

JISC DEVELOPMENT PROGRAMMES

Project Document Cover Sheet

STARGATE Extension FINAL REPORT

Project Acronym	STARGATE	Project ID	
Project Title	Static Repository Gateway and Toolkit: Enabling small publishers to participate in OAI-based services		
Start Date	9 th October 2006	End Date	22nd December 2006
Lead Institution	Centre for Digital Library Research, University of Strathclyde		
Project Directors	Gordon Dunsire (Management)/ Alan Dawson (Technical)		
Project Manager & contact details	R. John Robertson Centre for Digital Library Research, Department of Computer & Information Sciences, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH Tel: 0141 548 5854 Fax: 0141 548 4523 Email: robert.robertson@cis.strath.ac.uk		
Partner Institutions			
Project Web URL	http://cdlr.strath.ac.uk/stargate/		
Programme Name (and number)	JISC PALS II Metadata & Interoperability Projects ITT		
Programme Manager	Christine Baldwin		

Document

Document Title	<i>Stargate Extension Final Report</i>		
Author(s) & project role	R. John Robertson (Project Officer)		
Date	07/02/2007	Filename	StargateXFinalReport_1.1
URL	http://cdlr.strath.ac.uk/stargate/		
Access	Project and JISC internal	✓ General dissemination	

Document History

Version	Date	Comments
1.0	25/01/2007	Draft Final Report submitted to JISC
1.1	07/02/2007	Minor edits made, report published

Stargate Extension Final Report

Author: R. John Robertson

Date: February 2007

Version: 1.1

Centre for Digital Library Research (CDLR),

University of Strathclyde, Glasgow G1 1XH



Table of Contents

1	ACKNOWLEDGEMENTS	4
2	EXECUTIVE SUMMARY	5
3	BACKGROUND	6
3.1	Project Context	6
3.2	The development of static repository gateways.....	6
4	AIMS AND OBJECTIVES	8
4.1	Aim.....	8
4.2	Objectives	8
5	METHODOLOGY	10
6	IMPLEMENTATION	11
6.1	Surveying the use of gateways	11
6.2	Documenting the software installation.....	11
6.3	Estimation of running costs	12
7	OUTPUTS AND RESULTS	13
7.1	Overview of current use of static repository gateways.....	13
7.2	Software and hardware prerequisites.....	14
7.2.1	Hardware prerequisites	14
7.2.2	Software prerequisites.....	14
7.3	Software installation issues	14
7.4	Branding issues	15
7.5	Costing of static repository gateway deployment.....	15
7.5.1	Installation costs	15
7.5.2	Operational costs	15
7.6	An operational static repository gateway	16
8	OUTCOMES	18
8.1	Achieving the aim, meeting objectives	18
8.2	Assessing the extension.....	19
9	CONCLUSIONS	20
10	IMPLICATIONS	21
10.1	Direct implications.....	21
10.2	Related issues	21
11	RECOMMENDATIONS	22
12	REFERENCES	23

1 Acknowledgements

This project extension was funded under the JISC PALS II Metadata & Interoperability Projects programme.

We would like to thank the developers of the 'sreped' software and other gateway implementers and developers who contributed to the STARGATE project extension. In particular, Patrick Hochstenbach and Tom Habing.

2 Executive Summary

The extension to the STARGATE (Static Repository Gateway and Toolkit) project was funded by the Joint Information Systems Committee (JISC) to investigate the issues surrounding the deployment of static repository gateway software. Its intent was to provide guidance about the installation and branding of the software and so to support an assessment of the viability of the deployment of a static repository gateway by the publishing community, by JISC, and by other interested parties.

The extension has produced a functional branded gateway that the publishing community can use to explore the use of static repositories. It will be maintained for the next year. The gateway is available at <http://stargate.cdjr.strath.ac.uk/gateway/>.

The extension has produced installation and branding guides, and has carried an overview of existing gateway installations. This information should help those wishing to install and run a gateway, and will assist the relevant communities in assessing the deployment of static repository gateways.

The project concludes that although functional the software is not suitable for deployment by a novice user. It is also effectively still in at the beta stage of development and it has only been used in a limited number of settings.

The project further suggests that the creation and maintenance of gateway(s) within the publishing community may be more suitably carried out in the same way that DOI and Purl provision is offered through a third-party service provider willing to work with developing open source software. Any deployment of a gateway by JISC to support wider participation in static repositories should also engage with the gateway software developers.

3 Background

3.1 Project Context

This brief project extension builds on the findings of the STARGATE project, as such much relevant background information is provided in the final report (Stargate Final Report (2006), version 1.4,

http://cdlr.strath.ac.uk/stargate/StargateFinalReport1_4.pdf). This document assumes familiarity with that document, and will try to avoid duplicating its content.

