Increasing Repository Content through Automation and Services (IncReASe)

1. Introduction

1. This proposal is submitted by the Universities of Leeds, Sheffield and York – as the White Rose University Consortium. The three institutions have jointly developed a cross-institutional ePrints repository of research outputs – known as the White Rose Consortium ePrints Repository. The initiative began life under the JISC-funded SHERPA Project.
2. The repository is available at http://eprints.whiterose.ac.uk, and it utilizes the ePrints software.
3. The repository has achieved substantial success and impact within the three institutions. The repository currently holds over 1600 outputs, which compares well with other similar institutional repositories. Over 94% of content is available in full text, a large majority being peer reviewed post-prints. The Registry of Open Access Repositories (ROAR) lists the White Rose Consortium ePrints Repository as having one of the highest percentages of full-text content in the UK.[1]
4. The White Rose Consortium is a founder member of the ePrints Community[2], and is an active voice in the community. Current community activities include participation in the alpha and beta testing for ePrints v.3 (due early in 2007).
5. The partners are all committed to the implementation of institutional-level open access policies, and some progress has already been made with this within the institutions. Sheffield and York have already agreed policies which 'strongly encourage' deposit. In each case, these have been agreed at senior research committee level. Leeds is also close to a similar agreement being made. An example from Sheffield is available online.[3]
6. The repository has achieved a strong, steady growth rate; however, we recognise that the current content reflects only a very small percentage of the total output from the three institutions. So far, we have employed a strategy of 'mediated deposit' – that is, deposit is mediated by a full-time Repository Officer who is engaged in advocacy activities, encouraging academic staff to provide their content to her. Content is then uploaded on their behalf. The repository has so far only been able to identify a very small number of willing self-depositors amongst academic and research staff, and self-deposit itself can lead to specific problems with metadata, file versions and related issues.
7. The rationale for this Project is based on the experience of the wider repository community, which has indicated that filling repositories is proving particularly challenging, coupled with our own extensive experience with researchers across the three institutions. Our experience has indicated, several times, that, though there may be support in principle for open access and deposit in the repository, many academics are reluctant to spend time archiving their work – particularly if they feel that they are being asked to duplicate effort where their work is already available through their own web pages and where they have already been asked to input metadata about their publications in several places (for instance, the University publication database, a departmental publication list, information held by the Funding Councils). There are often problems with this approach – in particular, it provides a disjointed view of the overall institutional output as content is often scattered around a large number of unrelated sites, making it difficult to provide a single, joined-up view, thus reducing overall impact. Problems also emerge with searching across these distributed collections; they are rarely OAI compliant, and so are also unlikely to be picked up by national and international OAI services.
8. The Project proposes to explore, investigate and pilot an automated approach to the gathering, indexing and presentation of this content in the repository, using techniques and technologies for the automated acquisition, harvesting and metadata generation from document collections held in distributed environments. We will focus on an underlying principle of 'input once, use many', as the key rationale for the Project. The work will be undertaken in close collaboration

[1] http://archives.eprints.org/

[2] http://www.eprints.org/community/

[3] http://www.shef.ac.uk/content/1/c6/03/25/85/OpenArchiving.pdf

with researchers and academic departments. The relationship of the repositories with pre-existing publication databases will also be addressed. The Project will review whether repository deposit can be located within existing workflows. This will support the sharing and re-use of content, provide a sustainable route for achieving a critical mass of content in the repository, and, ultimately, improving the effectiveness of the repository as a tool for dissemination of scholarly communication.

9. The Project will also look at other output mechanisms – in particular in relation to the recent Research Council (RC-UK) statement on open access. The Project will consider the relationship between institutional and Research Council (funder) repositories, and establish a best practice model for workflow, export and harvesting by or from funder repositories. The Project will also work closely with researchers to assess what types of repository output will make the repository a more attractive and informative research tool, with a view to increasing researcher 'buy-in' to the system.
10. Appendix 2 provides an overview of the workflow and relationships between the repository and other systems. It also identifies a potential best-practice workflow model, which the Project will seek to build. The figure illustrates how content will be harvested from various internal sources, checked against external resources and organised within the repository buffer before being made live. The solid connectors trace routes into the repository. The dashed lines show ways in which the repository content may be fed into other internal and external systems.

