The International Journal of Digital Curation
Issue 1, Volume 3 | 2008

Challenges and Issues Relating to the Use of Representation Information for the Digital Curation of Crystallography and Engineering Data

Manjula Patel,
Digital Curation Centre, eBank-UK Project and KIM Project,
UKOLN, University of Bath

Alexander Ball
KIM Project
UKOLN, University of Bath

July 2008

Abstract

The OAIS concept of Representation Information (RI) is a potential strategy for the curation and preservation of all types of information. We share insights gained from our exploration of issues concerned with the capture of crystallography and engineering RI as well as its structuring, collection and curation. In addition, we discuss the supporting technical, IPR and globally collaborative infrastructures required to make such a strategy successful.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

The term *digital curation* has evolved to the stage where it is now generally accepted as including the active management of digital data and research results over their entire scholarly and scientific life-time, both for current and future use. The term also encompasses the notion of adding value to a trusted body of digital information as well as its reuse in the derivation of new information and the validation and reproducibility of scientific results (Beagrie, 2006)[1]. Curation, in the first instance requires a commitment to undertake duties of custodianship. However, such a commitment is influenced by a complex array of factors including: social, political, organizational, financial and technical issues.

Due to technological obsolescence (hardware, software and file formats), which constitutes a major threat to digital information, data can become inaccessible within a very short time. Moreover, much digital information requires software applications in order to make it accessible to humans.

A particular strategy concerned with mitigating the effects of technology evolution is based on the use of *Representation Information* (RI) – a concept used in the Reference Model for an Open Archival Information System (OAIS) (Consultative Committee for Space Data Systems [CCSDS], 2002). RI is all-encompassing; it comprises any information that is required to render, process, visualize and interpret data, and includes: file formats, software, algorithms, standards and semantic information. In addition, an OAIS *Archival Information Package* (AIP) comprises both RI and *Preservation Description Information* (PDI) and is in a sense, a form of encapsulation collecting together all the information relevant to the preservation, interpretation and reuse of digital data.

We provide a brief overview of RI and a Registry/Repository of RI (RRoRI) followed by two case studies, the first in the area of crystallography data and the second within the engineering domain. By choosing two domains with very different characteristics we hope to expose and explore a wide range of issues relating to the capture and use of RI for curation and preservation purposes. Finally, we discuss the insights that we have gained from the two case studies.

Representation Information

The OAIS model establishes a common framework of terms and concepts for use in the preservation of information. It has established itself as an important standard, influencing: the development of preservation metadata; architectures and systems of repositories; and conformance and certification criteria for archives.

Information in the OAIS model is regarded as being a combination of data and RI. An *Information Object* is composed of a *Data Object* that is either physical or digital, as well as the RI that allows for the full interpretation of the data into meaningful information. Furthermore, RI is recursive in nature; using one element of RI in a meaningful manner may well require further RI, so that a RI Network is created. It is expected that the recursion will terminate for a particular *Designated Community* when the RI can be understood using that designated community's *Knowledge Base*.

[1] The Digital Curation Centre (DCC) http://www.dcc.ac.uk/

It is therefore essential that RI itself be curated and preserved to maintain access to other digital data. The OAIS model identifies three main types of RI: structural, semantic and other. In a digital era, structural information manifests itself largely in the form of digital file formats for text, images, audio, moving images, datasets, 3D models as well as time-varying or dynamic data and can greatly affect the ease with which the information contained within can be accessed. Formal descriptions of file formats are useful in enabling automated processing. Semantic Information provides additional meaning to the contents of a digital object. This category includes data dictionaries and knowledge organisation systems such as schemata, ontology, metadata vocabularies and thesauri. Additional RI is classified under the "other" category and includes software, algorithms, standards, time-dependent information, actions, processes and so forth.

Registry/Repository of Representation Information (RRoRI)

The Digital Curation Centre (DCC)[2] and the CASPAR Project (Giaretta, 2007) are developing a registry/repository of RI (RRoRI)[3]. This is intended to be an authoritative source of RI for use by those responsible for the collection, curation and management of data. The work is heavily based on the ideas in the OAIS model; it centres on the notion that RI is critical to the long-term access of digital information (Giaretta, Patel, Rusbridge, Rankin, & McIlwrath, 2005; Giaretta, Rankin, McIlwrath, Rusbridge, & Patel, 2005). Collection and maintenance of suitable RI mitigates the difficulties related to the preservation of understandable information. Much emphasis is placed on interoperability and automated use, the vision being to have a global, distributed network of reliable and trusted RI on which others can rely, and which anyone can use.

