management in the larger context of reproducible data-driven research. The Hydrologist needs to do 4 broad activates some with sub-parts:

1. Acquire various data sets needed for research
    a. Pick the location of a stream gauge and a date
    b. Access USGS data sets to determine the watershed that surrounds the stream gauge
    c. Access USDA for soils data for the watershed
    d. Access NASA for LandSat data
    e. Access NOAA for precipitation data
    f. Access USDOT for roads and dams
    g. Project each data set to a region of interest
    h. Generate the appropriate environment variables
2. Execute a watershed analysis/simulation
3. Save the research results
    a. create a data object and/or workflow4
    b. Store the workflow, the input files, and the results
4. Enable another hydrologist to re-execute the analysis by detailing the above

4 Workflow is an important concept in these scenarios and is considered a stored or generatable data object itself

The graphic below shows some of the acquisition of data objects with information on its data type etc. from a variety of sources HIS, NHDPlus, DEM). It also shows creation of a wordfile5 to allow the execution of subsequent RHESSys simulations.



There is a corresponding set of Data Registry & Repository management operation since a collection of data is being assemble for analysis but also being made usable by others. A variety of accesses of a registry are likely to

- find an operation that can parse the data type
- find an operation that can create the desired data product
- determine whether the output from the data parsing operation can be used as input to the data product operation, and introduce transformation operations as needed

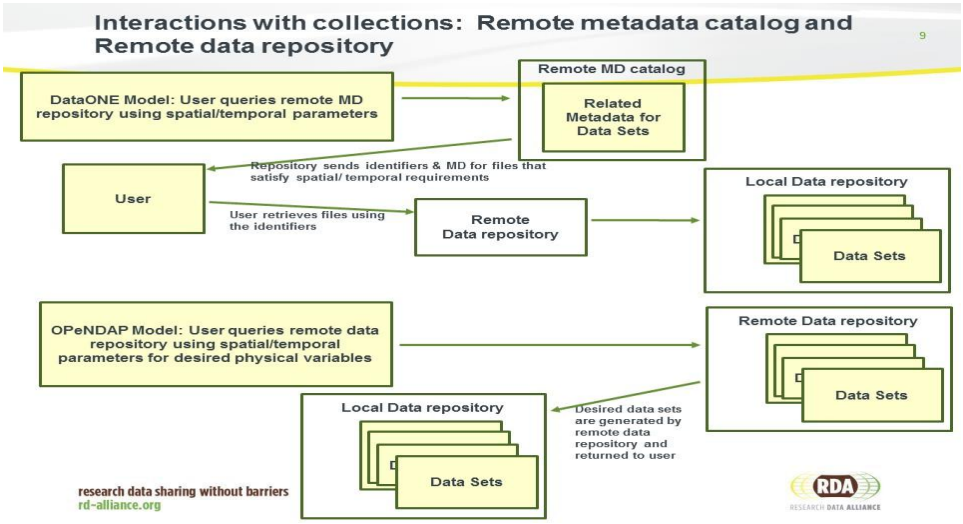
Repository operations which like the registry operations suggest some terms like" indexing" include:

1. .Authenticate the user
2. Authorize the deposition
3. Add a retention period
4. Extract descriptive metadata
5. Record provenance information
6. Log the event
7. Create derived data products (image thumbnails)
8. Add access controls (collection sticky bits)
9. Verify checksum
10. Version data
11. Replicate data
12. Index data
13. Choose a storage location

5 Workflows are detailed in wordfiles within Reagan's approach and thus used as an example here.

14. Choose the physical path name
15. Chain the operations together as a workflow and execute the workflow to create the desired data product.

Finally a graphical use case is provided to illustrate both local and remote approaches to retrieve appropriate data. As shown below DataOne queries a remote MD catalog with spatio-temproal parameters to retrieve appropriate IDs. These are in turn used to query remote data repository which sends data to local data directories.



