



Project Information			
Project Acronym	-		
Project Title	Shared OpenURL Data Infrastructure Investigation		
Start Date	1 st December 2008	End Date	31 March 2009
Lead Institution	EDINA, University of Edinburgh		
Project Director	Christine Rees		
Project Manager & contact details	Tim Stickland t.stickland@ed.ac.uk ; 0131 650 3309		
Partner Institutions	-		
Project Web URL	http://edina.ac.uk/projects/Shared_OpenURL_Data_Infrastructure_Investigation_summary.html		
Programme Name (and number)	JISC Repositories and Preservation Programme		
Programme Manager	Neil Jacobs		

Document Name			
Document Title	<i>Project Plan</i>		
Reporting Period			
Author(s) & project role	Tim Stickland, Leah Halliday, Christine Rees		
Date	31/1/09	Filename	
URL	http://edina.ac.uk/projects/Shared_OpenURL_Data_Infrastructure_Investigation_summary.html		
Access	<input type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
0.1a	Jan 2009	Draft
1.0	Mar 2009	Final

Overview of Project

1. Background

“Many UK universities and colleges maintain OpenURL link servers, which direct users to potential sources of the content they want. These redirect transactions might be logged and used in various ways. The MESUR¹ project has shown some of the possibilities that aggregations of such data might offer.”²

EDINA welcomes this invited proposal from JISC to identify what OpenURL linking data is available within the UK and to scope potential uses and highlight any organisational and legal issues around use of the data. During JISC-funded project activity, as part of the JOIN-UP 5/99 programme, EDINA scoped, implemented and now runs as a service the UK OpenURL router³. This service is also a potential source of UK OpenURL linking data. It has also provided EDINA with experience of and involvement with the OpenURL standard.

2. Aims and Objectives

The aim of the project is to scope a UK architecture for the analysis of OpenURL linking data. The architecture is intended to provide a basis for ongoing services.

1. Through desk research and consultation, to gain a good understanding of the challenges and opportunities related to sharing and using OpenURL link server data including technical, legal, and administrative.
2. Using data from the OpenURL router, to establish the feasibility of identifying user behaviour across sessions or within sessions and analysing and comparing those behaviours.
3. Through consultation with a few institutions, to identify the nature of the data that are currently collected within institutions, how they are used, whether they may be shared and on what condition, and the perceived benefits of sharing.
4. To assess the relative value and complementary value of data from the OpenURL router and from OpenURL resolvers within institutions.
5. To explore potential uses of these data through development of case studies and testing of the viability of these case studies through consultation with different stakeholder groups within the UK Research and Higher Education community.
6. To identify any legal obstacles to use of the data and explore how the data may be used whilst respecting legal boundaries and constraints.
7. To identify organisational, policy and cultural issues related to aggregation of data from different institutions and explore, through consultation, the viability of gathering and analysing aggregated data.
8. To make recommendations for future gathering and use of OpenURL link server data based on: a review of the architecture proposed by LANL with reference to the findings of this project; and assessment of the type of service that may feasibly be provided with reference to the requirements and benefits, the ability and willingness of UK institutions to share data, and the willingness of other stakeholders to participate.

3. Overall Approach

To achieve the aims and objectives, the project will:

Determine Context:

1. Conduct desk research and consult domain experts such as the LANL⁴ group to identify the main challenges and opportunities with respect to sharing and using OpenURL link server data. This

¹ <http://www.mesur.org/MESUR.html>

² Neil Jacobs, JISC: project description provided to EDINA

³ <http://openurl.ac.uk/doc/>

⁴ <http://lanl.gov/>

will include an assessment of the legal framework within which such sharing and use would take place.

Uses of UK level OpenURL linking data:

2. Consider potential uses of UK level OpenURL linking data; including, if appropriate production of use cases.
3. Validate potential uses with the academic community sector

Compare and contrast the OpenURL linking data sources:

4. In principle OpenURL linking data may be obtained by sharing institutional data, or from OpenURL Router logs. Key differences that must be considered include:
 - a. OpenURL Router data is a restricted sample. Whether a more representative sample could be achieved by aggregating institutional data is doubtful in the short/medium term, since the Router has around 90 registered resolvers. In the longer term this could become a constraint.
 - b. The OpenURL Router usage data are available at a very low cost and lower risk compared to aggregating institutional data.
 - c. OpenURL Router data cannot be used for applications that require analysis of the resolver response to an OpenURL request.

