UNDERSTANDING THE CRYMAN PROJECT DATA CASE

MANSUR DARLINGTON

erim7rep110314mjd10.pdf

ISSUE DATE: 21 APRIL 2011

Catalogue Entry

Title	Understanding the CRYMAN Project Data Case
Creator	Mansur Darlington
Subject	information management; data management; research data; representation information; preservation description information
Description	The ERIM Project has developed a method for the better management of research data for the purposes of re-use and re-purposing. This method includes making explicit the associations between data records, alongside other contextual information. To demonstrate the efficacy of this approach, one of the ERIM case studies is revisited; the contextual information available as a natural by-product of the researcher's workflow is compared to the contextual information that would have been available had the researcher used the ERIM data management method.
Publisher	University of Bath
Date	14th March 2011 (creation)
Version	1.0
Type	Text
Format	Portable Document Format version 1.4
Resource Identifier	erim7rep110314mjd10
Language	English
Rights	© 2011 University of Bath

Citation Guidelines

Mansur Darlington. (2011). *Understanding the CRYMAN Project Data Case* (version 1.0). ERIM Project Document erim7rep110314mjd10. Bath, UK: University of Bath.

1. INTRODUCTION

The difficulty of re-using and re-purposing research data after the research project in which it was created has been discussed in Howard, et al. (2010). Fundamental to its re-use and re-purposing is an understanding of the data, and fundamental to this is the availability of contextualizing data which 'explain' how the elements of a data case (that is the data records) relate to one another. As part of the ERIM Project a method has been developed for the better management of data for these purposes which provides the basis for creating and recording a context based on explicit association between data records through the use of metadata and other means.

The following report consists of an analysis of an existing Data Case which allows a comparison to be made between on the one hand the context-bearing information associated with the data case that is the natural by-product of the creation and organization of this data set and, on the other hand, the context-bearing information which might result were a better during-project management approach applied of the sort proposed using the RAID Modelling Method.

The data case used for the analysis and comparison is that associated with the CRYMAN Project which was carried out over a period of about six years in the I*d*MRC at the University of Bath. This data case has been selected from a set of data cases which were the subject of audit as part of the ERIM Project. This particular case has been selected as being representative in size and diversity of many data cases which result from engineering design and manufacture research. It is also characteristic of many research activities where an early understanding of the topic is first achieved through the process of doctoral research, which is then carried on as post-doctoral work.

2. THE CRYMAN RESEARCH

This research concerned the development of an understanding of the machining of soft polymers made hard by being cooled to very low temperatures and thus machinable using conventional machining methods,.

The research focused on two key aspects of the cryogenic machining process. The first concerned exploration of the cooling process necessary to provide a suitable and stable machining environment. The second concerned exploration of the machining process itself, in particular finding machine tool parameters appropriate to the nature of the material being cut. The activities which illuminated these aspects of the research topic resulted in the creation of a great many data records, a characterization of these being made below.

3. GENERAL CHARACTERIZATION OF THE CRYMAN PROJECT FOLDERS

As can be seen below, over 3000 data records of one sort or another were generated during the research activity. The records were organized in a conventional folder/file structure (see Figure 1) by the researcher and were related to a number of distinct and identifiable activities which influenced the folder structure and file naming:

Folder Levels 3:

L1 36 folders, 498 files

L2, 79 folders, 2632 files

L3, 19 folder, 144 files

Total Folders 134
Total Files 3288
Total file size 5,375 Mb
File types 50 (see Table 1)

In addition to this there are physical entities created by the research activity which constitute data, such as swarf, in the form of polymer material removed during the cryogenic machining experiments, and the end products of the machining process. These entities have not been catalogued since they are not representative of research projects in general except in as much as they represent evidence of the existence of physical data objects.