To enable access to RI by third parties, RRoRI uses an *RI Label*, with a *Curation Persistent Identifier* (CPID) used to connect RI to a particular data object[4]. The Label provides a mechanism for combining individual RI components and may be a structured digital object itself (i.e. it specifies a RI network).

Crystallography Data

High-quality data are the raw materials of contemporary e-science, used in scientific endeavour as well as verification and replication of scientific results. There are however, disciplines such as crystallography within which there is a bottleneck in the publication and accessibility of raw and derived data. Whilst the generation of raw data has accelerated with advances in instrumentation and computational tools, only a fraction is available in the public domain (Allen, 2004). This shortfall has arisen largely because current publication mechanisms have not managed to keep pace with increased experimental throughput, but also because research scientists typically do not archive and curate their data adequately enough to make it easily accessible and reusable by third parties.

[2] DCC Development Team: DCC Approach to Digital Curation http://dev.dcc.rl.ac.uk/twiki/bin/view/Main/DCCApproachToCuration
[3] Registry/Repository of Representation Information (RRoRI) http://registry.dcc.ac.uk/
[4] DCC Development Team: DCC Information Label Report http://dev.dcc.rl.ac.uk/twiki/bin/view/Main/DCCInfoLabelReport

In response, the eBank-UK Project[5] has constructed an institutional data repository (eCrystals[6]) to make available the raw, derived and results data from a crystallographic experiment (Coles et al., 2006; Duke, Day, Heery, Carr, & Coles, 2005). Building on the concept of open access, the project has focused on the laboratory-based experimental technique of chemical crystallography undertaken at the UK National Crystallography Centre (NCS). Following the creation of a completed crystal structure determination, data are uploaded into eCrystals and supplemented with chemical and bibliographic metadata. eCrystals is built using a modified version of the ePrints.org repository platform[7]. At present eCrystals has no formal long-term commitment to preservation and curation of its holdings; however this aspect is currently under investigation (Patel & Coles, 2007).

The designated community for the eCrystals data repository is well defined; it is the worldwide crystallography community and the broader discipline of Chemistry. In addition, crystallography has a tradition of sharing results data in an internationally accepted exchange format, the Crystallographic Information File (CIF)[8] supported by the International Union of Crystallography (IUCr)[9]. The Cambridge Structural Database (CSD) provides international facilities for the acquisition, storage, validation, retrieval, analysis and visualization of small-molecule crystal structures, again mostly available in CIF format. Furthermore, many crystallographic journals encourage or mandate the submission of structures in CIF format and the CSD acts as a data depository on behalf of a number of these journals. Other key stakeholders include the Royal Society of Crystallography and Chemistry Central (publishers) and the Reciprocal Net and Crystallography Open Database which enable data to be uploaded for open use.

Crystal Structure Determination Workflow

Crystallography is concerned with determining the structure of a molecule and its three-dimensional orientation with respect to other molecules in a crystal by analysis of diffraction patterns obtained from X-ray scattering experiments. In each experiment, the process relates to the determination of one structure, comprising both the molecular connectivity and the packing arrangements between molecules in the crystal being examined. The final result is a crystal structure in the form of a CIF file.

Procedures at the NCS indicate that a number of well-defined, sequential stages are readily identifiable and result in a workflow as shown in Figure 1. At each stage, an instrument or computational process produces an output, saved as one or more data files which provide input to the next stage. The output files vary in format, they range from images to highly structured data expressed in textual form; the corresponding file extension names are well established in the field. Some files also contain metadata, such as validation parameters, about the molecules or experimental procedures.

