Grey literature, institutional repositories and the organisational context

Simon Lambert, Brian Matthews, Catherine Jones

CCLRC

Business and Information Technology Department

Rutherford Appleton Laboratory

Chilton, Didcot, Oxon. OX11 0QX

UK

Email S.C.Lambert@rl.ac.uk, B.M.Matthews@rl.ac.uk, C.M.Jones@rl.ac.uk

Abstract. A wide variety of grey literature is produced during the work carried out at a large multidisciplinary scientific research organisation. This paper examines how the grey literature may be managed and the benefits that result. Trends in Technical Reports, which have always been an important medium for transmission of knowledge, are examined, and the use of an institutional repository is advocated for the future of the medium. Other kinds of grey literature produced in research projects are also described, and again the institutional repository is seen as an important mechanism for preserving and making accessible the knowledge they embody, particularly if it is coupled with other information systems in the organisation.

1. Introduction

CCLRC, the Council for the Central Laboratory of the Research Councils, is one of Europe's largest multidisciplinary research organisations supporting scientists and engineers world-wide. As well as participating in setting the priorities for the UK's science needs, it operates three research laboratories: the Chilbolton Observatory in Hampshire, the Daresbury Laboratory (DL) in

Cheshire and the Rutherford Appleton Laboratory (RAL) in Oxfordshire, and employs 1800 people. These laboratories offer facilities and expertise including ISIS, the world's most powerful pulsed neutron source; high-power lasers; space science technology including satellite and ground-based instrumentation; and information technology.

The operation and development of CCLRC's facilities—which set the organisation apart from most universities—require a high level of specialised knowledge and expertise that has been built up over many years. Additionally, research is conducted by members of staff and users of the facilities. Collaborative projects are undertaken, sometimes of a very large scale: for example, the Space Science and Technology Department was strongly involved in the Mars Express and Venus Express planetary missions. In information technology, there is a history of many years of international collaboration in European research and development projects, involving partners (companies, research institutes and universities) in many countries and producing software, standards and academic publications.

It can be seen that CCLRC is a knowledge-intensive organisation with some special attributes and requirements. Grey literature, as well as literature for publication, plays a key role in its business, and this paper will study that role, link it to the context in which the business is conducted, and highlight the importance of CCLRC's institutional repository in supporting it.

2. Technical reports as grey literature

CCLRC and its predecessors have produced Technical Reports since the organisation was founded. These are formally published by the organisation and are deposited, according to UK law, with the National Deposit Libraries. The format was designed to capture the pre-refereed version of journal articles or to capture technical details for posterity and dissemination. The first point was especially relevant for particle physicists, as rapid dissemination of information was, and still is, important to advance the boundaries of the field. The second is a recognition that not all the

knowledge gained from the scientific process is suitable for publishing in the scientific journal record but is still valuable. As one of CCLRC's main roles is the building and maintenance of large-scale scientific facilities, the organisation has detailed technical knowledge in very specialised fields. Building these facilities does not happen very often, but for example information contained in technical reports for the Neutron Spallation Source (ISIS) are being used twenty years later to aid the construction of a Second Target Station.

However both these roles have been affected by the changes in information dissemination that have occurred with the rise of the Internet. Pre-refereed paper versions have been completely overtaken by e-print repositories (such as arXiv[1]) and the print form is now redundant. Furthermore, the internal processes for producing technical reports have been a barrier in this electronic age and the format has been in decline. Figure 1 shows the trends in the production of formal reports over a ten-year timespan.