Using the OpenURL Router as a data source:

5. OpenURL Router usage data are immediately available in large quantities, and will be used to make initial investigations. These will cover issues that are relevant to shared institutional data as well as data available from the Router, including:
 - a. Exploring the feasibility of identifying the actions of end users over extended periods of time, and the actions of users within a single session, and comparing the value of each of these approaches; this will include investigating the usefulness of IP addresses, cookies or other methods of discovering which requests originated from the same end user.
 - b. A survey of items requested by end users, to estimate factors such as journal coverage (is it a broad sample of journal titles?), frequency of duplicate requests (are certain articles or journals discernibly more popular than others?), and the richness of metadata
 - c. A survey of referring service identifier ("sid") values ie an indicator of where the OpenURL request originated. Firstly it will be instructive to see how frequently these values are included in requests, and possibly whether HTTP Referrer headers provide useful additional data. Secondly, and assuming these values are available for a meaningful sample, an analysis can be performed of OpenURL requests from different referrers, to estimate the effectiveness of services in helping users locate full text. A comparison of subscription services with others (e.g. google) could be included. .
6. Determine to what extent OpenURL Router data are suitable for further work, and in what instances shared institutional data are required.

Using OpenURL linking data provided by UK institutions:

7. Analyse legal issues.
8. Consider issues related to aggregation of institutional data by undertaking informal consultation across the UK academic library community to:
 - a. assess their willingness to share this data
 - b. the risks and benefits that such sharing might entail
 - c. scope any technical and organisational / policy aspects
9. With a small number of institutions, investigate:
 - a. what OpenURL link server data is available
 - b. how it is used
 - c. whether/how it might be shared and under what conditions
 - d. what benefits they would anticipate from doing so

10. Undertake informal consultation across other interested groups within the UK academic community (for example, academics, students, national stakeholders) to assess their reactions to sharing this data, and the risks and benefits that might entail.
11. Gather information about options for data collection of the most common OpenURL link servers in the UK. Liaise with suppliers if necessary to identify their interests.

4. Project Outputs

The major deliverable of this scoping will be a report that outlines the findings of the project. It will also draw out the practical recommendations that can be made in order to contribute and use OpenURL linking data, including next steps and those targeted at particular organisations. In particular the report will:

- Review the architecture put forward by the LANL group in the light of the project findings.
- Assess what is feasible given:
 - a. the requirements implied by the anticipated benefits
 - b. the ability and willingness of UK universities to share the data; the availability of data from the OpenURL Router
 - c. the views of other stakeholders including suppliers of OpenURL link servers.
- Make recommendations on the best way to move to a position where OpenURL link server data are made available and useful services can be built around the data.

5. Project Outcomes

- An understanding of the views within UK HEIs with regard to sharing and making use of OpenURL usage data.
- Evaluation of the different potential options for aggregating usage data, and the legal, cultural and organizational to do so
- Understanding of the environment in which OpenURL resolvers operate and usage data might be sourced.

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
Library staff and managers	They manage the local services that collect usage data. They also seek effective, cost- and resource-efficient solutions that interoperate with the other elements in the information environment that they provide to their users.	<i>High.</i> Interest from this group, both a willingness to contribute and share usage data, and benefits services based on the data might provide, will be a major factor determining success.
Vendors of OpenURL resolvers	They provide the tools that gather usage data.	<i>High.</i> Co-operation may be needed to implement an efficient data export stream.
UK HE/FE projects considering use of information usage data, or new/existing services based on usage data.	Common issues related to data aggregation and opportunities for collaboration. Interest in range of potential services proposed by the project.	<i>Medium.</i>
JISC IE planners	JISC IE planners are interested in providing a joined-up infrastructure and services to meet the needs of the sector.	<i>High.</i> These planners will determine whether or not any UK service should be developed.
Academic staff and students in universities and research institutions, and in colleges	These are the potential end users of any services proposed on aggregated usage data.	<i>High:</i> This group is very important if the outcome of scoping is to develop services.