[5] The eBank-UK Project http://www.ukoln.ac.uk/projects/ebank-uk/
[6] The Crystal Structure Report Archive –eCrystals Data Repository http://ecrystals.chem.soton.ac.uk
[7] ePrints.org: EPrints for Digital Repositories http://www.eprints.org/
[8] CIF -The Crystallographic Information File http://www.iucr.org/iucr-top/cif/
[9] International Union of Crystallography (IUCr) http://www.iucr.org/

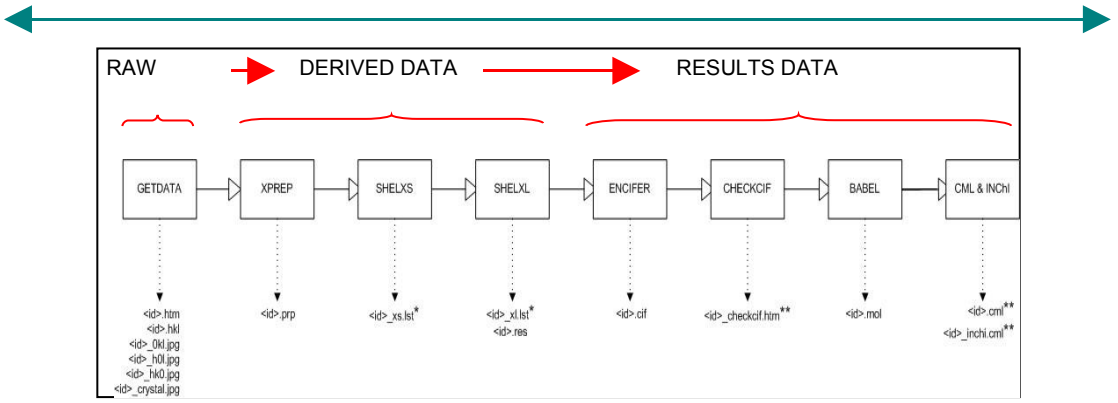

Figure 1. Workflow model of the EPSRC UK National Crystallographic Centre (NCS).

Data collection provides JPEG files as representations of the raw data, but also proprietary formats generated by specific instrumentation that may be in use. This stage may also include an HTML report file, providing information relating to machine calibrations and actions and how the data were processed. The main result of the processing stage is a standardized ASCII text file (HKL), which has become a historical *de facto* standard within the designated community through its requirement by the SHELXL suite of programs[10]. The solution stage results in a log file (LST) comprising information relating to the computer processes that have been run on the data by the SHELX software and a free-format ASCII text file (PRP), which is generated by software (XPREP). The SHELXL software produces both an output (RES) and a log file in ASCII text format.

The final three processes (final result, validation and refinement) produce community-adopted standard file formats[11]. CIF is a publishing format as well as being structured and machine-readable, using the ENCIFER software[12]. Associated with the CIF format is the widely used CHECKCIF service[13] for the validation of CIF files producing an HTML file as its result.

A Chemical Markup Language (CML)[14] file is also generated to provide automated exchange of the chemical structure information. Several file format conversions are performed according to well-defined standards using the open source toolbox, OpenBabel[15]. Finally, the International Chemical Identifier (INChI)[16] is generated to provide a unique text representation of the molecule.

Example RI Network for the CIF File Format

Within crystallography, we have initially chosen to examine the RI network associated with the CIF file format, since this appears to be the most important in the community at present. Space limitations and the recursive nature of RI networks mean that we are unable to reproduce the entire RI Network here (see Figure 2). However, a more complete (textual) version is available on the Web[17] and provides an indication of

[10] The SHELX software suite http://shelx.uni-ac.gwdg.de/SHELX/
[11] Example eCrystals Data Structure Archive Report http://ecrystals.chem.soton.ac.uk/300/
[12] enCIFer software http://www.ccdc.cam.ac.uk/free_services/encifer/
[13] IUCr checkCIF validation service http://checkcif.iucr.org/
[14] Chemical Markup Language (CML) http://www.ch.ic.ac.uk/rzepa/cml/
[15] Open Babel: The Open Source Chemistry Toolbox http://openbabel.sourceforge.net/wiki/Main_Page
[16] International Chemical Identifier (INChI) http://www.inchi.info/
[17] Example Representation Information Network for CIF file format
http://homes.ukoln.ac.uk/~lismp//IDCC2007/RINetCIF.html

the complexity and granularity of the information required. We are currently engaged in experimenting with a prototype ingest tool in an attempt to populate RRoRI with the CIF and IGES 5.3 (see section *Example RI Network for IGES 5.3 File Format*) RI Networks.