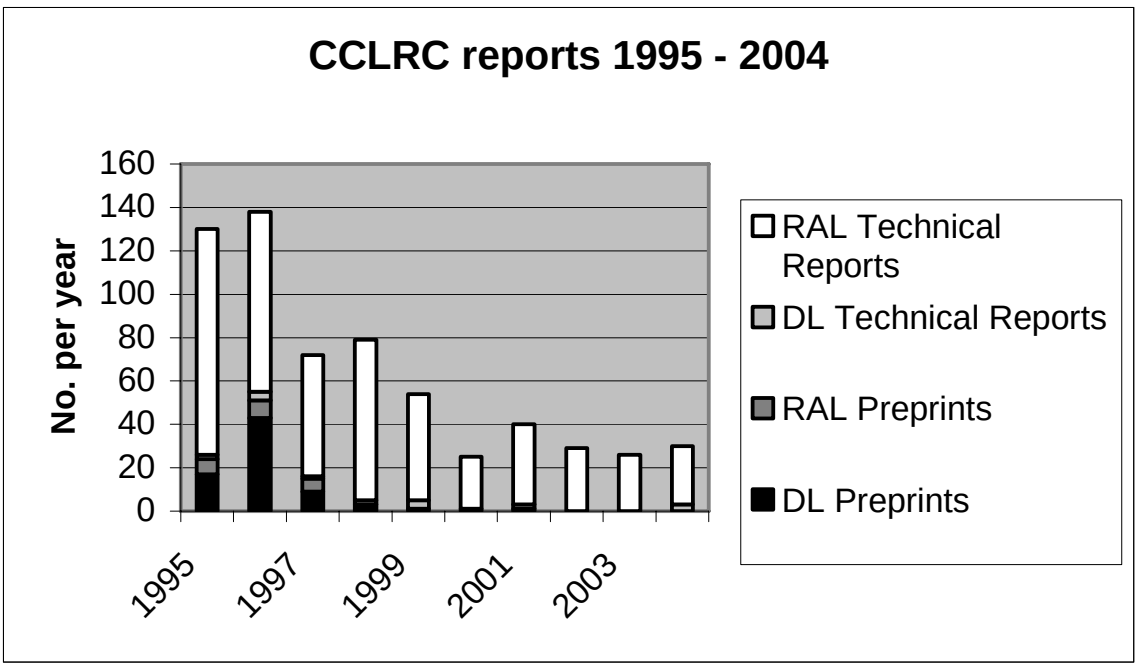

Figure 1: The decline of formal Technical Reports at CCLRC

The production of the technical knowledge has not stopped, but the barriers of the internal process have made it more attractive to produce 'grey' versions of the material. This limits the

dissemination and resource discovery potential of the material, especially if it now resides in a variety of systems such as internal file stores. It can also lose the clarity of versioning: is this copy in location A the same as this copy in location B?

3. Reclaiming Technical Reports through an institutional repository

The previous section discussed some of the reasons that Technical Reports have been encouraged to go grey. It is a contention of this paper that by using an institutional repository the organisation and wider community can 'reclaim' Technical Reports. An institutional repository is a formal and managed archive of research output in the form of digital documents that is operated by a particular institution such as a university. In some cases the focus is only on e-prints[2]—pre- or post-publication versions of refereed papers. However, the remit of CCLRC's institutional repository, known as ePubs[3], is the scientific and technical output of CCLRC; it does not distinguish in its collection remit between formally published or grey literature. It is concerned with the dissemination of the knowledge discovered by CCLRC and the users of its facilities rather than the final location of the information.

By encouraging technical report writers to deposit their material within ePubs, and making the internal process recognising that work as a technical report more efficient and effective, it becomes more visible and useful to the wider community. ePubs is committed to the long term preservation and curation of Technical Reports and by depositing such work in ePubs then its long term future is assured. Further, ePubs uses the IFLA Functional Requirements for Bibliographic Records to represent the work and its version relationships. It is possible to distinguish between different versions to bring clarity back into this area. This is discussed further in section 5 below.

4. Varieties of grey literature

Technical Reports exemplify a type of literature with a potentially long useful life becoming progressively greyer as time goes on. The longevity of this material suggests a refinement to the established definition of grey literature for a research institute like CCLRC. The established definition is: *'Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body.'* [4] It is accepted that this definition needs narrowing for particular contexts, and such a refinement is: *'information produced in a specific working context which is, or might be of value outside that context.'* This additional specification is intended to capture the idea that the material, though not intended for publication in the formal sense, may profitably be 'published' to different working contexts.

