

Stargate project extension

Overview of existing Static Repository Gateways

Author: John Robertson

Date: January 2007

Version: 1.0

Centre for Digital Library Research (CDLR),
University of Strathclyde, Glasgow G1 1XH



1	Introduction.....	3
2	Overview.....	4
2.1	Gateway locations.....	4
2.2	Administrating the gateway.....	4
2.2.1	Number of static repositories mediated.....	4
2.2.2	Number of static repositories registered per annum.....	5
2.2.3	Time spent on gateway administration per year.....	5
2.2.4	Typical frequency of static repositories update.....	5
2.2.5	Purpose of gateway provision.....	6
2.2.6	Gateway service provision guarantee.....	6
2.2.7	Source of gateway budget.....	6
2.3	Technical overview of gateways.....	7
2.3.1	Software used for static repository gateway provision.....	7
2.3.2	Server provision for static repository gateways.....	7
2.3.3	Application and server maintenance requirements.....	7
2.3.4	Are you aware of any outstanding technical issues with the gateway software that you use?.....	8
2.4	Other comments from the gateway administrators.....	8
3	Summary.....	10
	Appendix 1 Sample Questionnaire.....	11

1 Introduction

This document, produced as part of an extension to the Stargate project, provides a brief overview of the deployment of Static Repository Gateways in December 2006. The gateway and repositories created in the Stargate project itself have been excluded from this overview. The overview focuses on publicly available gateways – there may be gateways used within large organizations that the project is not aware of.

It is based on brief email questionnaire exchanges with the administrators of existing gateways and a small amount of further investigation.

A survey of a number of registries of repositories identified operational static repository gateways. Operational gateways are hosted by the Los Alamos National Laboratories (LANL), the University of Illinois Urbana Champaign (UIUC), the University of Strathclyde (OAIScotland), and the University of Ghent (Ghent). Responses were received from the administrators of three of these gateways, and estimates were able to be made for some details about the LANL gateway. Another possible static repository gateway was identified at School of Agronomy of the University of Uruguay, but no contact was able to be made with this installation's administrators.

The Open Language Archives Community (OLAC) hosts a comparable service based on the same principles as static repository gateways but, unfortunately, it was unable to respond to the survey. Their implementation combines results from static and normal repositories – this has the indirect effect of reducing the information that is publicly available about their use of static repositories.

2 Overview

2.1 Gateway locations

Two of the gateways are in Europe and two in the US.

Three are located within universities and the other in a national research laboratory.

The base URLs and administrators of the known Static Repository Gateways are:

1. *UIUC*

Tom Habing

<http://imlsdcc.grainger.uiuc.edu/gateway/oai.asp>

2. *Ghent*

Patrick Hochstenbach

<http://purl.oclc.org/NET/ugent/lib/srepod/>

3. *OAIScotland*

Gordon Dunsire/ Alan Dawson

<http://oaiscotland.cdrl.strath.ac.uk/gateway/>

4. *LANL*

<http://purl.lanl.gov/NET/srepod/>

2.2 Administrating the gateway

2.2.1 *Number of static repositories mediated*

1. *UIUC*

UIUC mediates 9 repositories.

2. *Ghent*

Ghent mediates 9 repositories

3. *OAIScotland*

OAIScotland mediates 3 repositories (4 repositories were recently moved to the STARGATE gateway)

4. *LANL*

LANL mediates around 58 repositories (estimate from the <friends> response). Many of these however, appear to be demonstration or

experimental. For example, a number include the word test in their name and a number of the listed repositories produced no response.

2.2.2 Number of static repositories registered per annum

1. UIUC

On average UIUC registers 2-3 each year.

2. Ghent

Ghent registers 4 each year.

3. OAIScotland

OAIScotland registers 2 each year.

2.2.3 Time spent on gateway administration per year

1. UIUC

UIUC spends approximately an hour per registration/ termination. This can be much longer however, if a registrant needs help with their XML. Given the estimation of 2-3 repositories per annum this might equate to 2 days per annum.

2. Ghent

Ghent spends approximately 2 days per annum administering the gateway

3. OAIScotland

OAIScotland spends approximately 2 days per annum administering the gateway.

2.2.4 Typical frequency of static repositories update

1. UIUC

Typically UIUC's repositories are updated at least once after registration but usually only within their first year.

2. Ghent

One of Ghent's repositories is updated weekly, one monthly, and the others annually.

3. OAIScotland

Some of OAIScotland's repositories are updated quarterly, others annually.

2.2.5 *Purpose of gateway provision*

1. *UIUC*

UIUC is operating as both a research gateway and service provider

2. *Ghent*

Ghent is operating as a research gateway

3. *OAIScotland*

OAIScotland is operating as both a research gateway and service provider

2.2.6 *Gateway service provision guarantee*

All the respondents offer some informal service guarantee; none of them are aware of any gateway with a formal service guarantee. They each describe their informal service guarantee as follows:

1. *UIUC*

“We tell people that we plan to maintain our gateway into the foreseeable future and will continue to mediate any registered and conforming repositories as long as we do, but we do not guarantee any consistent level of up-time. We try to run the service as a production system, but occasional system upgrades or other issues may require us to take the service down for periods of time.”

