



# RepoMMan Project

---

Deliverable R-D4

Report on research user requirements interview data

Richard Green

November 2005  
Version 1.1 January 2006



## The RepoMMan Project

<b>Project Director:</b>	Ian Dolphin, Head of e-Strategy, University of Hull (i.dolphin@hull.ac.uk)
<b>Project Manager:</b>	Richard Green (r.green@hull.ac.uk)
<b>Technical Lead:</b>	Robert Sherratt (r.sherratt@hull.ac.uk)
<b>Repository Domain Specialist:</b>	Chris Awre (c.awre@hull.ac.uk)

The Repository Metadata and Management Project (RepoMMan) at the University of Hull is funded by the JISC Digital Repositories Programme. The project is being carried out by the University's e-Services Integration Group (e-SIG) within Academic Services.

## Introduction

The RepoMMan Project Plan requires the development of a front-end interface to the Fedora repository software for use by researchers working on their own behalf or collaboratively. Clearly, before work on such an interface can progress very far it was necessary that the development team understand what it is to "do research" in the academic sense.

## Methodology

The work is being informed by a two-pronged investigation of research methods. The first approach was to interview a small number of University of Hull researchers at length about their working practices, the second approach was to ask similar questions of a wider (inter)national audience using an on-line survey. From the data thus collected generalised scenarios and use cases can be developed. This document focuses on the interviews conducted with researchers. A later project document will bring together the data from the interviews and the on-line survey.

## The interviews

Six researchers from the University of Hull were interviewed at length, taking between 75 and 90 minutes each. The outline structure of the interviews is described in project document R-D2 "Criteria for research user interviews". The researchers were drawn from five different disciplines, Biology, Law, Business, History, and Medical Research in an attempt to get a variety of viewpoints on what it is to 'do research'. In the event, there is much commonality between their approaches.

The main part of the interview concerned researchers' current practice and their responses therefore focus largely on the process of developing a journal paper rather than on other forms of publishing that may be available in the future. Some of the interviewees were aware of the possibilities offered by on-line publishing and of the possibility of publishing other forms of research output - for instance, supporting datasets. (Indeed, one interviewee was very positive about the benefits of (rapid) electronic publishing in his fast-moving field.) The initial stages of the RepoMMan project seek to understand and cater for current research practice before developing support for a wider range of methodologies and research outputs.

## The research process

The research process starts with an idea. The researcher then spends some time checking that the idea is one worthy of further investigation by reading around it. The 'reading around' process is elaborated later in this narrative.

Depending on individual circumstances it may be that the researcher needs a grant to support his or her work. That being the case, a likely funding body must be identified. Depending on their guidelines a funding proposal will be developed and submitted. Depending on the outcome of this request, the research may go ahead, another funding body may be approached or the idea may be abandoned.

Once the research starts in earnest a period of development ensues during which the idea and/or the data is developed. The data may be developed 'in-house', perhaps by scientific experiment, or it may be obtained elsewhere, as in the case of written primary sources. This process is generally accompanied by a period of reading around the topic, looking at other's work in the area. In the specific area of medical research involving patients, there is an absolute requirement for specific, detailed record-keeping.

Location of remote data and relevant papers is increasingly done through the internet. The searching process may be accomplished using a range of tools from the generalised (eg

Google), through the general academic (eg Google or Google Scholar) to the highly subject specific (eg Westlaw in the case of legal texts). More often than not it seems that researchers download the materials thus located and store them locally in digital form, although some also print them, finding the paper format easier to work with. Little mention was made of the older method of locating a journal article in a library and photocopying it for reference. The researchers interviewed all made some attempt to save these digital materials in a structured way although this varied from quite a simple structure to something that was approaching a semantic web.

The material thus gathered, synthesised and developed eventually becomes a draft paper. The process of generating the draft may well precipitate more investigation but eventually a document will be produced that the researcher wishes to share for comment. As the paper develops all our researchers had methods for keeping successive versions of the file although the frequency with which this was done varied. Likewise, all our researchers had some method for ensuring periodic backup of their works.

The process of sharing a draft paper for comment can take a number of forms. There is the specific case where a paper is being developed collaboratively with other researchers; this raises some specific issues which will be dealt with separately. Generally the process involves inviting others in the field to comment on the draft, a process which may be initiated in a number of ways, some manual and some electronic, and which may likewise be accomplished in a number of ways including telephone discussion, handwritten comments and digital in-document comments. Unusually, only our researcher from the Business School mentioned this, the draft paper may be turned into a conference presentation and its ideas tried out on peers in that way.

The process of sharing with collaborators follows the general sharing process described in the previous paragraph but it raises an additional issue related to versioning. If a draft paper is being comment on it seemed to be agreed amongst the researchers who commented that the primary copy must be untouched during that process and that, if multiple collaborators are involved, the process of commenting should be a serial one and not parallel. In this way the primary document is protected against possibly conflicting modifications.