There are many other varieties of grey literature produced at CCLRC, but one that is worth singling out is that produced by collaborative R&D projects in the form of deliverable documents, working reports, etc. Normally the distribution of these documents is specified by the contract under which the project is conducted, but it will usually be the case that some of the documents have unrestricted distribution. Furthermore, they embody valuable knowledge that might be applicable in wider contexts. For example, final reports of one project may serve as input to the baseline for the next project, or surveys of products or state of the art produced for internal decision making may retain their value for some time and have a wider applicability beyond the project that created them.

Figure 2 shows a schematic (and incomplete) depiction of parallel processes in an organisation such as CCLRC, and the ways in which grey literature is generated and consumed among them. These processes are represented as coarse-grained work flows, which, while not constraining the activities of scientists or managers, describe the stages that they pass through in a particular project.

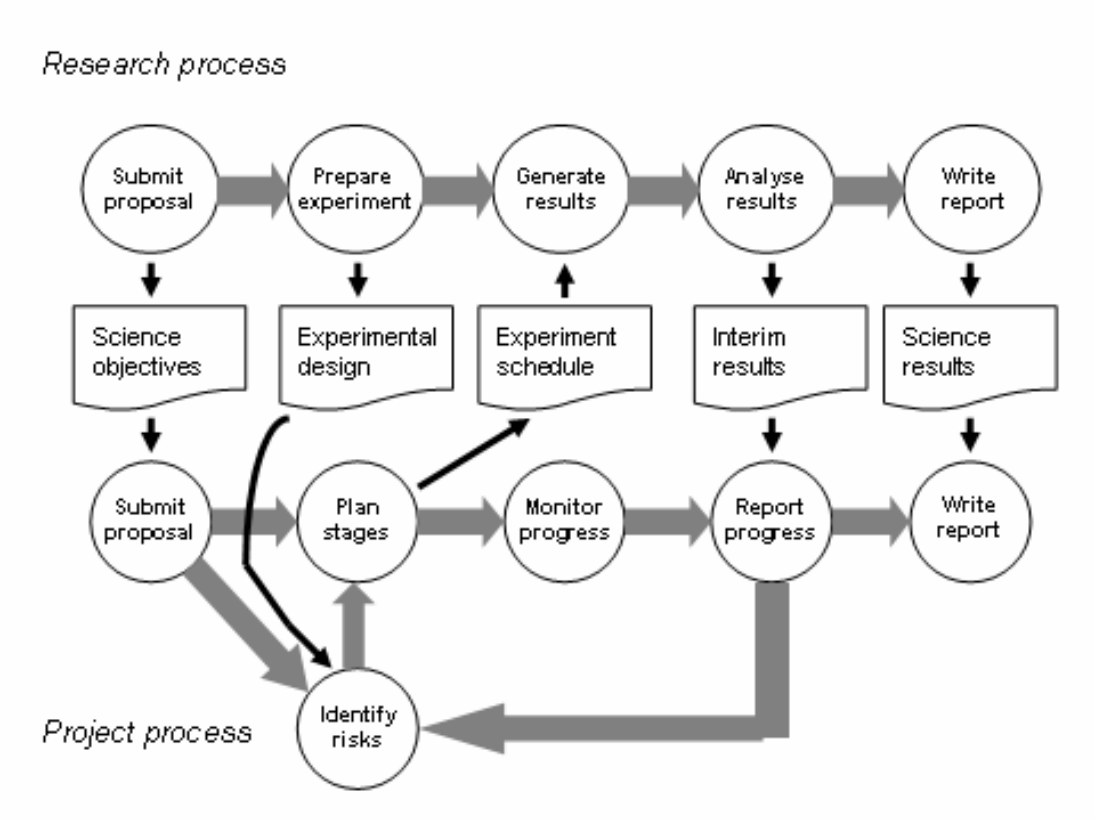

Figure 2: Exchange of grey literature between processes

The research process is what drives the production of scientific knowledge, and interlocked with it is the project process, which can be understood as part of the organisation's *business process* since it is concerned with receiving and using resources with which results are returned and disseminated.

These processes interlock in the sense that outputs from one stage serve as inputs to another stage in a parallel process. The analysis of results feeds into the progress monitoring; the scientific report forms part of the final project report; and so on.

It is also possible to characterise in a similar fashion other processes such as the digital curation process[5] and the publication process.