1.2 Project Objectives

11. In summary, the Project will:
    - Explore mechanisms for the automatic harvesting of metadata and full-text content from a range of external sources, including departmental websites and databases and personal home pages.
    - Test and pilot a number of mechanisms for the automatic population of repository content and metadata, and make recommendations for the consequent impact on institutional roles in relation to support for repositories.
    - Where appropriate, organise ingested material by publisher to take advantage of economies of scale; publisher to be identified by automated journal title checking with the RoMEO database
    - Review the relationship between institutional repositories and national and subject repositories, and explore the workflow implications for the population of Research Council repositories in particular.
    - Explore issues for academic and research staff around the research and publication lifecycle, and make recommendations for the optimal point at which research outputs should be deposited in both subject and institutional repositories.
    - Facilitate a transition from uncoordinated deposit across distributed network to a coordinated centralised deposit model for the White Rose Institutions.
    - Offer outputs from the repository to feed local systems such as departmental and individual web pages, and Portals.
    - Offer customized statistics on repository growth and usage that can be tailored for individual, departmental and institutional level needs.

1.3 Key Project Outcomes:

12. The key Project outcomes, and benefits to the partners, will be:
    - Overall improved management of institutional content and intellectual output.
    - More effective integration between the repository and local systems and practices across the three White Rose institutions.
    - Increased understanding of the relationship between subject and institutional repositories – especially in relation to the emerging Research Council repositories – how this relationship will work in practice, and how duplication of effort can be avoided.
    - Clarification of roles and responsibilities in supporting the repository at an institutional

level.

- Increased content available in the repository with a consequent positive impact on overall credibility of the repository within the institutions.

1.4 Benefits to the JISC community

13. Key benefits for the JISC community will include:
    - A best practice model for the management of workflow within an institutional context – which will be applicable to the wider JISC community as a model which could potentially be adopted and used by others.
    - A best practice model for managing the relationship between institutional repositories and funder repositories – which could be adopted and used by others.
    - Software and integration tools which could be used and adapted by others to facilitate the automated population of repository content, and integration with other campus systems.
    - Comparison reports of the issues raised by and the effectiveness of different harvesting methods.

1.5 Strengths and Experience of the Consortium

14. The University Libraries of the Universities of Leeds, Sheffield and York are already collaborating on a number of initiatives – including the development of the White Rose Consortium ePrints Repository, which was originally developed under the JISC-funded SHERPA Project, and is now being taken forward as a fully supported service by the three institutions. All three partners are working toward institutional-level policies on open archiving.
15. We have built a strong and effective partnership based around this initiative – including the ongoing support for a Repository Officer post shared across the three institutions. We have fostered a proactive approach to working in partnership through shared ownership of both the strategic and operational development issues for the repository, and co-ordination of all activities through a Repository Steering Group which meets on a regular basis to plan for developments. The White Rose Consortium has also been active in repository developments on a national level; and has provided a substantial level of input into other related developments – including the JISC-funded SHERPA-DP and STORE Projects.
16. In the context of the Project, the three institutions each have differing local practices and circumstances. For example, only one of the partners has a centralised publication database. This will help us to identify a range of issues and solutions which may be applicable across different institutions.
17. The White Rose Consortium ePrints Repository has a strong growth rate, adding 50-60 full text items each month. The repository holds in excess of 1600 full text items and has achieved over half a million hits and 150,000 full-text downloads over the past two years. Most referrals to the repository come through external search engines such as Google, Google Scholar and Yahoo. The repository holds research outputs from across many of the Schools and Faculties in the White Rose universities; we have made efforts to engage with social science and arts and humanities disciplines as well as science, technology and medicine. Approximately 22% of our content is from arts, humanities and social sciences.
18. The repository has been actively promoted, and awareness of the repository at senior levels in the consortium is high. We have established contact with a variety of academic departments, and are therefore in a good position to move forward with repository development and to actively engage with the research community in the three institutions to inform and guide the Project.
19. The University of Leeds is also the lead partner in the JISC-funded EVIE Project (funded under the Virtual Research Environments Programme). EVIE has identified a research lifecycle, which will be used as the starting point for mapping workflow relationships. Leeds has also built up effective and close working relationships with researchers, which will provide an additional strength for the Project.