Data file suffix	**Application**	**Proprietary/generic/ researcher's own**	**Comments**
System-identified			
Avi	Generic	Generic	Audio/video file
Csv	Generic	Generic	Comma separated variable file
Db	Microsoft	Proprietary	Thumbnail image file
Doc	Microsoft Word	Proprietary	
Docx	MS Word	Proprietary	Open XML, 2007 suffix
Ico	MS	Proprietary	Icon image file
Ini	Generic	Generic	Program initialization file
Jpg	Generic	Generic	
Pdf	Adobe	Proprietary	Portable document format file
Ppt	MS Powerpoint	Proprietary	
Pptx	MS Powerpoint	Proprietary	Open XML, 2007 suffix
Rar	RarLab	Proprietary	Compression file
Tif	Generic	Generic	image file
Txt	Generic	Generic	ASCII file
Vsd	Visio	Proprietary	
Xls	MS Excel	Proprietary	
Xlsx	MS Excel	Proprietary	Open XML, 2007 suffix
Zip	Generic	Generic	Compression file
Identified by researcher			
3dm	Rhino CAD	Proprietary	Rhino 3D model
Ccs	Delcam CopyCAD	Proprietary	
Dgk	Delcam PowerMill	Proprietary	Delcam process plan data
Dmt	Delcam CopyCat	Proprietary	Delcam machining triangle file
Igs	Not known	Generic	Generic 2D/3D Drawing
Lockfile	Delcam PowerMill	Proprietary	
Log	IGES	Proprietary	Unigraphics to IGES conversion log file
M	MATLAB	Proprietary	M-file

Data file suffix	Application	Proprietary/generic/ researcher's own	Comments
Mat	MATlab	Proprietary	variables binary file
Mpf	Siemens controller	Proprietary	Machine file for Siemens 840D controller
Pmldat	Delcam PowerMill	Proprietary	
Pmlent	Delcam PowerMill	Proprietary	
Pmllist	Delcam PowerMill	Proprietary	
Pmlms	Delcam PowerMill	Proprietary	
Pmlmt	Delcam PowerMill	Proprietary	
Pmlpar	Delcam PowerMill	Proprietary	
Pmlprj	Delcam PowerMill	Proprietary	Project file
Prt	Unigraphics	Proprietary	Part file
Stl	Delcam CopyCAD	Generic	Stereo-lithography
Stp	STEP	Generic	3D CAD file
Triton	TritonDMA	Proprietary	Material characterizing software
TritonDMA	TritonDMA	Proprietary	Material characterizing software
Uncertain identification			
Asv	Not known	probably generic	Perhaps generic autosave file?
db?encryptable	Not known	–	Probably related to move from PC to Mac platform
Def	Not known	probably generic	Probably related to move from PC to Mac platform
igs_tmp	Not known	probably generic	Probably related to move from PC to Mac platform
Unidentified			
Cylinder	Not known	possibly researcher's own	
cylinder 1	Not known	possibly researcher's own	
Fig	Not known	possibly researcher's own	
Sim	Not known	possibly researcher's own	
Stp	Not known	possibly researcher's own	
Tap	Not known	possibly researcher's own	

Table 1. Analysis of the applications and file types represented in the CRYMAN data case

3.1 Understanding the CRYMAN Data Case

A number of interviews were carried out with the researcher who was principally responsible for creation of the CRYMAN Data Case in order to understand the data, the data records, their relationships and the context in which the research occurred.

During these interviews a number of important things became clear. The CRYMAN data case constitutes evidence of the physical data records that were in some way a direct outcome of the research project. There are however, two classes of context-bearing information which are not represented in the Data Case but which, nevertheless, would be instrumental in aiding an understanding of the research work and of the recorded data. The first of these is the information which is a precursor to the research work and which informs the researcher in such a way as to facilitate the work, but has been recorded and stored elsewhere and, indeed may neither be referred to directly during the research nor identified as a source. The second class of information is that which constitutes the researcher's working knowledge, which is used to facilitate the work, but is not recorded at all. It is important to, at the least, acknowledge the existence of these classes of contextualizing information in order to understand fully the management of information for data re-use and repurposing.