Inasmuch as the output of the stages is recorded in some form, it is by definition grey literature. As noted above, an additional specification may be introduced that the information is, or might be of value outside the context in which it was produced. This can mean two things:

- the information is useable elsewhere in a parallel process (for example, transfer from the research process to the project process);
- the information is useable in a different (later) process instance (for example, in a different project).

Figure 3 shows a variant of the diagram in which the research process is now concerned with collaborative development of software technologies (as opposed to the experimental science represented in Figure 2) and grey literature is generated that may be of value in other such projects. The broken arrows indicate such feeds into the future.

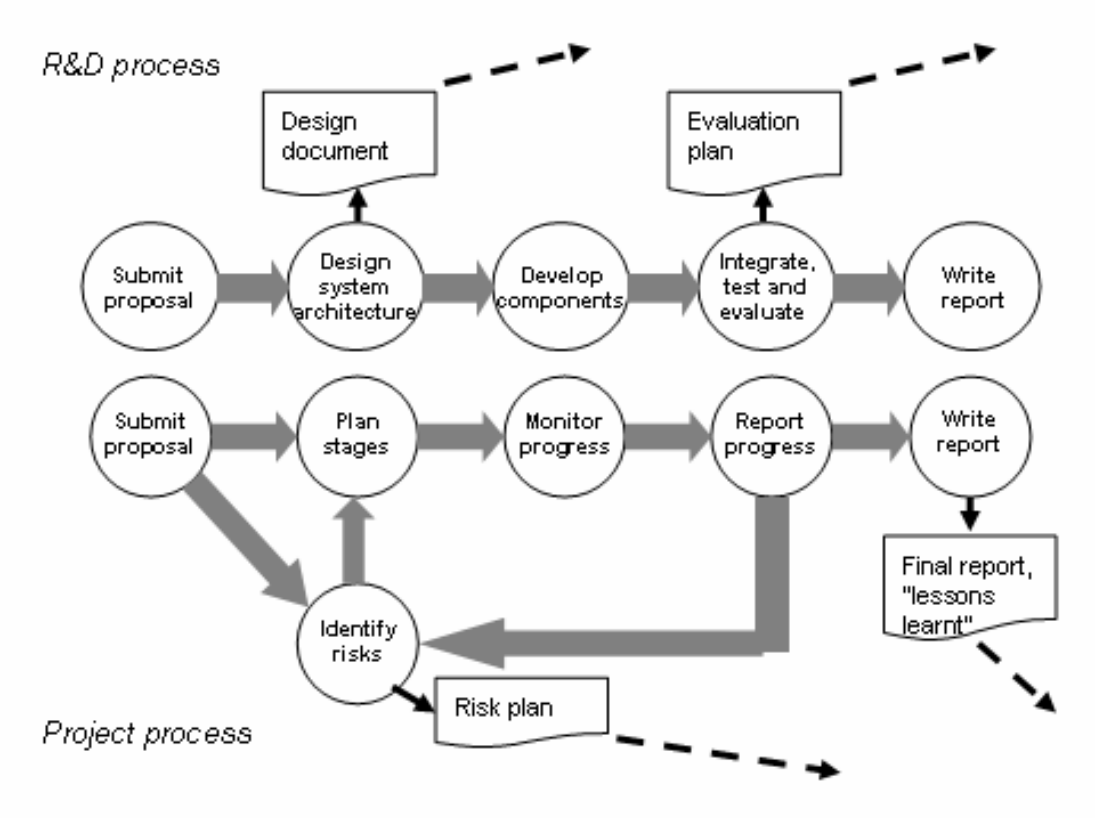

Figure 3: Grey literature available outside the immediate processes

5. Grey literature and the institutional repository

It is a contention of this paper that some of the grey literature produced within the organisation may be profitably stored in an institutional repository, so as to enable the kind of reuse

outlined above. Particularly relevant are the institutional repository's capacity for long-term storage and for clarity in versions of material submitted. At CCLRC a programme of digitisation of Technical Reports is under way, with the ePubs repository playing a central role in storing and indexing the documents. This is part of the effort to have ePubs as a long-term archive for curation and preservation of the research output of CCLRC. The deposit of some project documents such as final reports is encouraged.