3.2 The development of static repository gateways

Static repositories were developed as part of version 2.0 of the OAI-PMH; they emerged in response to a perceived need for a simpler approach to participation in OAI-PMH-based services, and were based on work done on virtual data providers within OAI 1.0 by the Open Language Archives Community (OLAC).

The functional requirements of a static repository gateway were developed as part of the static repository specification. An initial software implementation of the static repository gateway specification, described in was developed at LANL as a prototype and released as Open Source software through Sourceforge. An experimental Gateway based on this software was made available by LANL to support the use of static repositories. The software and Gateway service were updated in line with the release of version 2.0 of the OAI protocols. The initial software development 'srepod' continues to be periodically developed and forms the basis of a number of operational static repository gateways.¹ The development of the specification and software is explained in more detail in Hochstenbach et al. (2003) [1].

At least one other implementation of the Gateway specification has been developed – the 'ASP OAI 2.0 SR Gateway'. This implementation was developed as part of the University of Illinois Open Archives Initiative Metadata Harvesting Project (<http://uilib-oai.sourceforge.net/>). It created a gateway in ASP to run under Microsoft's IIS web server technology. The software was designed to conform to an alpha version of the

¹ It should be noted that the actual full title of the srepod software is 'OAI-PMH Static Repository Gateway' – srepod is the name of the downloadable file; it will however, be referred to as the 'srepod' software to avoid confusion with the gateway specification or other implementations of gateways.

static repository gateway specification. Although the last public release of this software was in October 2003, University of Illinois Urbana Champaign (UIUC) have continued to develop this software for their own use and hope to be able to update the public release in the future.

4 Aims and Objectives

4.1 Aim

The Stargate project's primary aim was to lower the technical barriers to the implementation of OAI-compliant repositories, thereby enabling small publishers of electronic resources to participate more readily in OAI-based disclosure and delivery services.

This project extension has enhanced the impact of Stargate by further investigating the project's second recommendation (Stargate Final Report (2006), p 24):

“The publishing community, in so far as it has a communal interest in OAI exposure, should consider funding a static repository gateway with some degree of service guarantee. It would support the development of publisher-specific services and allow a degree of branding to be attached to the metadata provided therein. At this stage, it is unclear who would host or support this but there are a number of companies who already provide technical services to the sector (the gateway software required is open source, stable, and complete).”

It has done so by examining in more detail the requirements of the static repository gateway software, by documenting its deployment, and by presenting informed estimates for viability of such a service being run and maintained by the publishing community. This review of the static repository gateway software will also be informed by an overview of current gateway deployment. This software infrastructure is a precondition of the use of static repositories.

The project extension supports the publishing community's assessment of the viability of static repository technology and informs other communities about the issues and costs involved in running static repository gateways.

4.2 Objectives

1. The project has created a demonstration static repository gateway for the publishing community and document the process. This involved:
 - a. Setting up a web server
 - b. Setting up a gateway
 - c. Branding/ customising the look and feel of the gateway
 - d. Documenting and explaining the above process

2. The project also investigated the cultural preconditions and financial cost of setting up and running a static repository gateway. This has entailed:
 - a. Estimating the set-up costs of a gateway
 - b. Estimating the maintenance and running costs of a gateway (the relative importance of this objective of the project has diminished in light of the findings of 1b – Setting up a gateway)
 - c. Investigating, as far as possible, the operating conditions of existing gateways
 - d. Providing recommendations to the publishing community

5 Methodology

In studying the deployment of static repository gateways, the project extension carried out two strands of enquiry; it engaged with other organisations running gateways and it documented the installation process and investigated the customisability of the default interfaces.

For the survey, gateways were identified by searching a number of repository registries and services for the distinctive url structure of static repositories and by looking for url components distinctive to known pieces of gateway software. Where possible the homepages for these gateways were located and administrators identified. Brief surveys about their gateway were then sent to these individuals. The results of these surveys provided the basis for an overview of the static repository gateway community.

The other strand of the study was to install, document, and re-install the static repository gateway software on a new server. This allows the creation of an enhanced installation guide. This process also allowed the identification of how the gateway interfaces are created and allowed an examination of what within them would be customisable to allow the development of branding for gateways.

Within the timescale of the project these approaches appeared to provide the best investment of time. In a longer investigation additional work could have been carried out to test and adapt the installation guide with a number of novice users and to try to hold interviews or meetings with other gateway administrators.