It also became apparent that many of the data records were directly related in some way, but that little or no effort had been expended in making these relations explicit through some associative means. There are, for example, no external contextualizing files by which the data in many of the research data records may be understood, nor for the very large number of image files. Likewise, there are no *intentional associative research data records* in the data case. For the purposes of the research work, the provision of such contextualizing information had neither been necessary nor, from the point of view of the researcher, desirable, since to do so would have required the investment of a good deal of time. At the same time, some data, which had been judged to be worthless, have been deleted. This includes data which were overwritten as file content was modified in some way. In addition to this, there are no records in the Data Case which can be said to be that of the process: in fact, as the researcher himself said, 'the cryogenic research was carried out in a 'hands-on' way with little formal recording of the process itself'.

It was also clear from discussion with the researcher that the CRYMAN data case consisted in reality in a multitude of smaller data cases, each one of which was related to some activity, possibly carried out in parallel with others, which was not related in any way to the other activities. An example of this is the set of scanning electron microscope images of chip material removed during machining. These images are related to the machining activities which produced them (of which data records exist concerning machining parameters, and with which they are related) but are not related to the data logs which record the thermal profiles of materials when subjected experimentally to the cryogenic cooling. The two distinct research activities can be seen represented as Data Case 1 and Data Case 2 in Figure 2 below.

This notion of carrying out parallel activities during research is characteristic of much research in engineering design and manufacture research and can thus be said to be characteristic of much engineering design and manufacture research data. Yet, in the CRYMAN data case very little of the relations are made explicit through association, intentional or otherwise.

The following sub-section considers in what ways and to what extent the relations between the data records have been made explicit or suggested by contextualizing information.

3.2 Context-bearing information

In order to make sense of a Data Case it is necessary to have available some contextualizing information that is additional to the data themselves. Context-bearing information is any information that provides clues about the data records and the relationships between the data therein and other data or data records. In addition, the context-bearing information provides knowledge about the data themselves and their history in respect of how they have been collected, created and manipulated and, in some cases, how they should be interpreted or decoded. There are a number of dimensions which support inference about the context of the records and which can be found in general associated with the CRYMAN data case. These include:

- The structure and names selected for the folders
- The file names
- File type or format
- System and application metadata including that within the 'document properties' of Microsoft Office related records such as file name, creation date, location, size, etc.
- Standard user-contributed metadata such as author name, company name, keywords, etc.
- Customized user-contributed metadata, such as that relating to approval and sign-off, versioning, etc.
- Intentional or non-intentional context information carried in the file content.

Some of these items are the result of, often fairly standard, application behaviour and to a lesser extent system behaviour. Table 1 shows the diversity of applications that have been involved in record creation; some of these – for example Microsoft Office applications – will be similar in the way that they create, gather and record metadata and in the spectrum of metadata that is recorded, unless advantage has been taken of the customizable properties facility that might be offered by these applications. Other applications, especially bespoke software, will have uniquely specified and specifiable metadata. Other context-bearing elements are those that have been provided at some point by the user. The user in this case is the researcher who has been instrumental in the research by which the data have been created.

Although each data record may be inspected individually and the available information used to make associational inferences where and to the extent possible, this is a laborious process and one that for more than trivially small data cases is impracticable. It is only necessary to inspect the CRYMAN data case, through the medium of a standard folder/file organisational structure, to see what a daunting task would be the gaining of an understanding of the data within the case. That this is a fact is attested to by a second researcher who was passed the data by the data creator; he was able to understand the data and data records transferred to him only because he had some knowledge of the processes which had produced them; even so, he required clarification from the original researcher. As he attested himself, without that prior knowledge, making sense of the data records would have been intractable, and attempting to use them impossible.

As shown above there are over 3000 individual data records. Association is suggested by the folder organization, and further relations by the means bullet-pointed above. However, this information is not integrated. It is possible to inspect the contextual

information for one data record, and then use this in respect of another data record to draw some inferences about their relations. This can, in principle, be done in a stepwise manner to build up a picture of how the data fits together. It is clear, however, that this will be practical only in a small proportion of cases, and even then, only a limited amount of information can be drawn from the limited contextual data that have been captured. This is because the CRYMAN Project is one, characteristic of many, where during-project management of data for the purposes of use by others at a later data has been limited.

3.2.1 The folder structure and naming

The folder structure is informal and organic; informal because the researcher followed no specific organizational method or form, and organic, because the folders were created as suggested and dictated by the acquisition of the files placed therein. The folder structure (at the first level only) can be seen in Figure 1.