Regarding versions, ePubs partially implements the IFLA Functional Requirements for Bibliographic Records[6] model. This model specifies four levels of description for any particular work. The *work* itself is an abstract concept of a distinct intellectual or artist creation. A particular work can be realised as an *expression*, or series of *expressions*. Each expression will be physically embodied as a *manifestation*, or series of *manifestations*. A single exemplar of the manifestation is known as an *item*. ePubs implements the Work, Expression and Manifestation elements at present. Figure 4 shows an example of different expressions of a single work: in this case, a paper as published in conference proceedings, and the PowerPoint presentation that was actually shown at the conference.

Publications

Publication Type	Reference	URI	Local file(s)	Year
Presentation	*9th DELOS Network of Excellence thematic workshop: Digital Repositories: Interoperability and Common Services, The Foundation for Research and Technology - Hellas (FORTH), Heraklion, Crete, 11-13 May 2005*	http://www.ukoln.ac.uk/ev...	jones-mascord.ppt [Microsoft Powerpoint, 707 KB] [42 accesses]	2005
Paper in Conference Proceedings	Proc. *9th DELOS Network of Excellence thematic workshop: Interoperability and Common Services, Heraklion, Crete, 11-13 May 2005*		delos-jones-20050425.pdf [Adobe PDF, 134 KB] [19 accesses]	2005

Figure 4: An example of different expressions of a single work in ePubs

Of course storage of the grey literature is only one half of the problem. It is also necessary to retrieve it effectively according to context. There can be no substitute for human expertise here, but

the expertise can be augmented by suitable metadata describing content and the working context in which it was produced. Another contention of this paper is that the institutional repository could and should be integrated with other information systems of the organisation, so as to provide some of the context needed for retrieval. These systems include:

- personnel systems;
- financial systems;
- project management/monitoring systems;
- security infrastructure.

Such integration can offer benefits including accuracy of information (for example, allocation of publications to individuals), quality control (ensuring that all necessary stages are met), efficiency of metadata collection (for example, collecting metadata at project application stage and tracing through), adherence to policies and procedures (for example, approvals for paper submission), traceability and accounting, and report generation (for example, for performance assessment)

Integration of this kind will require integration of repository metadata with metadata associated with the other systems (HR-XML, CERIF, PRINCE 2 standards, …). Moreover, to assist with retrieval, it is also necessary to describe the stage in the processes at which the document was produced, and to characterise those processes themselves in terms of the domain (scientific specialism). There are two dimensions of similarity of context: the content itself (for example, documents relating to data mining or to neutron scattering spectrometers) and the stage in the process (for example, a risk assessment, or a final report). The ePubs system currently supports the first of these through optional keywords associated with entries, though not the second, which has clear links to the project process.

6. Conclusions

An institutional repository, being a central point within the organisation for literature and data, is a component of the integration of processes, which promises benefits both to the organisation itself and to the researchers within it. Grey literature can be recorded and retrieved according to accurate and up to date personnel, project and business unit structure, leading to benefits in accuracy of reporting, quality control, etc. From the researchers' point of view, the repository assists them by storing multiple expressions or manifestations for the different parts of the process, and so brings clarity to work and version relationships. It also allows retrieval of grey literature from other projects that might otherwise be invisible to those who could benefit from it. In short, the institutional repository is integrated as a part of the overall institutional memory.

References

1. arXiv.org e-Print archive, http://www.arxiv.org
2. Supporting Open Access, http://www.eprints.org
3. Jones, C., Mascord, M.(2005), 'Experiences of building an Open Access Institutional Repository in a UK Scientific Research Organisation', Proc. 9th DELOS Network of Excellence thematic workshop: Interoperability and Common Services, Heraklion, Crete, 11–13 May 2005. To access the ePubs system see: http://epubs.cclrc.ac.uk
4. Definition of grey literature at http://www.greynet.org, accessed November 2005.
5. Open Archival Information System (ISO 14721:2003): Reference Model for Data Curation
6. IFLA Study Group (1998), *Functional Requirements for bibliographic records: final report*, Saur, Munich, http://www.ifla.org/VII/s13/frbr/frbr.pdf