2. Project Description

Work-package 1: Project Management

20. This work package will provide overall management of the Project and ensure that the outcomes are shared widely. Key deliverables include the production of project management documents including work-plans and targets, monitoring progress against objectives, managing the budget, planning and implementation of dissemination activities.
21. Timescale – entire duration of the Project.
22. Resources - Project Advisory Board, Project Manager, Project Officer.

Work-package 2: Survey of existing activities

23. The Project will survey the range and extent of existing sources of metadata and full text content currently available through the institutional websites. This will include capturing and mapping the extent to which departmental and personal websites are being used as 'personal repositories' at present, identifying other sources of information – such as departmental publications databases, and institutional-level publications databases. There may also be records of content held at an institutional level – eg: for RAE purposes. The Project will develop an overall picture of key practices in this area, which will be used to inform the scoping of Work-package 3 (see below). The work-package will also provide a mapping of existing workflow practices – eg: the point at which academic staff (or others) will usually deposit their outputs or metadata on a website or departmental or institutional database. We will establish the triggers for deposit, how the process is handled, and by whom – and from this identify the key opportunities within the workflow process where content could be solicited from a potential depositor.
24. Timescale – project months 1-4.
25. Resources – Project Manager, Project Officer.

Work-package 3: Harvesting of data

26. The Project will pilot the harvesting of content and associated metadata from a range of the identified sources – including departmental websites and publication databases, institutional publication databases, sources of RAE data etc – in order to establish the most effective mechanisms and circumstances for doing this.
27. The Project will utilize and exploit new tools and features available in ePrints v.3, which is currently in beta form.
28. As part of this work-package, the Project will produce and test a set of software tools which can be used to undertake the harvesting. These will be tested across a range of content from the three institutions. We will develop a set of scripts which can be used to parse specified web pages for possible citations. These will then generate lists of citations which can be automatically checked against external sources (eg: RoMEO) in order to ensure copyright compliance before deposit in the repository.
29. For dynamically generated web pages (eg: from departmental publication databases), we will explore export via conversion into XML, and bulk import into the ePrints database. We will utilize new developments supporting bulk import and export, which will be available in ePrints v.3.
30. We will also explore the utilization of tools produced by other projects – for example, the Daedalus Project, which has produced a tool for importing content from Reference Manager databases[4], and the extent to which this could be used in the manner described above.
31. We will also explore the extent to which content and metadata can be harvested from 'personal repositories' – eg: personal network drives or shared departmental drives. In this context, we will explore the use of WebDAV, as well as other possible tools and approaches.
32. We will also seek to provide enhanced user services, such as ability to specify pages from which to upload at the point of user registration, set preferences for automated or mediated harvest from URLs, and automated feedback to users.
33. Software tools and scripts will be made available to the wider community.