7. Risk Analysis

Risk	Probability	Severity	Mitigation
OpenURL data not available from resolver applications	Very low: virtually all applications of this type enable usage logging	Medium: OpenURL Router data are an alternative, but may not provide all information	Use OpenURL Router data
Difficulty identifying users who made requests	High: experience suggests this data will be unavailable or of low quality	Medium: options available to mitigate	Find useful applications for data that do not require end users to be identified; add authentication to resolvers; investigate methods for using IP addresses or cookies
Institutions are unwilling to consider contributing, possibly due to legal issues	Medium: likely to encounter at least some issues.	Medium: any problems are unlikely to affect all institutions.	Provide funding to institutions; anonymize usage data; agree permissible use(s) of data sharing
Project staff unable to secure decision to share on time to act within project schedule	Medium: library staff are very busy and such decisions usually require consultation which can delay decision making	Medium: some institutions will be able to make decisions on schedule.	Project team will build on existing relationships to secure participation.
Staff within institutions resistant to participating due to institutional priorities and believe that participation would require time and effort.	High: staff within institutions are often overstretched.	Medium:	Project staff will seek to minimise the effort required to participate and will secure commitment for participation from senior staff via existing relationships.
Institutions are unable to share data for technical reasons	Low: not technically demanding	Medium: any problems unlikely to affect all institutions	Be flexible; do not require optimal solution (e.g. OIA/PMH), accept simple methods like FTP/HTTP downloads.
Data not available in standard format (e.g. XML Context Object)	High: anticipate this will be the case	Low: data may be transformed	Data can be transformed to the desired standard.
Failure of aggregation mechanism design or procedures	Medium: not especially complex or technically demanding, but must meet contributing institutions' needs	Very high: aggregation is a fundamental requirement	Liaise and negotiate with institutions at all stages of project; minimize technical requirements placed on institutions.

Risk	Probability	Severity	Mitigation
Aggregated data are no more useful than individual institutional data sets	Medium: various suggested applications are largely academic exercises, unproven in value-for-money terms	Medium: though value-for-money may be hard to justify, total failure to find applications is unlikely	Investigate feasibility of applications prior to creating costly, full sized infrastructure; collaborate with academic researchers to find novel applications
Lack of expertise in techniques required for applications of the data	Low: EDINA has experience in the areas that we think should be in scope	High: in-depth understanding is necessary	EDINA has contacts with specialists in informatics.

8. Standards

Name of standard or specification	Version	Notes
OpenURL (Z39.88) San Antonio Profile (SAP)	1.0	The purpose of a resolver is explicitly to handle OpenURL requests. The San Antonio Profile (SAP) is the only community profile to date that can handle bibliographic metadata.

9. Technical Development

No significant technical development will be involved at this stage. The purpose of the project is to decide feasibility, and the approach that might be used.

10. Intellectual Property Rights

The outputs of this project will be documents: a scoping report, architecture and evaluation of business models. These are deliverables of the project and will be provided to JISC on schedule. In accordance with the grant letter, IP will remain with the University of Edinburgh while the documents will be made available to interested stakeholders throughout the UK HE/FE community as appropriate. At present, the project team foresees no need to use materials for which IP is owned by third parties.

Project Resources

11. Project Partners

The project will be undertaken by EDINA staff.

12. Project Management

It will be reviewed in consultation with the JISC Programme Manager. The project will be conducted in accordance with the plan; any proposed deviation will be agreed by the EDINA Director. Exceptions/risks will be reported to the Director with a view to adjusting the plan to accommodate these whilst still meeting the project deadlines within budget. This is a small project; there will be no Management Board or Steering Committee.

The project team comprises the following:

Project Management

Tim Stickland 0131 650 3309; tim.stickland@ed.ac.uk

Consultation work

Leah Halliday 0131 650 4616; leah.halliday@ed.ac.uk
with contributions from Tim Stickland and Christine Rees 0131 650 3333; Christine.rees@ed.ac.uk

Investigation & evaluation of data aggregation options

Tim Stickland and
Des Reid 0131 650 2914; des.reid@ed.ac.uk

The project has no training needs.

13. Programme Support

The project team would value support from the Programme Manager to facilitate liaison with contacts throughout the UK academic community, and other others working in the area - JISC projects and other groups beyond the UK.

14. Budget

See Appendix A.

Detailed Project Planning

15. Workpackages

See Appendix B.

16. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
Spring 09	Publication of article in EDINA Newslite of the results of this project.	Site representatives at EDINA subscriber institutions (UK HE/FE)	To publicise the work within the potential user community.	Findings of scoping report.
Spring 09	Various news and blog items e.g. on JISC website, JIBS blog,	Librarians and other interested parties in UK HE/FE	To publicise the work within the potential user community.	Findings of scoping report.

17. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Project Report inc. scoping, architecture and business model	Will be delivered to JISC and made available directly from EDINA to interested parties.	If appropriate (depending on output from scoping), proposal to JISC for development of service or further investigation for UK HE/FE.
Learning within project team	Project team will present findings to colleagues within EDINA to ensure that future projects benefit from learning. Project team will also report through EDINA Newslite, and on the Project Website.	
Learning	Will be disseminated	

Appendixes

Appendix A. Project Budget

[Omitted]

Appendix B. Workpackages

WORKPACKAGES	Month:	1	2	3	4
1: Project Management & Liaison					
2: Consult stakeholders					
3: Determining context					
4: Determine requirements for successful sharing of OpenURL usage data					
5: Investigation of OpenURL Router data					

Project start date: *1st December 2008*

Project completion date: *31 March 2009*

Duration: *4 months*

Members of Project Team:

Leah Halliday: LH

Christine Rees: CR

Des Reid: DR

Tim Stickland: TS

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Responsibility
<p>WORKPACKAGE 1: Project Management & Liaison</p> <p>Objective: To deliver the objectives of the project within the agreed timescale and budget.</p> <p>1. Project planning; assigning tasks; monitoring progress & reporting to JISC</p> <p>2. Dissemination of project findings</p>	1 December 2008	31 March 2009	Project documentation Project Website Project Report	TS & CR
<p>WORKPACKAGE 2: Consult stakeholders</p> <p>Objective: Determine what data are collected, how they are used, attitudes to sharing and potential uses, benefits and risks as well as barriers to sharing.</p> <p>3. Initial contact with librarian to determine what data are currently collected by libraries and how they are used and to explore attitudes and any barriers to sharing including any technical and policy issues and to identify any further issues or concerns of librarians.</p> <p>4. Consultation within team to identify risks and benefits related to institutional resolver data and router data.</p> <p>5. Web form questionnaire used to assess willingness to share data, perceived risks and benefits (including uses to which shared data may be put), and to determine any technical,</p>	15 December 2008	16 January 2009.		LH
	12 Jan 2009	23 Jan 2009	Will inform development of draft case studies.	
	19 Jan 2009	6 Feb 2009	Contribution to project report.	

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Responsibility
organisational and policy issues associated with sharing.				
6. Interviews with librarians at four institutions (preferably not all using the same proprietary resolver).	2 Feb 2009	13 Feb 2009	Contribution to project report.	
7. Develop case studies on sharing.	16 Feb 2009	20 Feb 2009	Draft case studies as focus for consultation.	
8. Conduct two focus groups with members drawn from different stakeholder groups to test cases.	16 Feb 2009	6 March 2009	Revised case studies in project report.	
9. Contribute to project report recommendations	16 March 2009	20 March 2009	Contribute to recommendations in project report.	
WORKPACKAGE 3: Determining context				LH
Objective: determine the challenges and opportunities associated with sharing and using OpenURL link server data.				
10. Conduct desk research.	5 January 2009	27 February 2009	Contribution to project report	
11. Identify and consult domain experts (email and telephone)	15 December 2009	27 February 2009	Contribution to project report.	
WORKPACKAGE 4: Determine requirements for successful sharing of OpenURL usage data				TS
Objective: To identify the problems that must be overcome in order to successfully share OpenURL usage data, action required to address these problems, risks faced and possible mitigation.				
12. Identify problem areas & consider action	5 January 2009	13 February 2009	Contribution to project report.	

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Responsibility
required		2009		
13. Identify risks, and possible mitigation				
14. In light of requirements identified, compare likely usefulness of data aggregated from institutional with data from OpenURL Router logs	16 February 2009	27 February 2009		
WORKPACKAGE 5: Investigation of OpenURL Router data				
Objective: To see what can be achieved using the OpenURL Router logs. These data are likely to offer less value in certain respects, but are available immediately and in a large quantity.	2 February 2009	20 March 2009	Contribution to project report.	DR
15. Investigate ways to associate requests originating from particular users/sessions				
16. Survey bibliographic metadata				
17. Investigate ways of evaluating outcome				