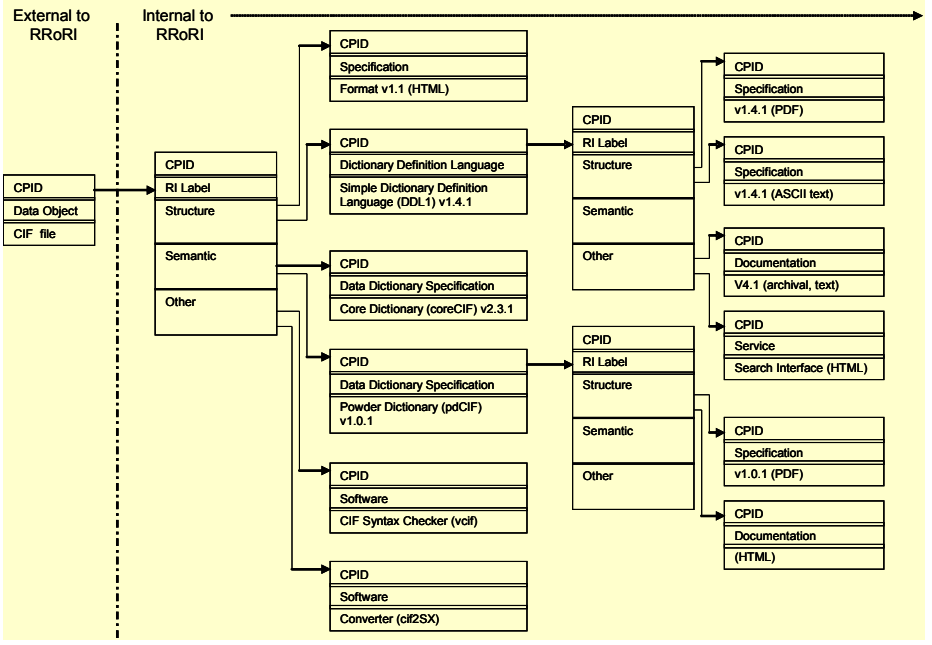

Figure 2. Graphical visualization of part of an RI Network for the CIF file format.[18]

Engineering Data

Engineering as a discipline covers a wide variety of industries (mechanical, electrical and civil engineering, for example, as well as architecture and the built environment) each with their own characteristics and working practices. However, they have in common the concepts of a product and a product life-cycle: in the course of engineering activity, a product is designed, produced (manufactured or constructed), put into service, maintained, refitted/adapted, and finally retired and disposed of.

Within the product life-cycle, the two principal sources of engineering data are the design/redesign phase, when product models, testing data and so forth are generated, and the in-service phase, when performance is monitored and recorded. It is important that both types of data are properly curated throughout the lifetime of the products concerned, for reasons of both regulatory compliance and practicality. In-service data can inform not only future maintenance cycles but also feed back into future designs, while design data can assist maintenance activities as well as future designs and redesigns.

In areas of engineering where products are typically commissioned or procured through high-value contracts – aerospace, defence and construction are typical – there is an increasing trend for the products to be purchased as services rather than artefacts. For engineering companies, this means that their responsibility is not merely to produce and deliver products, but to ensure that the capabilities of those products are delivered throughout the life of the contract; in practice this means taking over

[18] Example Representation Network for IGES 5.3 file format
http://homes.ukoln.ac.uk/~lismp/IDCC2007/iges.html

responsibility for the maintenance, servicing, refitting and (in some cases) replacement of the products. This 'product-service' way of working can be characterized as a new paradigm within the engineering domain (Cambridge Manufacturing Review, 2003; Davies, Brady, & Tang, 2003; Oliva & Kallenberg, 2003).

This new paradigm comes at a time when radical changes are being made to the way product data are recorded and used. Traditionally, computer-aided design (CAD) packages were used to create the technical drawings that defined the design of a product. From the turn of the twenty-first century, however, CAD packages started integrating with downstream applications such as numerically controlled machine tools and inspection routines, meaning that the CAD model took over as the definitive expression of the design.