2. *Ghent*

“I'll support static repositories as long as I can manage or find funding.”

3. *OAIScotland*

“Yes, [an informal service guarantee is offered]”

2.2.7 *Source of gateway budget*

1. *UIUC*

UIUC has no specific budget, the operational costs are absorbed

2. *Ghent*

Ghent has no specific budget, the operational costs are absorbed

3. *OAIScotland*

OAIScotland has no specific budget, the operational costs are absorbed

2.3 Technical overview of gateways

2.3.1 *Software used for static repository gateway provision*

1. *UIUC*

UIUC run their own static repository gateway software based on ASP technology. (NB their operational version is more up-to-date than the available open source version. The open source release may be updated in the future to better reflect the operational version.)

2. *Ghent*

Ghent uses LANL srepod customised with UGhent branding

3. *OAIScotland*

OAIScotland uses LANL srepod customised with OAI Scotland branding

4. *LANL*

LANL presumably uses LANL srepod

2.3.2 *Server provision for static repository gateways*

1. *UIUC*

UIUC's gateway is hosted on a general purpose server; the gateway is a fairly minor service compared to others that are hosted on the same server.

2. *Ghent*

Ghent's, this server also runs UGhent search engines.

3. *OAIScotland*

OAIScotland's server is likely to run other applications in the near future.

2.3.3 *Application and server maintenance requirements*

1. *UIUC*

UIUC carries out server maintenance and upgrades on the static repository gateway server as part of a regular administration cycle. Occasional software bugs with the gateway require perhaps 4 days a year ("probably less than 8 hours per quarter").

2. *Ghent*

Ghent spends 15 days a year on application and server maintenance.

3. *OAIScotland*

OAIScotland spends 5 days a year on application and server maintenance.

2.3.4 *Outstanding technical issues with the gateway software*

1. *UIUC*

UIUC's asp-based software does not support the initiate or terminate operations. Therefore, these actions are all manually mediated via email exchanges between the administrator and potential static data providers.

2. *Ghent*

(N.B. Patrick Hochstenbach is the srepod software developer)

The srepod software could have:

- Better support for namespace declarations
- Support for resumptionTokens
- Easier validation and feedback
- Easier upgrade and installation
- Support for local static repositories

3. *OAIScotland*

Updating to new version of srepod had some problems with missing files.

2.4 **Other comments from the gateway administrators**

1. *UIUC*

- a) "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."
- b) "We took one down static repository because the provider had developed their own full-fledged data provider and didn't need the static one any longer."
- c) "Note that we do not actively advertise this gateway. It was originally developed to support data provider partners for an IMLS research grant we had. However, knowledge of our gateway has informally spread and we do accept new registrations."

2. *Ghent*

“I'm working now on a Java port of the Static Repository software that might interest you. This could also be a way to incorporate your branding work in a new version.”

3 Summary

Static repositories currently operate on a small scale when compared to the deployment of full OAI repositories; they do however, allow around 80 repositories to participate in OAI-based services and fill a necessary gap in the information environment. These repositories are made available through four static repository gateways. The LANL gateway is the most visible and active, with the others averaging 7 mediated repositories each. The number of static repositories is growing steadily with an average of 3 repositories being added to each gateway per year (NB, growth figures for LANL unknown). All of the gateways exist in research environments, but some of them identify themselves as service-providers as well.

The administrative and technical commitment to running a gateway is quite small – administering repositories typically requires 2 days a year and supporting and developing them averages 8 days a year (including server backup and maintenance). The gateway does not need its own server however, and the server support costs could be shared by all the applications on the server. Although many of the gateways were established under development grants, their ongoing costs are absorbed into general budgets.

It should be noted that both pieces of gateway software in use are still under development. Their deployment in their current form assumes a degree of technical competence and requires more human involvement in repository registration than might be the case when more mature software is released.

UIUC's use of static repositories as a teaching tool for their Library and Information Science students is notable as an illustration of how simple static repositories are to create.

Appendix 1 Sample Questionnaire

1) Questions about administration

- a) How many static repositories do you mediate (discounting those obviously created for test purposes)?
- b) Approximately how many static repositories do you register per annum?
- c) Approximately how many man-days do you spend on administering the gateway per year? - approving static repository registration etc.?
- d) Do you know how frequently your static repositories typically update (weekly, monthly, quarterly, annually, never)?
- e) Are you operating as a research gateway or as service provider?
- f) Are you aware of any gateway operating with any level of formal guaranteed service commitment?
- g) Do you offer any informal service guarantee for your gateway?
- h) Is supporting and running your gateway a specific budget item or is it cost that is absorbed into another budget (general, research, IT, other)?
[if you are willing to share specific budget details they will only made available to JISC to inform future development in this area]

2) Questions about technical support

- a) What software do you use?
 - LANL srepod
 - UIUC asp SR Gateway
 - Other _____
- b) Have you customized the gateway software that you use?
- c) Do you run other applications on your gateway's server?
- d) Approximately how many man-days do you spend maintaining your server per year (including any standard server maintenance such as backups)?
- e) Are you aware of any outstanding technical issues with the gateway software that you use?