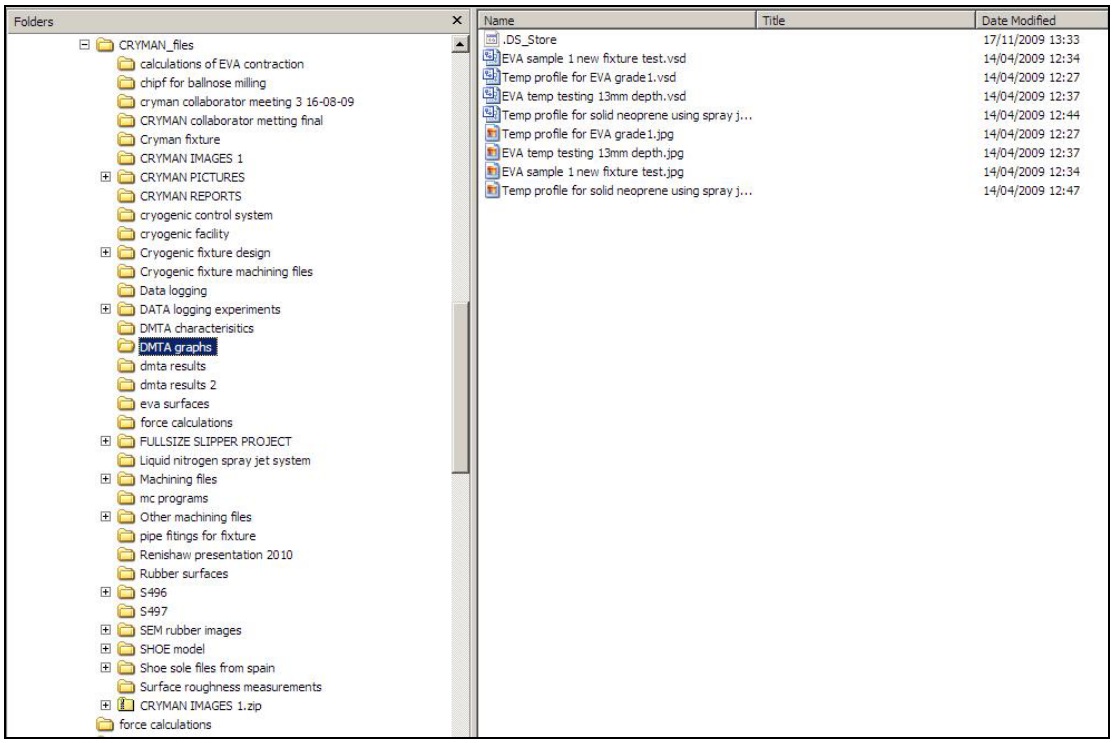

Figure 1. The CRYMAN Project data record folder structure

3.2.2 The file names

At project start no systematic method was adopted for file naming; rather files were named on the hoof by the researcher in a manner which seemed useful at that time. The naming of some sorts of files became more systematic as the research went on, the researcher providing in the file names some contextual information. In addition, the researcher revisited some key files at various points in the research to make them more instantly recognisable and these were then provided with more systematic names. An example of contextual information can be seen in the data logging file name '2spray_jet_t1_neo.csv'; the record contains test rig output data in comma separated variable form. It is the second test of the day associated with the cooling spray jet; is the first test for the material identified as neoprene. This does not conform to a formal file naming system because the elements and meaning of the elements are

not specified; furthermore, as show by inspection, other data records containing similar data content have file names constructed in a somewhat different manner. Thus, interpretation is ad hoc. As a result of this, interpreting the contextual information, and thus making inferences about the record content, may be difficult after the event for the researcher and impossible post-project for an individual with no knowledge of the project.

3.2.3 File types

As can be seen in the analysis, many file types are represented in the data case. File type information (usually encoded in the file suffix or extension) provides information conventionally about the format of the content and/or the application used to generate the file. Many of the file types found in this data case are standard and universally recognisable; other are less common (often relating to a particular software application which may be familiar in the research discipline or subject area) or entirely bespoke.