These two trends together – companies having to commit to supporting particular designs for much longer periods, and the designs being defined by CAD models – are in conflict with the state of the CAD software industry, which is dominated by ephemeral, backwards-incompatible, proprietary applications and file formats. This issue is one that has been recognized for quite some time, although solutions have traditionally been geared towards the exchange of product model data between contemporaneous designers.

One of the earliest attempts to provide an open, neutral format for CAD was IGES, the Initial Graphics Exchange Specification (US Product Data Association [USPRO], 1996), first published in 1980. While promising, this format proved to be too flexible, allowing CAD vendors to produce mutually incompatible subset implementations. In response, the Standard for the Exchange of Product Model Data (STEP) was introduced in 1994; published as ISO 10303, this standard has 454 current parts and corrigenda, and is constantly being updated and expanded (South Carolina Research Authority, 2006). While certainly addressing the issues, STEP is a victim of its own breadth and depth; its application profiles often fail to interoperate and it is expensive to implement.

Given that engineering is primarily a commercial enterprise rather than a public service, the concept of a designated community for engineering data is a little more complex than for academic disciplines. The immediate designated community for design information is the design section that produced it. Design teams in other organizations are strictly excluded on the basis of commercial sensitivity, although they may be occasional consumers of parts of the data when collaborating on a given project. This immediate community may be subject to sudden and sweeping change as companies merge, sub-divide or are subsumed into larger companies.

Beyond the immediate designated community there is a regulatory designated community: regulatory authorities, accident investigators, etc.

As engineering companies are responsible for more of the product life-cycle than ever before, the processes and tools by which they manage the life-cycle must become ever more sophisticated. A scenario has been devised by the KIM Project[19] to illustrate ideal information flows through the product life-cycle model.

To summarize, there are flows of information from the original designers to the refit designers, from both sets of designers to in-service engineers and incident investigators, between in-service engineers, from in-service engineers to incident investigators, from incident investigators to refit designers and from incident investigators to future designers. In all cases the flow of information must be managed to ensure that exactly the right amount of information is transmitted and adequately understood.

Example RI Network for IGES 5.3 File Format

The Initial Graphics Exchange Specification (IGES) version 5.3 (USPRO, 1996) has been chosen as an example in the engineering domain because IGES was the first popular exchange format in this area, and is likely to remain popular until STEP implementations are mature. IGES 5.3 is the last official version of the format, published as a standard in 1996. Again, space restrictions mean that we are unable to reproduce the whole RI Network in Figure 3, but a more complete (textual) version is available on the Web[20].

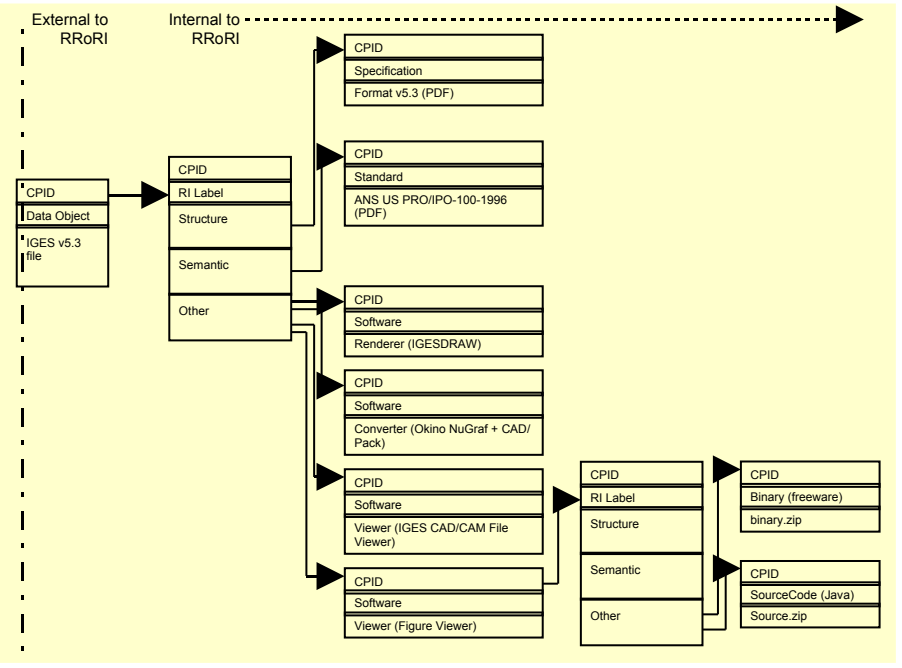

Figure 3. Graphical visualization of part of an RI Network for the IGES 5.3 file format[20].