6 Implementation

6.1 *Surveying the use of gateways*

As outlined in the methodology, static repository gateway providers were identified through manually scanning various registries of repositories (The Open Archives registry and the University of Illinois' registry) for the distinctive double urls of mediated static repositories. The process was made simpler as the standard installation of the various pieces of gateway software creates a consistent folder structure which is then replicated in the base url. For example, the srepod software, creates a folder named 'srepod' which is then part of the gateway's url and so is included in the url of any mediated repository. This is seen in the following sample url:

http://purl.oclc.org/NET/ugent/lib/srepod/www.progettocaere.rm.cnr.it/databasegestione/A_C_oai_Archive.xml?verb=Identify .

The identified gateway providers were emailed a brief questionnaire asking for some key details about their mediation of static repositories. The community of gateway providers is small; there are around half a dozen static repository gateways in operation - allowing the project to quickly gain an overview of the current use of this technology. Unfortunately one of the identified gateways was unable to respond to our questionnaire – some of their details were, however, able to be identified from the gateway itself.

6.2 *Documenting the software installation*

After a new Linux server was set up to host the gateway, the latest version of the 'srepod' software was downloaded and installed. This process was documented to produce the accompanying guide. During this process, however, some critical issues emerged for the use of the software. These included the identification of further software prerequisites (and their subsequent installation) and the replacement of a component file. As detailed in the installation guide the downloaded software package required the prior installation of other software which was neither mentioned in the installation documentation provided with the software package nor on the software's Sourceforge homepage. A more complex issue was that one of the files

included in the software package didn't work and the software developer had to be contacted to acquire a replacement file. The developer was extremely helpful and the replacement allowed the subsequent installation to proceed smoothly.

The installation of the additional software required and the replacement of the component file were both relatively straightforward problems to address, but their occurrence increased the technical skills required of the installer and the time commitment involved. The installed software was tested by submitting valid and invalid static repositories and by deregistering mediated static repositories and it was found to work normally.

After installation it was confirmed that the web interface to the gateway software was based on static webpages. This allowed the project to examine the default html files to see how they could be customised to support the branding of gateways. Trial interfaces were developed and successfully substituted for the default ones. This process, including a careful note of the required web page elements, was documented to produce the attached guide.

6.3 Estimation of running costs

The survey responses provided a basis for an estimate of the technical and administrative support commitments of running a gateway. The financial costs involved in setting up and running a gateway were estimated from this data and from the project's initial start up costs. Some initial estimates of the cost of web-hosting were investigated, but the size of the application and the expected traffic meant that the costs involved were relatively negligible. A related consideration of the developmental nature of the software suggested that local hosting of the software was more likely given the requirements of administering the software.

7 Outputs and Results

7.1 Overview of current use of static repository gateways

The project carried out a brief survey of existing static repository gateways in order to estimate the current use of static repositories and gain an estimate of the level of support required to provide a gateway. A concise summary of the overview is presented here and the full overview report is available on the project website (<http://cdlr.strath.ac.uk/stargate/>).

There are around 80 static repositories mediated by four static repository gateways (based at: LANL, UIUC, University of Ghent (Ghent), and the University of Strathclyde (OAIScotland)). All of the gateways exist in research environments, but OAIScotland and UIUC also identify themselves as informal service-providers. LANL mediates the most static repositories, with other gateways averaging 7 repositories each. About 3 repositories are added to each gateway per year (the figure for LANL is unknown).

The administrative and technical commitment to running a gateway is quite small. Repository administration (registering and terminating repositories) typically requires 2 days a year and supporting and developing the gateway software averages 8 days a year (including server backup and maintenance). The gateway software can easily share a server with other applications however, allowing the maintenance costs to be shared.

It should be noted that two of the three respondents are responsible for the creation of gateway software. Both pieces of gateway software are still under development. All of the gateway administrators noted that in the software they use there are still features that require development and a small number of bugs that continue to be discovered.

7.2 Software and hardware prerequisites

7.2.1 Hardware prerequisites

The hardware requirements of the software are minimal, a demonstration version created prior to the extension ran happily, though slowly, on a Pentium 3; consequently, the lowest specification of a new machine should be more than adequate. The project is providing the interim publisher's gateway on the following specification of machine:

- Processor: P4 3.2 Ghz 1mb cache
- Memory: 2Gb DDR2
- Hard disk: 2x 250Gb 7200rpm SATA disk drives with a raid controller
- Onboard graphics and sound.