Of the 50 file types found approximately half were associated with 'standard' applications (e.g. MS Office products); of the remaining the creating researcher was able to identify all but ten. Of these four were identified with uncertainty; he was unable to identify at all a further six.

3.2.4 System and application metadata

A random inspection of within-document metadata showed that with few exceptions no use had been made of the opportunities for capturing context data using the facilities supplied by the applications. The exception was the, presumably system-provided, identification of the author, in Microsoft Office records. Even then, the author name did not always accord with the actual creator of the document. Context date associated with each record was limited almost exclusively to the applications- and system-generated metadata such as creation date/time, modification data/time, file location and file size and so on.

Similarly, little opportunity had been taken to populate the standard record properties nor to increase the system-provided contextual information that is available through the use of application-based metadata capture and no use had been made of the 'custom' properties facility provided by some applications.

3.2.5 Intentional or non-intentional context information carried in the file content.

There is some evidence, in specific file types, of intentional contextual information being provided as part of the content in some data records, especially where this resulted from following good experimental methodology. As an example, many of the records which contained logged data relating to a specific experimental event contain a worksheet which recorded the test rig experimental parameters. Whilst the intention of making this record was almost certainly to provide experimental validity and repeatability, nonetheless the parameter data (which includes such things as the tester's name, the data the experiment was run and so on) is clearly contextual information of the sort necessary to interpreting at a later time the logged data. This data might, then, be regarded as being 'unintentional contextualizing information'. This type of content-based contextualizing data is rare in the data case, and is limited to just few types of record. As a result understanding how the data and data records relate, and interpreting much of the data is very difficult.

It should be observed, however, that in general the context-bearing information that is available will depend on the circumstance of the generation and use of the research data. Data records that have been produced within a document management environment are likely to have more context-bearing information associated with them than otherwise, and the environment will provide some integration. Nevertheless, the purpose of document management environments is predominately not for the purposes of managing data for re-use or re-purposing but rather for general process management.

4. CONTEXTUALIZATION OF THE CRYMAN DATA CASE(S) USING THE RAID APPROACH

The RAID modelling method introduced in Howard, et al. (2010) and elaborated in Darlington (2011) provides the basis for the sort of during-research data record management that the authors believe is necessary to support data re-use. Such support could only be achieved by the *implementation* of the modelling method, which has yet to be carried out. Thus it is not possible to apply this method wholesale to the large CRYMAN data case.

The following treatment, however, with a small sub-set of records from the CRYMAN data case, demonstrates the additional capacity for understanding that can be achieved using the RAID modelling method. Principally, the method relies on the semi-automatic capture of metadata which describes individual data records and their associations with other data records.

By way of example a small number of data records have been selected which are known (by the researcher involved) to be related in some way. A comparison will be made between the context-bearing information available at the time of hand-over, and that which would, in principle, be available were the RAID method of association implemented.

The RAID approach involves the automatic generation of associational and other metadata during the development of research information during the research project. Information that is unavailable to the system may be augmented by additional information that is input manually by the researcher, often as a result of system prompting. Some part of the metadata (to a certain extent customizable by the user) which describes a data case will be usable to generate a visualization of all or part of the data records within the data case together with their important relations, thus promoting an understanding of the data and data records and the way they fit together.

For the purposes of illustration a small set of CRYMAN Project data records are shown in Figure 2. The metadata revealed here is only a very limited subset of the total that might be specified in a RAID application implementation. The extent of the metadata that might be gathered and recorded to provide contextualization is given in the RAID Associative Tool Requirement Specification (Darlington, 2011). The representation shown in Figure 2 is a prototype and should be seen as suggestive of the sort of diagram that might be available on demand in an implementation of the tool. It can be seen clearly, however, that a number of data records are identified, together with the context in which they have been developed. This should be compared with the limited information that is actually contained as metadata within each of the existing CRYMAN files and which might be used in conjunction with the CRYMAN Project files store folder structure suggested in Figure 1 above to provide some basis for understanding the data case.