Challenges and Issues in Capturing Representation Information

By investigating two domains with very different characteristics we are able to

[19] Knowledge and Information Management (KIM) Grand Challenge Project https://www-edc.eng.cam.ac.uk/kim/

[20] Example Representation Network for IGES 5.3 file format http://homes.ukoln.ac.uk/~lismp/IDCC2007/iges.html

expose and explore a wide range of issues relating to the capture and use of RI for curation and preservation purposes. Crystallography, within an academic environment, is a strongly bounded domain. It has a limited number of stakeholders and a strong tradition of using open standards and software, as well as a culture for sharing data. Engineering, on the other hand, is a broad area with vested commercial interests and a proliferation of proprietary file formats and closed software solutions.

Constructing RI Networks

Without a doubt, in practice, documenting and structuring RI networks is a non-trivial task, not only because of the sheer amount of information that is required, but also because it is difficult to know when to end the recursion. The expectation of the OAIS model is to use the knowledge base of the designated community to stop the recursion–this presupposes a well-defined knowledge base at any particular time. The tacit, implicit and dynamic nature of knowledge means that a knowledge base may be difficult to define; however, one possibility is to describe it as a set of familiar software applications, community standards, contextual descriptions and topic categorizations. The CASPAR Project is currently investigating a more explicit technique for capturing a user's knowledge profile (Giaretta, 2007).

In addition, since comprehensive RI is needed to preserve access to information, it is necessary for those involved in creating RI networks to understand the variety of forms RI may take and to identify what is a composite part of the data object, and what is required by the end user in order to be able to work with the data.

It should also be borne in mind that there may exist more than one RI network, each equally applicable to a particular digital object.

Building robust RI networks requires effective search and retrieval techniques for RI further down the chain, ideally from a single interface. Without such a facility there is likely to be a huge duplication of effort.

Creating RI networks is tantamount to taking tacit, unstructured and dynamic knowledge within a domain and making it explicit and structured with encoded relationships to enable automated processing. In a sense, this is analogous to the work required to create the Semantic Web (Berners-Lee, Hendler, & Lassila, 2001) and is likely to share similar issues.

Although such an approach is heavy on human effort, on the positive side, once an RI network is available, it can be used to process all digital objects of a particular type. Additionally, it may become possible to mine the information in the RI networks to make inferences with respect to curation and preservation strategies.

Classification of RI

Of necessity, the classification scheme in RRoRI is expressed at a high level since the aim is to capture all types of RI. Such coarse granularity, though, is unlikely to be sufficient for certain individual disciplines, which will have their own view of the universe and may have a requirement to capture particularly significant characteristics of specialist data.

Furthermore, without consensus on the semantic definitions, the interpretations placed on the elements of the classification scheme will be subject to variation. The

effectiveness of the classification will also have an impact on the relevance of search and retrieval of RI. Once again, parallels can be drawn with issues relating to knowledge representation and the construction of an ontology for use in the Semantic Web.

IPR and Rights

Over the broader picture of digital preservation, significant issues have been raised with respect to the restrictions that IPR and copyright laws place on archival activities. The legislation is out of date and does not make any provision or concessions to the copying of digital information for preservation and curation purposes. More recent frameworks such as the Creative Commons and the Science Commons, although not specifically geared to preservation issues, are more flexible in nature.

The capture and maintenance of RI in domains such as crystallography, where open standards and software are prominent, is considerably easier and more likely to be successful than, for example, in engineering which is rife with closed formats and software (the pharmaceutical industry is similar to engineering in this respect, and thus represents a converse example to that of crystallography).