[4] https://dspace.gla.ac.uk/handle/1905/175/

34. Timescale – project months 3-18.
35. Resources – Project Manager, Project Officer, Technical Officer.

Work-package 4: Metadata enhancement

36. The Project will explore the requirements for metadata enhancement in order to enable content and metadata to be transferred between institutional and Funder Repositories. We will work with the ESRC Repository as a test-bed in order to establish likely baseline requirements in this area.
37. Certain key metadata fields – eg: funder information – will be implemented in order to facilitate the import and export of content between repositories. A field to capture funder information currently exists in the University of Leeds Publication Database, and may be stored in other central databases at Sheffield and York. However, in the University of Leeds database, this field is rarely used. We will explore the extent to which centrally held funding information (eg: in Research Support Units) can be matched with eventual outputs, in order to identify where a particular output should be deposited, according to the terms and conditions of the Research Council grant. We will also investigate the use of the JULIET database in this context[5].
38. We will also explore a range of other key metadata enhancements which would facilitate the automatic population of content. There are particular workflow benefits and economies of scale to be had through enhancement of metadata – in particular, this will assist with better management of the ePrints submission buffer, so that content can be organised by publisher, facilitating easier automatic checking against the RoMEO database[6]. Other enhancements might include – implementation of ePrints Application Profile and automatic population of DOI field, utilizing CrossRef[7].
39. Timescale – project months 3-18
40. Resources – Project Manager, Project Officer, Technical Officer.

Work-package 5: Integration with Funder Repositories

41. In light of the Research Councils-UK published position on Open Access, the ESRC Awards and Outputs Database (AOD) has recently been extended (October 2006) to incorporate OAI-PMH.

42. The Awards and Outputs Database (AOD) contains records from ESRC's managed information systems on individual awards going back as far as 1975, and where possible also includes any outputs that are generated during the course of funding, and beyond. Award Holders are automatically reminded (on an annual basis for the first five years after funding ends), to keep their information updated, as, quite often, outputs only arise several years after the research itself.

43. ESRC's OA vision is to support social science research activity by presenting the AOD as a subject repository which is interoperable with institutional repositories. In particular *ESRC Society Today*, as a subject repository, will provide ESRC award holders with an easily accessible means for preserving their e-prints, or aim to provide the necessary linkage facilities if they are depositing materials in another OA repository.

44. A second development project is underway to specifically harvest and link materials that have arisen from ESRC funding - with a pilot for this querying the D-Space repository at the Teaching and Learning Research Programme.[8] Once this has been successfully integrated, ESRC will be keen to target other repositories used by its award-holders, with a view to incorporation/linkage.
45. IncReASe will work with the ESRC Repository as a test-bed to establish best practice for linking

[5] http://www.sherpa.ac.uk/juliet/

[6] http://www.sherpa.ac.uk/projects/sherparomeo.html

[7] http://www.crossref.org/

[8] http://www.tlrp.org/dspace/index.jsp

between institutional and Funder Repositories. We will establish a pilot integration which will enable content and metadata to be transferred between the two systems in the most appropriate way.

46. In establishing the pilot system, we will liaise with the Deposit API[9] work which has been undertaken as part of the JISC Digital Repositories Programme. Of particular interest in this context is likely to be the work on SRW Update[10], which may provide an opportunity for developing services for automatic harvesting of content between repositories. We will explore this work and make recommendations as to it's relevance in this context.
47. With regard to the White Rose Consortium ePrints Repository, the ESRC will undertake to ensure its OAI-PMH gateway remains readily available, and is willing to liaise on areas in which this interface could be improved for greater good (whether through manipulation of meta-data, or through functional means).
48. Timescale – project months 5-18
49. Resources – Project Manager, Project Officer, Technical Officer.