Following on the comment from peers, the document may well be revised - a process which may involve further development of ideas and/or data and more reading around - but eventually it will be in a form that the author(s) consider fit for publication.

The document is now submitted to a journal where it may be accepted in principle or rejected. If rejected (not something that our researchers had experienced, in general) it is presumed that another journal would be approached. The journal will generally submit the article to referees who, together with the editor, are likely to have their own suggestions for revision. In addition, the editor may have views about the length of the submission.

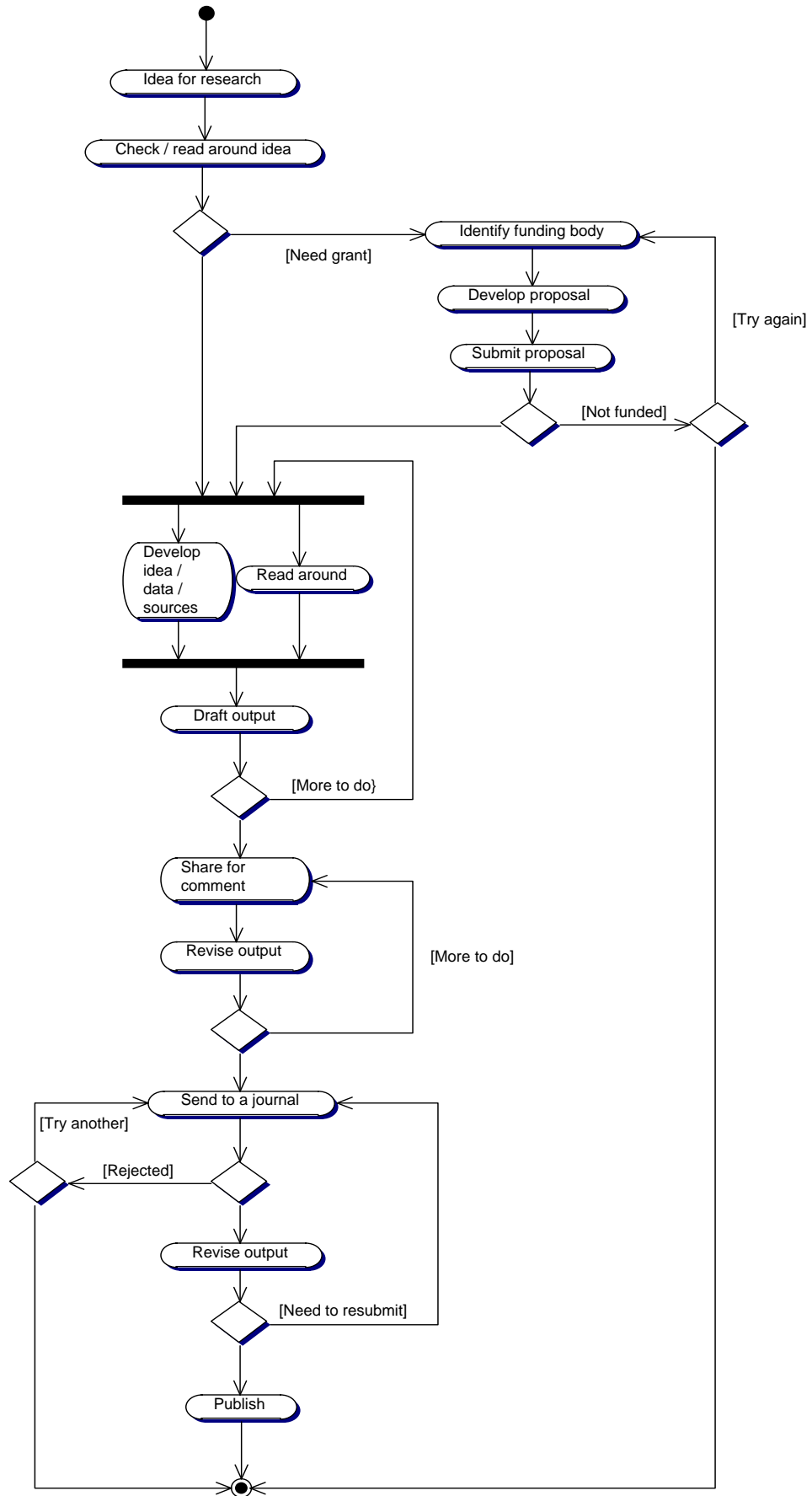
Again a process of (possibly cyclical) revision ensues until the paper is eventually accepted for publication. The author then has little further involvement in the process other than to agree a copyright position with the publisher; something that may apparently happen without much thought.

All our researchers felt that they were in a position personally to sign some sort of copyright agreement (notwithstanding the University's official position on IPR). Most were happy to accept the publisher's standard copyright contract (although they did not read it closely) with only our researcher in Law taking a keen interest in, and potentially challenging, its provisions.

At the end of a particular research project our researchers generally kept the materials that they had downloaded.