7.2.2 Software prerequisites

The gateway software requires:

- An operating system; the software works on the following: All POSIX (Linux/BSD/UNIX-like OSes), Linux, Solaris.
- Apache, or other web server software
- Makeinfo (part of the Texinfo documentation software)

Optional recommended software

- Pico or PFE or other text editor

7.3 Software installation issues

The project produced a guide to the installation of the srepod software to supplement the installation notes provided with the software. It provides a walkthrough of the installation process and some troubleshooting assistance. It is available on the project website (<http://cdlr.strath.ac.uk/stargate/>).

The guide outlines the installation process step by step, providing some screenshots and examples to makes the process of software installation easier. It should be noted however, that the software is still a relatively early beta – installing and running it requires some Linux proficiency, root administration rights on the server the

software is being installed on, and may require some engagement with the developer.

7.4 Branding issues

The project produced a guide to branding the srepod software. Starting with the default interface, it outlines the core code that any web interface must have to function, and then outlines the parts of the html that can be easily adjusted. It goes onto provide sample interfaces with Cascading Style Sheets (CSS) implemented and also to provide sample Dreamweaver templates for the interface. The guide, sample interfaces, and templates are available on the project website (<http://cdlr.strath.ac.uk/stargate/>). It should be noted that as long as the core code (which gathers the repository's url and submits it to the software) is in place the register and terminate interfaces can be in any website structure or page format.

This aspect of deploying a static repository gateway is accessible to any user who is happy to edit webpages. The sample interfaces using CSS provided alongside the guide support the creation of accessible webpages and the Dreamweaver templates allow the interfaces to be edited in industry-standard software (although any html editor could be used for the CSS samples).

7.5 Costing of static repository gateway deployment

From the survey of operational gateways and the installation guide the following can be said about the costs of deploying a gateway.

7.5.1 Installation costs

The software requirements are such that a basic desktop machine (as specified above) can be used as a server. The typical cost of such a machine might be £600.

The installation process for the gateway software is likely to require a between a few hours and two man days for a programmer with Linux experience.

7.5.2 Operational costs

These estimates are drawn from the experience of existing gateways; it should be noted that a static repository which is providing information about a frequently

updated resource (such as an ongoing run of journals), will have higher administrative overheads.

Administering the gateway is likely to cost around 2 man days per three repositories per year. The initial validation and registration of a repository will often represent the bulk of this work. This role requires knowledge of XML, some library and information science skills, and may require familiarity with Linux.

Technical support is likely to cost around 8 man days per year. This role requires some Linux programming skills and server maintenance skills.

The bandwidth requirements to run the software are not high. The maximum size of a static repository is 2mb and it can be estimated that the average repository updates at most twice a year. If a gateway has 7 repositories this suggests it needs about 28mb to update itself per year, and a comparable figure for each harvester. Basic hosting packages examined measure bandwidth in gigabytes so bandwidth usage is not likely to be a significant factor in the short term. A more significant issue is the cost of having root access to the server.

However, in both the survey of existing gateways and the review of the installation process it has been noted that the software is still in early beta, is still primarily used in research environments, and that a degree of Linux knowledge is required to set it up and administer it. This suggests that the software is not suited to being commercially hosted, unless the external host is also providing most of the technical support and administration. Consequently a detailed costing is not useful beyond the above estimates of man days – the gateway software is still at a stage of development that precludes novice use; it is more suited for use by intermediaries offering it as a service, at as yet undetermined market rates. This type of service provision may be seen as parallel to the DOI service offered by Crossref or OCLC's PURL service.

7.6 An operational static repository gateway

The project has created an operational static repository gateway for journal publishers. The gateway is available at <http://stargate.cdli.strath.ac.uk/gateway/> and

has been branded with the project’s logo. It will be available and maintained until December 2007 at least. The provision of this gateway will allow the publishing community to explore the use of static repositories and consider the deployment of a gateway on a more permanent basis.



Figure 1 Screenshot of STARGATE's gateway

8 Outcomes

8.1 Achieving the aim, meeting objectives

The project's primary aim outlined in the project plan was to "support the publishing community's assessment of the viability of Static Repository technology. It will also inform other communities about the costs involved in running Static Repository Gateways".