Work-package 6: Integration with local systems

50. The Project will seek to integrate the repository with institutional Portal systems at the three institutions. Currently both the University of Leeds and the University of Sheffield have implemented the Luminis Portal, developed by Sungard. Luminis is built around open standards and will adhere to the JSR-168 standard, thus making this work of interest to the wider JISC community. The University of York is currently considering procurement/development of a Portal.
51. Informed by the outputs of WP1, the Project will look at how repository content can be offered back to individuals and departments so that it becomes the central place of deposit, feeding content back to local systems.
52. Integration will require the repository to utilize central authentication systems at the three institutions (likely to be LDAP).
53. We will also investigate the development of a portlet interface for the repository which would enable it to be directly surfaced within a Portal environment. This will enable the repository to be better embedded in the working practices and routines of academic staff. We will be able to demonstrate the transferability of this work through working across the three institutions. This work will draw on and utilize the Deposit API work undertaken as part of the JISC Digital Repositories programme, in particular, in relation to proposals from the ePrints development team for a Web Services interface to ePrints[11] (originally planned for early 2006).
54. As part of this work-package we will also explore provision of statistical data. Our experiences show that there is significant interest from individuals about usage of their own material. The project will investigate methods of capturing and presenting repository usage statistics.
55. Timescale – project months 5-18
56. Resources – Project Manager, Project Officer, Technical Officer.

Work-package 7: Evaluation and dissemination

57. The Project will produce an evaluation and dissemination plan, which will ensure that key success criteria have been met, and will rigorously test the outputs and establish their impact on the user community. An overall evaluation will be incorporated into the project final report.
58. Dissemination activities will take place to alert the wider community to the value of the software tools produced, and to encourage their take-up and use. Where possible, the Project will also seek to join-up dissemination activities with the other key initiatives in this area – including the work of the ePrints Application Profile Project. Key outcomes will include a series

[9] http://www.ukoln.ac.uk/repositories/digirep/index/Deposit_API

[10] http://www.loc.gov/standards/sru/

[11] http://www.ukoln.ac.uk/repositories/digirep/index/Deposit_API

of dissemination events, publications, workshops and focus groups. The Project will maintain a development WIKI, and a project issues blog in order to ensure rapid and ongoing transfer of the learning back into the wider community.

59. Timescale – project months 1-18; summative dissemination activities to take place in months 17-18.

60. Resources – Project Manager, Project Officer, Technical Officer.

3. Budget

4. Project Management

61. The Project will be managed in accordance with JISC project management protocols. It will be informed by formal project methodologies and will conform to recognised good practice in the field.

4.1 Project Directors

62. The Project will be jointly directed by the Steering Group members of the White Rose Consortium ePrints Repository. The Directors will oversee the work of the Project and act as an advocate for it to the external community. The people identified for this role are Tracey Stanley, Head of eStrategy at the University of Leeds; Elizabeth Harbord, Head of Collection Management at the University of York; and Peter Stubley, Assistant Director for Academic Services at the University of Sheffield.

4.2 Project Steering Group

63. The Project will utilize existing Steering Group arrangements for the current management of the White Rose Consortium ePrints Repository.

64. The Steering Group meets on a monthly basis and is currently used to oversee the strategic development of the Repository, and to provide a steer on key issues. This role will be extended to include the IncReASe Project. The Steering Group will play an active role in the Project and will drive activities and provide expert input, as well as acting as the key sign-off for all Project outputs and for Project evaluation.

65. Steering Group membership is as follows: Tracey Stanley, Peter Stubley, Elizabeth Harbord, Rachel Proudfoot (Repository Officer). The Project Officer will also be asked to join the meeting, and the Technical Officer will be invited for specific discussions.

4.3 Staffing arrangements

66. The following posts are required for the Project:

- Project Manager (0.3 FTE for 18 months) – to provide day-to-day project management, including development of a project plan, detailed work-packages breakdown, communication plan, risk log, ongoing issues log, provide key reports to JISC, evaluation and dissemination plan, lessons learned log and final report. To provide line management for the Project Officer, and ongoing supervision of the work of the Technical Officer. The person identified for this role is Rachel Proudfoot, White Rose Consortium ePrints Repository Officer.
- Project Officer (0.5 FTE for 18 months) – to lead the work on work-packages 2, 3, 4 and 5. To provide input on project management and evaluation/dissemination activities. Recruitment will take place to fill this post.
- Technical Officer (1 FTE for 18 months) – to lead the technical development activities of the Project. To provide input on project management and evaluation/dissemination activities. Recruitment will take place to fill this post.