In the area of CAD, rights are a particular area for concern. In order to cope with the increasing complexity and informational scope of CAD models, many CAD vendors have shifted the burden of representing the information away from the file format and towards the software used to render and edit it. Partly this is achieved by using 'features': predefined parts that a designer can adapt to a particular situation by adjusting certain parameters. Since these features are the intellectual property of the CAD vendor, the intellectual ownership of models that use these features is less clear than before.

From a preservation perspective, the intimacy of this connection between the files representing a model and the software used to create them multiplies the vulnerability of these models to recoverability. The file formats used by most CAD software are both proprietary and unpublished, but even if they were published and freely implementable, the specifications alone may not be sufficient in order to interpret the files. In many cases the code for the software is needed as well, and this is (understandably) kept closed by the vendors.

Although CAD is used here as the prime example of the problem in engineering, there are other areas of engineering information, such as numerical control programs for manufacturing robots, which may also prove to be problematic.

Continuous Monitoring

The notions of a designated community and its associated knowledge base in the OAIS model are both dynamic concepts. A designated community has, at any particular time, a particular knowledge base. For a specific designated community this knowledge base will evolve over time. In addition, the definition of the appropriate designated community for a dataset may also change over time. The implication is that both the designated community and the knowledge base need to be continuously monitored, which entails liaison and collaboration between previously divorced roles in the creation, publication and long-term curation of digital information.

Technical Infrastructure

As recognized by the DCC and the CASPAR Project, in order to use RI successfully as a means of curating digital information, a reliable and robust technical infrastructure is necessary. As a minimum repositories and archives must record at least one CPID as part of the preservation metadata for a specific digital object. The CPID needs to be supported through a resolver service to enable the automatic traversal of an RI network.

Moreover, due to the continuously changing nature of designated communities, knowledge bases and technologies, it is imperative that RI and RI networks be maintained and updated. There is a danger here of that frustrating situation arising, familiar from the Web, known as link-rot or the 404 problem.

Furthermore, the traversal of a RI network requires the curation and maintenance of the RI Label and the RI network, being digital objects in their own right, as well as the actual RI. As for all types of digital information, this includes bit-preservation, encompassing secure storage, error detection and correction, verification, as well as data integrity, authenticity and validity checks.

While reliance on the IT infrastructure is necessary for automated processing – large volumes of digital data leave us with no other choice – there are many potential risks involved in relying on the technological environment; this itself exacerbates the problem of keeping digital information alive and fit for purpose and reuse.

Cost/Benefit Analysis

We have found that the actual task of creating and maintaining RI networks is time-consuming and non-trivial. It is therefore important to resource such efforts adequately – in this respect models such as that produced by the LIFE Project (Wheatley, Ayris, Davies, Mcleod, & Shenton, 2007) are likely to become increasingly relevant in forecasting costs.

Conclusions

To maximize the potential and investment of digital information, digital curation principles must be exercised throughout the useful lifetime of digital data. This means that curation should be planned for from the outset to ensure longevity and sustainable access. A preservation strategy based on the use of RI depends on a global, well-engineered and distributed network of RI. Coordination and collaboration on a global scale are the hallmarks of such a network. Hence, the proponents of the RRoRI are working with initiatives such as PRONOM and the Global Digital Format Registry; both efforts focus on the provision of details about file formats (structure RI).

The creation of comprehensive RI networks at the necessary level of detail and granularity requires domain expertise to be trustworthy. To harness the power afforded by the Web, simple and automated procedures and tools are required to enable domain experts, whose primary interest may not be in curation and preservation issues, to contribute to the global effort.

It is worth noting that there will inevitably be gaps in the global network of RI, since not all disciplines will be inclined to participate. For example, one of the hardest challenges for collecting RI of common interest to engineers is that much of it cannot

be generated without using real world data, formats and software.

While the business case for using a store of RI is clear, the case for submitting information to it is less clear, since the immediate beneficiaries are likely to be other (possibly competitor) organizations. In order to foster some degree of collaboration in sectors such as engineering and the pharmaceutical industry, a compelling incentive is required.