## **Activity diagram**

In outline, the process outlined above can be represented in an activity diagram such as that shown overleaf:



## Scenarios

From the interview transcripts, it is possible to reduce the researchers' needs to five scenarios:

Steven uses a range of specialised search tools to inform the development of his research papers. He organises the materials that he finds in a highly structured manner using sophisticated cross-referencing and he uses advanced indexing software to help him find references on his computer's hard drive. When his draft is far enough advanced he consults colleagues at Hull and further away who comment using pen or Word 'track changes'. When the article is finished he submits it to a publisher electronically. During the development of the paper Steven starts a new copy of the file each time there is a significant structural alteration; he takes periodic backup copies on CD. At the end of a research project he retains all his research materials. Steven generally signs any copyright agreement with his publisher without looking at its detail.

Tony uses a small number of search tools to inform the development of his research papers. He organises the materials that he finds in a simple structure. As he develops a paper Tony uses version numbers which advance each time he makes more than a trivial change; he normally has at least one backup and usually two. He does not generally share his work until it is at an advanced stage at which point he presents his thoughts at a conference to obtain feedback. When his paper has been published Tony retains all his research materials. Tony generally reads and signs any copyright agreement with his publisher but has not considered the implications of this for any 'personal' publishing that he may wish to do.

Charles uses general and subject-specific search tools to inform the development of his research papers. He keeps the materials that he finds in an ordered manner on his computer but frequently prints them in order to work with and annotate the text. Once the paper is well advanced he may share it with colleagues elsewhere for comments which are generally made using the Word 'track changes' facility or else by telephone. As he develops a paper he uses version numbers which advance each time he makes more than a trivial change; there is always at least one backup of the current version. At the end of a research project he may keep his research materials on disk for a time but is more likely to print them out and file them. Charles carefully reads any copyright agreement with his publisher and may challenge its provisions if they do not suit his purpose.

Darren uses a small number of search tools to inform the development of his research papers but increasingly tends towards Google Scholar as his tool of first choice. He keeps the materials that he finds in an ordered way. As he develops a paper he shares it with colleagues elsewhere for comment which is generally done using the Word 'track changes' facility. Every time he alters the developing paper he gives it a new filename which includes the date; there are always multiple copies for backup. At the end of a project he makes his paper available via the Departmental website and would like to be able to provide accompanying data. He retains his research materials on disk in a structured way but also keeps printed copies. Darren reads any copyright agreement with his publisher and signs it but believes that there is generally a private understanding that he will also post a version of his paper on the Departmental website.

The last scenario is similar to the third although, being about medical research, there is an important addition about record keeping.

Peter, a medical researcher, uses general and subject-specific search tools to inform the development of his research papers. He keeps the materials that he finds in an ordered manner on his computer but frequently prints them in order to work with and annotate the text. Peter develops a lot of his work using empirical research data about patients and is subject to strict rules about the way this data is handled, stored and retained. Once the paper is well advanced he may share it with colleagues elsewhere for comments which are generally made using the Word 'track changes' facility or else by telephone. As he develops a paper he uses version numbers which advance each time he makes more than a trivial change; there is

always at least one backup of the current version. At the end of a research project he may keep pdfs that he has downloaded as background on disk for a time but is more likely to print them out and file them. Records and data relating to his own research are carefully preserved in their original paper form. Peter usually reads any copyright agreement with his publisher and has a general idea of its provisions.

## Basic requirements

There is a great deal of commonality in the way that each of our researchers works and so it is relatively straightforward to draw out the requirements that their current practice would have of a repository which supports the process as well as the end products of research:

- we take in as a *sine qua non* that a repository interface should not make it difficult to do something that is currently achieved easily
- the repository interface must allow structuring of a user's personal storage space and have the capacity to hold potentially large numbers of objects, possibly of a range of differing types, for each user
- the repository should provide an easily usable versioning facility (it must be easy to version a file *and* to revert to an earlier version)
- the repository should allow sharing of a private document with a closed group of collaborators and should provide some sort of locking facility so that conflicting revisions cannot occur
- the repository must make public exposure of content easy and controllable, taking account of digital rights issues as part of that process

These requirements are an initial set that would support the current practice of a small number of researchers; when the results of our on-line survey are known they may need to be modified to take account of the way that other researchers work.

Once these basic requirements have been dealt with the project will consider what 'added value' facilities might be provided for researchers and will consider how best to support an effective workflow which includes, amongst other things, the necessary stages of metadata population and the grouping of related research outputs.



## **References**

ASK Project: [http://tadpole.oucs.ox.ac.uk/repowiki/index.php/Main\\_Page](http://tadpole.oucs.ox.ac.uk/repowiki/index.php/Main_Page) (Validated October 2005 - RG)

JHU project: <https://wiki.library.jhu.edu/display/RepoAnalysis/ProjectRepository> (Validated October 2005 - RG)