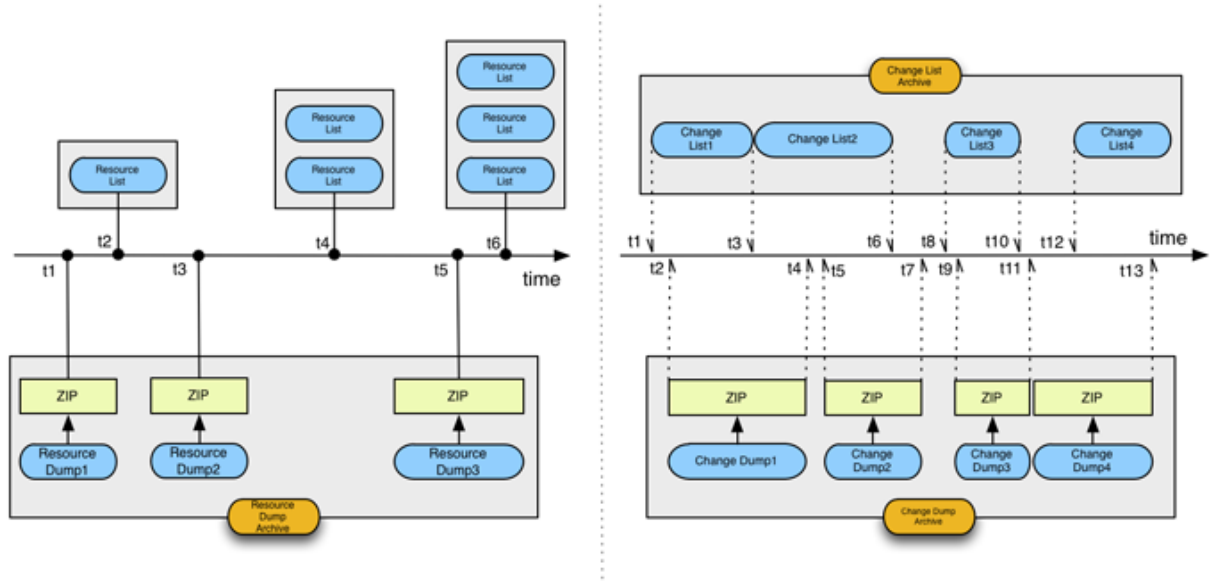
OPeNAP uses a different model of querying a remote data repository which returns data to a local data repository.

We should note from this use case and some others mentioned earlier that local and remote repositories need to be distinguished adding another dimension of complexity.

Summary: This use case is an excellent illustration of how complex practical policies can become and what kind of terms are relevant if such a processed is spelled out in great detail. As indicated the intention is to make the terminology used in DFT and in this use case compliant. In addition, this use case will be an inspiration for further work on core terminology in the RDA realm and it will inspire also the work in the Data Fabric interest group.

6. Web Resource Synchronization

Web resources and linked data are increasingly useful sources for research. But they often employ different standards such as URIs than traditional data. After creation, like centrally managed data they may be, updated, or deleted but also copied or moved. Both the Europeana project model supporting cultural heritage collections of institutions throughout Europe and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) model have been including in the Model Analysis. One use case relevant to them is synchronizing resources in a type of collection at Portals. An example would be the transfer/collection the metadata describing cultural artifacts from a participating institution to a collection portal and later copies of this resource ported out to new sites. To support this some mechanism is needed for transfer of actual data (e.g., images, videos, audio) and to keep this data permanently in sync with the metadata at the Portal.



The ResourceSync model is one that provides a mechanism for keeping track of the when and what of resource changes. As part of copying on an image resource, for example, a lower definition image may have been copied from the Portal. When updates from the original institution are made to the high quality image at the portal requests for copies must take this update into account based on the image quality of the destination. Example sequences from Time T1 to T6 are shown in the graphic above. Some update is provide to some resources at T3 and synchronization of this at T4 is managed through the use of a Resource List published then keeping info on the resource Source up to date. The Resource List enumerates and describes its resources and for each resource an URI and optional metadata and links are maintained. A requesting “Destination” uses a Resource List to synchronization with the stated of the Resource by dereferencing a listed URI.

Summary: The ResourceSync model being discussed here describes a modern protocol is not in conflict the DFT core model. The ResourceSync model basically describes a set of structures that can be used for executing a replication process. In doing so it is agnostic with respect to how the resources – be it data or metadata – are being organized on the originating and the replication side. Thus it is a model that is beyond the DFT core model and can thus serve as an inspiration for further work.

References

Chalmers, Iain, and Paul Glasziou. "Avoidable waste in the production and reporting of research evidence." *Obstetrics & Gynecology* 114.6 (2009): 1341-1345.

Open Archives Initiative Protocol for Metadata Harvesting, http://www.openarchives.org/pmh/

Reagan Moore’s Use Cases for RDA DFT, https://rd-alliance.org/filedepot?cid=100&fid=462