Acknowledgements

We are grateful to David Giaretta, Steven Rankin and Brian McIlwrath (DCC, CASPAR Project, STFC) for the development of RRoRI and the corresponding ingest tool, as well as discussions relating to RI. We would also like to thank Simon Coles (eBank-UK Project, EPSRC UK NCS) and Chris McMahon (KIM Project, University of Bath) for discussions relating to curation and preservation in crystallography and engineering respectively. The DCC and eBank-UK projects are supported by the Joint Information Services Committee (JISC). The KIM Project is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and the Economic and Social Research Council (ESRC) under Grant Numbers EP/C534220/1 and RES-331-27-0006.

References

Allen, F. H. (2004). High-throughput crystallography: the challenge of publishing, storing and using the results. *Crystallography Reviews, 10*, pp.3-15.

Beagrie, N. (2006). Digital curation for science, digital libraries, and individuals. *International Journal of Digital Curation, Vol. 1 (2006)*. Retrieved July 13, 2008, from http://www.ijdc.net/ijdc/article/view/6

Berners-Lee T., Hendler J., & Lassila O. (2001, May). The semantic web. *Scientific American*.

Cambridge Manufacturing Review. (2003, Summer). Living without production, if we can… . Institute for Manufacturing (IfM): University of Cambridge. Retrieved July 14, 2008, from http://www.ifm.eng.cam.ac.uk/service/cmr/03cmrsummer/productionsummer03.pdf

Coles, S., Frey. J., Hursthouse, M., Milsted, A., Carr, L., Gutteridge, J., et al. (2006). Enabling the reusability of scientific data: Experiences with designing an open access infrastructure for sharing datasets. In *Designing for Usability in e-Science*. (January 26-27, 2006). Edinburgh, UK. Retrieved July 13, 2008, from http://eprints.soton.ac.uk/24059/

Consultative Committee for Space Data Systems. (2002). Reference model for an open archival information system (OAIS). Retrieved June 10, 2008, from http://public.ccsds.org/publications/RefModel.aspx

Davies, A., Brady, T., Tang, P., Hobday, M., Rush, H., & Gann, D. (2003). *Delivering integrated solutions* (Monograph). SPRU-CENTRIM: University of Brighton, UK.

Duke, M., Day M., Heery R., Carr L., & Coles S. (2005). Enhancing access to research data: the challenge of crystallography. *JCDL (2005), Digital Libraries: Cyberinfrastructure for Research and Education*, *Denver, Colorado, USA*. June 7-11, 2005.

Giaretta, D. (2007). The CASPAR approach to digital preservation. *International Journal of Digital Curation*, *Vol 2 (1)*. Retrieved July 13, 2008, from http://www.ijdc.net/

Giaretta D., Patel M., Rusbridge A., Rankin S., & McIlwrath B. (2005). Supporting e-research using representation information. *Proceedings of UK e-Science All Hands Meeting*. Retrieved July 13, 2008, from http://www.allhands.org.uk/2005/proceedings/papers/447.pdf

Giaretta D., Rankin S., McIlwrath B., Rusbridge A., & Patel M. (2005). Representation information for interoperability now and with the future. *Proceedings MSST 2005*, (June 20-24, 2005). International IEEE Symposium on Mass Storage and Systems: Sardinia, Italy, pp. 54-58. Retrieved July 13, 2008, from http://www.soe.ucsc.edu/~elm/msst05/MSST05-Sardinia-Proceedings-2.pdf

Oliva, R., & Kallenberg, R. (2003). Managing the transition from products to services. *International Journal of Service Industry Management, 14(2),* 160–172.

Patel, M., & Coles S. (2007). *A study of curation and preservation issues in the eCrystals Data Repository and proposed federation*. eBank-UK phase 3, scoping report (July 2007).

South Carolina Research Authority. (2006). *STEP Application Handbook, ISO 10303, Version 3*. Retrieved July 15, 2008, from http://www.tc184-sc4.org/SC4_Open/SC4_Standards_Developers_Info/Files/STEP_application_handbook_63006.pdf

US Product Data Association. (1996). *Initial Graphics Exchange Specification (IGES) v5.3*, ANSI. Retrieved July 12, 2008, from http://www.uspro.org/documents/IGES5-3_forDownload.pdf/view

Wheatley, P., Ayris, P., Davies, R., Mcleod, R. & Shenton, H. (2007). *The LIFE Model v1.1.* Discussion paper. London: LIFE Project. http://eprints.ucl.ac.uk/archive/00004831/