4.4 Risk Management

Risk	Categorisation of risk	Action to prevent / manage risk
Staffing is not recruited in time for start of the Project	High	The Project Directors will develop job descriptions and advertise as soon as possible upon hearing that the funding bid has been successful. The Project Directors will co-ordinate the initial project set-up work until staffing appointed.
Staff members leave during the course of the Project	Medium	Team members will have 3 month notice period so some overlap of contracts may be possible. We will ensure that other key staff are closely involved in the Project, so disruption is minimal. We will ensure that the Project is well documented so new staff could take over if required.
Key stakeholders do not buy in to/support the Project	Low	To ensure regular information flow to all stakeholders, and seek feedback on direction and progress at every opportunity. To develop a communication strategy for the Project
Expectations from stakeholders are higher than we can deliver.	Medium	Ensure regular information flow, develop communication strategy. Set clear boundaries on project scope.

4.5 Evaluation

67. Project evaluation is the subject of work-package 7, and is detailed in that work-package, above. Formative evaluation will consist of ongoing analysis of the Project, and formal consultation with stakeholders. A final report will be produced at the end of the Project.

4.6 Dissemination

68. A key element of the Project will be the sharing of learning outcomes with the wider community. Dissemination will be an ongoing activity throughout the Project and will take place within the framework of a Dissemination Strategy devised in the early stages of the Project. Dissemination will be co-ordinated by the Project Manager but will be carried out by all main participants in the Project. Key dissemination activities are likely to include a series of dissemination events to be held at different institutions; the Project web site - which will be regularly updated by the Project Manager and will include news of new developments; presentation of public Project documents on the web site; email announcements on email discussion lists such as lis-link; engagement with other projects in the JISC Programme; publications for the professional and academic literature; conference presentations - various key conferences will be targeted for dissemination opportunities; blogs and WIKIs.

4.7 Quality assurance

69. All work packages will be subject to quality assurance activities, including production of end stage reports and detailed testing and quality assurance audits. QA will be the responsibility of the Project Manager, who will ensure that requirements are identified, monitored and reviewed.

4.8 Standards and technical development

70. The Project will work with JISC-identified standards in the JISC Information Environment, wherever possible.

4.9 IPR

71. The Project will be managed according to JISC guidelines for the management and handling of intellectual property. Outputs will be made available free of charge to the UK HE and FE

community in perpetuity, and will be available for wide dissemination in partnership with JISC.

4.10 Exit Strategy

72. This work is of considerable value to all of the institutions involved, as well as being of significant value to the wider JISC community. In particular, the work will inform developments in best practice for the local management of repository content, and the workflow issues associated with this. The Project will feed into institutional strategies for further embedding of the repository, and will provide long-term benefits for all three institutions, in terms of raising the profile and impact of the repository, Leeds will commit to maintaining the project outputs, including the Project website, documents and software and scripts.

4.11 Stakeholders

Stakeholder	Interest level
JISC Digital Repository Programme	High
Repository community within the UK	High
Global repository community	Medium
Research Councils	Medium
Academic staff within institutions	Medium

5. Key Personnel:

Rachel Proudfoot, Repository Officer, Universities of Leeds, Sheffield and York:

Rachel has worked on the White Rose Consortium ePrints Repository for the last two and a half years; the repository has steadily increased over this time period from five items to over 1,500 items. She has worked closely with researchers, library staff and research office staff across the White Rose consortium and has actively contributed to several JISC funded projects, including SHERPA, SHERPA-DP and StORe. Previously, Rachel has worked within the HE sector in library and student advice roles and has worked for the RNIB providing information on access technology.