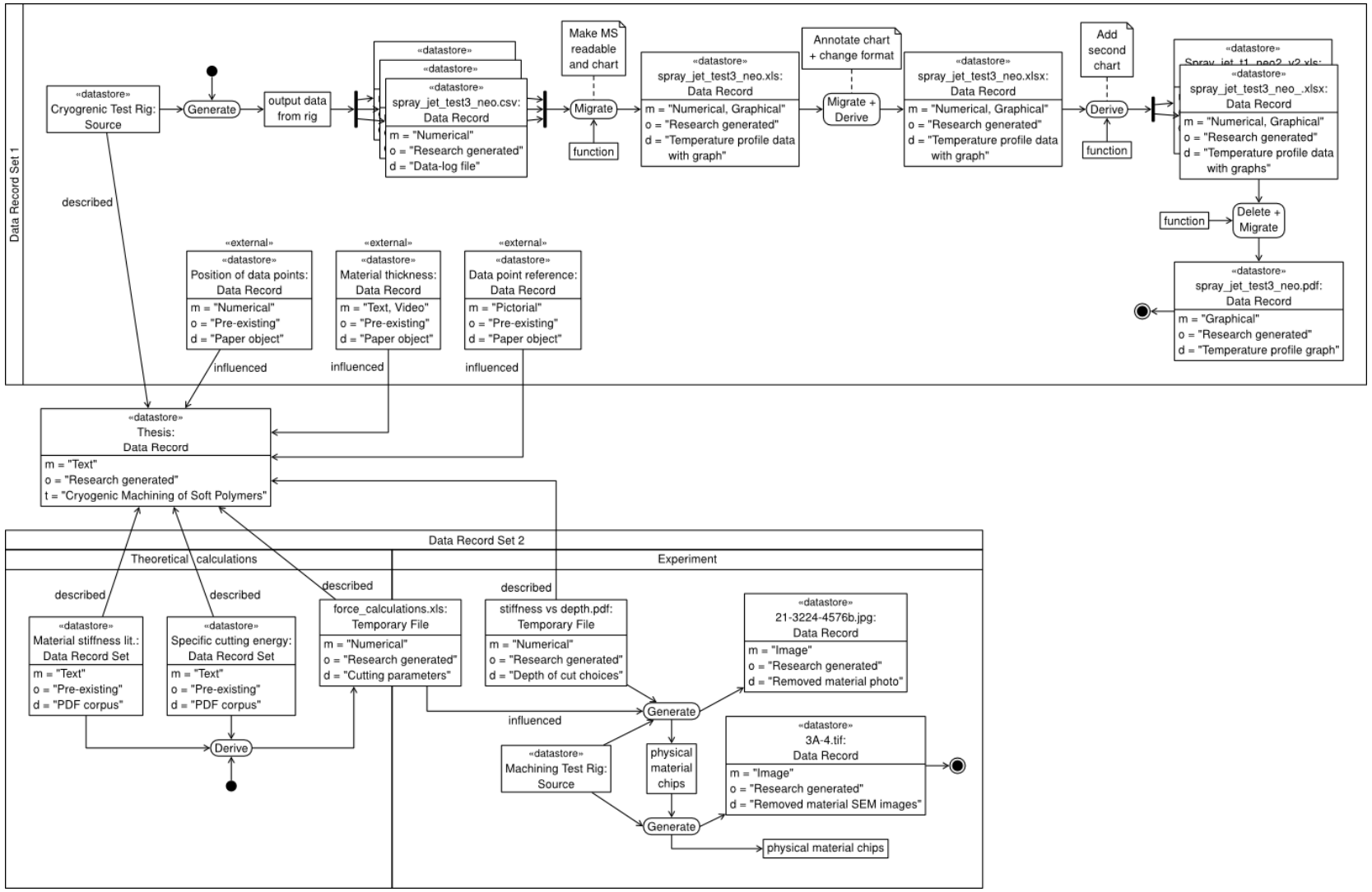

Figure 2. A RAID Diagram of selected related data records in the CRYMAN data case revealing important associational information

4.1.1 The existing contextualizing metadata

The contextualizing metadata carried by the existing files from the CRYMAN Project is illustrated below in the case of representative data records. It should be noted that the data record with the file name 'spray_jet_test3_neo.xlsx' exists in two locations and with different content. It is also the case that often there will be evidence of only one instance of a file since the researcher followed the practice often of overwriting the file each time it was modified in some way. In contrast to this, the RAID diagram suggests the preservation of each version of a data record with appropriate amendment to the file name and metadata so that data losses (e.g. State Loss and Information Loss) are avoided.