This aim was set out in the following objectives:

1. "The project will create a demonstration Static Repository Gateway for the publishing community and document the process. This will involve:
 - a. Setting up a web server
 - b. Setting up a Gateway
 - c. Branding/ customising the look and feel of the Gateway (as far as possible)
 - d. Documenting and explaining the above process
2. The project will also investigate the cultural preconditions and financial cost of setting up and running a Static Repository Gateway. This will entail:
 - a. Estimating the set-up costs of a Gateway
 - b. Estimating the maintenance and running costs of a Gateway
 - c. Investigating, as far as possible, the operating conditions of existing Gateways
 - d. Providing recommendations to the publishing community"

Alongside the installed gateway for publishers, the two guides to the software that the extension produced (*Static Repository Gateway Software Installation* and *Static Repository Gateway Branding and Customisation*) met the first of project's objectives. The second objective has been met by the *Overview of existing Static Repository Gateways* and by sections of this final report. One area in which the project exceeded its expectations was that a relatively detailed response from a number of operational gateways was gained – resulting in a more comprehensive overview than the project had expected. The conclusions of this overview and the installation guide resulted in the estimation of the set-up and running costs to be less detailed than had been expected. The costs are provided in terms of man days rather than as absolute costs as it quickly emerged that the gateway software is not yet suitable for off the shelf deployment but is more likely to be a provided service – at an undetermined market cost.

8.2 Assessing the extension

The outputs of this extension should inform JISC, the publishing community, and other interested parties about the practicalities of deploying static repository gateways. The benefits for the teaching learning and research communities exist in so far as such bodies chose to support the use of static repositories by providing gateways. The extension's work informs that choice.

The provision of gateways is necessary to support the use of static repositories which represent a key component in an OAI-based information environment as they allow metadata about sets of resources to be easily exposed using desktop technology. Static repositories are not going to be the dominant form of OAI provision but they do present a simple way for some types of content to be exposed. The simplicity of static repository technology is highlighted by a comment from one of the Gateway providers based at UIUC - "An instance of our gateway has been used as a teaching aid in our Library School. Students are asked to create an OAI static repository and then register it with the gateway, and harvest the records."

9 Conclusions

The project extension has explored issues relating to the deployment of static repository gateways. It has provided a gateway which the publishing community can use to explore the suitability of static repositories. It has also provided an installation guide, a branding guide, and an overview of existing gateways to inform the publishing community and other communities about the process and costs of deploying a gateway.

The extension has concluded that although the 'sreped' software is entirely capable of fulfilling its mediatory role in the information environment it is currently not yet suitable for off-the-shelf deployment. It is rather more suitable for deployment as part of an ongoing participation in static repository research or in the context of a service provider used to software development or supporting the deployment of beta software. It may not yet be particularly suitable for deployment by a commercial publishing organisation not used to open source software.

The software is however, capable of customisation for branding purposes and as development work continues, and new versions are released, it may reach a point where off the shelf deployment is possible.

10 Implications

10.1 *Direct implications*

Within the publishing community suitable technical providers need to be identified to provide gateways services. Such providers should expect to need to become involved in the gateway development community as the experience of other gateway administrators and the software developers is an important resource.

The srepod gateway software is not as polished as commercial providers may expect but the core functions are stable and reliable. Development and troubleshooting are ongoing and gateway administrators should subscribe to the sourceforge's rss feed for file releases in this software project (the project's news feature is not updated). The srepod gateway software is however, quite capable of being branded and incorporated into an existing website.

The provision of a gateway service within the JISC information environment is likely to boost the use of static repositories and indirectly encourage the development of the software. With some support, the gateway software is ready for deployment in that setting.

10.2 *Related issues*

In the course of the project extension, another of the journals that had participated in the original project switched to using the Open Journal System – a open source software product supporting the journal production workflow - that can also expose metadata as an OAI repository. This reinforces the observation in the original project's final report that for some data providers using a static repository will be a transitional step.

If the number of gateways increases the issue of which gateway a repository should be registered in will emerge. Registries of, and service descriptions for gateways, will be helpful.

11 Recommendations

If they wish to use static repository technology the publishing community should seek to identify suitable service providers willing to support 'beta' software.

Subject to their review of this project's outputs, JISC should deploy a static repository gateway to support the use of static repositories within the information environment. Such a deployment should be engage, as possible, with other gateway providers.

The Stargate project extension should attempt to make its outputs as available as possible to OAI community by ensuring the software developers are aware of them.

Future work on developing either srepod or the asp-based software should be supported. In particular, more detail about the Java version of srepod which is under development should be sought by any future efforts in gateway provision and UIUC should be encouraged to release the updated version of their ASP-based software. The project notes that direct JISC support for these efforts is unlikely as neither of them are based within the UK. More general support for the use of static repositories is however, likely to encourage development.

A scoping of how to identify which gateway a static repository should be registered in, may be required. This may involve registries such as IESR, but is likely to also need some form of human readable view of this data that could be incorporated into a gateway websites.

12 References

[1] Hochstenbach, P. et al. (2003) *The OAI-PMH static repository and static repository gateway*. Available from:

<http://public.lanl.gov/herbertv/papers/jcdl2003-submitted-draft.pdf>