Tracey Stanley, Head of eStrategy and Development, University of Leeds:

Tracey is responsible for the development of Library strategy for electronic services and electronic resources. She is the Project director for the University Institutional Portal Project, and is responsible for the University Virtual Learning Environment Service. Tracey is currently Project Director for the JISC-funded EVIE Project and for the JISC-funded MIDESS Project, and previously Project Director for the JISC-funded PORTOLE Project. She is leading the White Rose Consortium ePrints Repository at Leeds, and has also been involved in other JISC projects – including CEDARS and ILEJ.

Peter Stubley, Assistant Director for Academic Services, University of Sheffield:

Peter has responsibility for the management of library support to academic departments for learning and teaching, and research. In this role, he is taking a fresh look at the way in which the library needs to be organised to support new methods of learning and teaching, in addition to being responsible for research provision and being the lead for the White Rose Consortium ePrints Repository at Sheffield. He has extensive experience of participating in and leading JISC-funded projects having been Project Director for the eLib-funded RIDING clumps project and leading the Feasibility Study into a National Union Catalogue for the UK in 2000/01. He has published widely.

Elizabeth Harbord, Head of Collection Management, University of York:

Elizabeth is responsible for the strategic and operational management of information resources, both print and electronic and leads the White Rose Consortium ePrints Repository work at York. Previously she worked at the University of Newcastle, responsible for library information systems and services. She has been involved with several JISC projects such as SHERPA, SHERPA-DP and LOCKSS, and is a member of the StORe Project management group. She has also recently managed two online cataloguing projects; the National Railway Museum Library Collection (funded by the AHRC) and Yorkshire Country House Libraries (funded by MLA Yorkshire).

Appendix: Further information on capabilities of partners

The University of Leeds

The University of Leeds is one of the largest research Universities in the UK. It provides courses in all core disciplines to an increasingly diverse student body, a growing number of whom are recruited through flexible and lifelong learning initiatives as part of the widening participation agenda.

The University Library has a significant record of involvement in a wide number of relevant initiatives and projects. It was one of the largest recipients of e-Lib/Follett funding, and has recently received funding for a range of projects including MIDESS (digital repositories) and EVIE (development of a virtual research environment). The Library also received JCALT funding (joint with Manchester Metropolitan University Library) to participate in "The Big Blue", a study of information skills training.

The University of York

The University of York is one of the smallest universities in the UK but is regularly ranked in the top ten UK universities for research and teaching and is a founder member of the Worldwide Universities Network of sixteen research universities. It is currently embarking on a period of expansion to increase the number of staff and students by 50-70% over the next 7 years. It has strong departments across the humanities, social sciences and science, including medicine and nursing, with two new departments of Theatre, Film and Television, and Law starting in the next two years.

The University has made significant investment in the Library in recent years to build its collections, especially in the humanities; funding from a benefactor and the HLF has enabled a new library building. The Library has in recent years received project funding from HEFCE (retroconversion of York Minster Library), RSLP (resource sharing in archaeology), AHRC (retroconversion of the National Railway Museum Library) and for several JISC repository projects (SHERPA, SHERPA-DP, StORe).

The University of Sheffield

The University of Sheffield celebrated its centenary in 2005 and is one of the UK's leading universities, with an international reputation for excellence in both teaching and research. With over forty departments across seven faculties – Architectural Studies; Arts; Engineering; Law; Medicine; Pure Science; Social Sciences – the University offers a wide range of discipline-based and interdisciplinary courses. There are 25,000 full-time and part-time students studying in the University, of whom 19,450 are undergraduates and 5,550 postgraduates. Included in this figure are more than 3,800 international students from 116 countries. The University Library is a multi-site system with 8 service points. There are over 180 staff, and some 1.4 million items of stock. The Library makes extensive use of IT-based systems, including the Talis library management system, and is a member of the Consortium of Research Libraries (CURL) and the US-based Research Libraries Group (RLG).

Appendix 2: Workflow diagram

Figure 1: Repository ingest and outputs