Spray_jet_test3_neo.csv
File type: Microsoft Comma Separated Value file
Created 03 Feb 2009
Modified 03 Feb 2009
No summary information
Spray_jet_test3_neo.xls (MS Excel-readable plus chart)
File type: Microsoft Excel Worksheet
Open with MS Office Excel
Created 04 Feb 2009
Modified 04 Feb 2009
Printed 03 February 2009 (sic)
No summary information
Spray_jet_test3_neo.xlsx (change of format and annotate chart)
File type: Microsoft Office Excel Workbook
Open with MS Office Open XML Converter
Created 14 April 2009
No Summary information
Spray_jet_test3_neo.xlsx (as earlier version but with addition of the original chart)
Microsoft Office Excel Workbook
Open with MS Office Open XML Converter
Created 14 April 2009
Modified 14 April 2009
Spray_jet_test3_neo_.pdf (contains the graphical representation alone of the annotated chart)
Title: spray_jet_test3_neo.pdf
Author: administrator
Pdf Version 1.3 (Acrobat 4.x)
Pdf produced by PrimoPDF

Table 2. Sample of contextualizing metadata found in a representative series of related data records

Clearly the metadata that is found embedded within these data records provides a limited basis only for making associations between them. If, as has been assumed, it is

inferred that the records are related by virtue of their file names, then some sort of ordering is permitted given this information together with information about opening the records for inspection and use.

4.1.2 Metadata representative of that provided by a RAID associative tool

By way of contrast, the second MS Office Excel XML data record is treated below in a fashion that would be the case assuming metadata capture is carried out as prescribe for the RAID associative tool. This particular file has been selected because its relation to and association with precursor and successor files may be demonstrated.

Attribute Label	**Value**	**RAID-related Label Definition**
Data development process type	Derivation	One of a set of processes that are commonly carried out during the Research Activity which changes or adds to the Research Data associated with a research activity or project.
Data record type	research record	One of five types of data record specified in the RAID model.
Data object type	Digital	Either a Physical Object or a Digital Object.
RAID datatype description	data record	Content-level classification of a data record.
1st-Generation Data Type	research-generated	The Data resulting either from Data Collection or from Data Generation.
Media type	Numerical + pictorial_2D	n/a
Data format	Microsoft Office Excel 2007	The organization of digital information according to a preset (sometimes application-specific) specification.
Reality	Real	Event classification reflecting conditions in which research data were collected.
Stability	closed {dormant}	Dataset status relating to expectation of future dataset development.
Collection method	experimental rig	The means by which the research data were collected, generated or recorded.
Repeatability	Repeatable	A measure of the practical possibility of a Generation process being repeated such that Data Reproducibility is possible in principle.
Interpretation	Objective	Indication of whether derived data were the result of objective or subjective interpretation of research data.
Dublin Core Sub-set		

Attribute Label	**Value**	**RAID-related Label Definition**
Creator	Vimal Dhokia	Probably the researcher or research team responsible for making or collecting the data record.
Description	The point data from an experimental rig measuring the temperature over time of a matrix of sensors embedded in a cryogenically cooled work piece	A brief account.
Date	3 April 2009	The date of acquisition/generation into the data case.
Identifier	spray_jet_test3_neo.xlxs	An unambiguous reference to the data record within a given context; preferably based on a formal naming system such as a file naming protocol.
Source	spray_jet_test3_neo.xlxs (file name used twice)	A reference to a data record from which the present data record is derived through a Research Data Development Process.
Subject	cryogenic machining, soft polymers	The topic represented in keywords.
Title	Cryogenic Freezing Temperature Profiling of Neoprene	A name given to the data record.
PREMIS sub-set		
Creating application	Microsoft Office Excel 2007 XML; unknown version	The application which created the data record, together with version.
originalName	n/a *(applicable only to pre-existing and renamed data objects)*	The name of the pre-existing object as submitted to or collected by the researcher (team), before any renaming has been carried out.
storageMedium	hard disk	The physical medium on which the object is stored.
Dependency	n/a	Information about a non-software component or associated file needed in order to use or render the representation or file, for example, a schema, a DTD, or an entity file declaration.
relatedEventIdentification	n/a	
DataCite v2.0 sub-set		
Relationship sub-type:	n/a	
IsSupplementTo	n/a	Use to indicate the relation to the data record to which this data record is a supplement.

Attribute Label	**Value**	**RAID-related Label Definition**
IsSupplementedBy	n/a	Use to indicate the relation to the data record(s) which are supplements of this data record.
IsContinuedBy	Spray_jet_test3_neo.pdf	Use to indicate the data record is continued by the data record referenced by the related identifier.
Continues	n/a	Use to indicate the data record is a continuation of the data record referenced by the related identifier.
IsNewVersionOf	Spray_jet_test3_neo.xlxs	Use to indicate the data record is a new edition of an old data record, where the new edition has been modified or updated.
IsPreviousVersionOf	n/a	Use to indicate the data record is a previous edition of a newer data record.
IsPartOf	n/a	Use to indicate the data record is a portion of another data record.
HasPart	n/a	Use to indicate the data record is a container of another data record.
IsReferencedBy	n/a *(it would be useful if there were such data records which provided information about the column and row headings in the spreadsheet)*	Use to indicate the data record is used as a source of information by another data record.
References	n/a *(comment as above)*	Use to indicate the relation to the data record which is used as a source of information of the data record.
IsDocumentedBy	n/a	Use to indicate the data record is documentation about/explaining the data record referenced by the related identifier.
Documents	n/a	Use to indicate the relation to the data record which is documentation.
IsCompiledBy	n/a	Use to indicate the data record or data is compiled/created by using another data record or dataset.
Compiles	n/a	Use to indicate the data record is used for creating another data record or dataset.

Attribute Label	Value	RAID-related Label Definition
isVariantFormOf	n/a	Use to indicate the data record is a variant or different form of another data record, e.g. calculated or calibrated form or different packaging.
isOriginalFormOf	n/a	Use to indicate the relation to the data records which are variant or different forms of this data record.

Table 3. Example contextualizing metadata for the data record *spray_jet_test3_neo.xlxs* using RAID associative tool-specified metadata

The metadata properties suggested above are taken from the RAID Associative Tool Requirement Specification. This set should be regarded as a starting-point for the selection of metadata that would best support the (semi-) automatic acquisition of information necessary to provide optimal contextualization of research data for the purposes of use, re-use and re-purposing. To the metadata given above should be added whatever data entities are required in the implementation of the RAID Association Tool, for example, those data entities required to provide a coherent visualization as exemplified in Figure 2. In addition, it should be assumed that the commonly encountered 'system' or application metadata which identifies such things as creation date, modification date and so on would augment those specified.

5. CONCLUSION

The above analysis has provided a comparison between the contextualizing information available in two cases relating to an example data case.

Whilst representative only, the information provided in the metadata values given in **Table 3** clearly provides the basis for understanding the relationship between the example data record and other data records with which it is related. Were this approach adopted, then, it would be possible to scrutinize and query the entire metadata set which associates all the data records in a data case, providing a sound basis for transmission of the sort of knowledge about the data that would be required for a good understanding and as a basis for its re-use and repurposing.

6. REFERENCES

Howard, T., Darlington, M., Ball, A., Culley, S. and McMahon, C., 2010. *Understanding and Characterizing Engineering Research Data for its Better Management.* Project Report. Bath, UK. University of Bath. URL http://opus.bath.ac.uk/20896/

Darlington, M. 2011. *RAID Associative Tool Requirements Specification*. Project Report, Bath, UK, University of Bath. ERIM Project Document erim6rep101111mjd10. URL http://opus.bath.ac.uk/22811