

Stargate project extension

Static Repository Gateway Software Installation

Author: Alan Dawson

Date: February 2007

Version: 1.1

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



1	Introduction	2
2	Hardware prerequisites	3
3	Obtain software	3
4	Software prerequisites	3
4.1	Required Software	3
4.2	Optional recommended software.....	3
5	Copy software to Unix server.....	3
6	Install software	4
6.1	Unzip the software.....	4
6.2	Set the installation directory	4
6.3	Install the software	4
7	Configure software	7
7.1	Establish a link between the web server and the gateway software	7
7.2	Edit the main configuration file.....	7
7.3	Edit the HTML files	8
8	Set permissions.....	9
9	Start the gateway software.....	10
10	Register a static repository	11
11	Testing repository registration.....	11
12	Troubleshooting.....	12
12.1	XML errors.....	12
12.2	BaseURL validation	12
12.3	Email notifications.....	12
12.4	Log files.....	13
12.5	Restarting the gateway	13
13	Harvesting a static repository	14
14	Updating a static repository	15
15	Terminating a static repository	15
16	Further information	16

1 Introduction

This document, produced as part of an extension to the Stargate project, is a guide to installing the experimental Static Repository Gateway software developed by Patrick Hochstenbach, Henry Jerez and Herbert Van de Sompel. The software was developed as a prototype at the Los Alamos National Laboratories to demonstrate an implementation of the Open Archives Initiative's Specification for an OAI Static Repository and an OAI Static Repository Gateway. Subsequently, Patrick Hochstenbach has continued to support its development. The software 'sreped' is available as open source software through Sourceforge (<https://sourceforge.net/projects/sreped/>).

The software continues to be under periodic development; one consequence of this is that an update was released during the final stages of the creation of this guide. As a result, this guide does not refer to the most recent version of the software. It should be noted however, that the new software release contained a few bug fixes and cosmetic changes only – the installation process and requirements will have largely remained the same.

Although this guide aims to be accessible to interested novice readers, a degree of familiarity with Linux or Unix operating systems is essential before any actual installation is attempted. As the installation of the software requires root (i.e. full/ system administrator) privileges on the server, the guide's target audience is therefore users with such knowledge and experience (those without such familiarity should **not** have root access or attempt to install this software).

It should be noted that although the software is functional, effective, and supports the real use of OAI static repositories, it is still under development and should be considered as an early beta version. Consequently, as indicated in the following guide, installing and maintaining the software may involve interaction with the software developer(s).

R. John Robertson

Stargate Project Manager

January 2007

2 Hardware prerequisites

The hardware requirements of the software are minimal, a demonstration version ran happily, though slowly, on a Pentium 3; consequently, the lowest specification of a new desktop machine should be more than adequate. The typical cost of such a machine might be £600.

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.

3 Obtain software

Download software from

```
https://sourceforge.net/projects/srepod/
```

The downloaded file will be called srepod-2.1.66.tar.gz or something similar.

4 Software prerequisites

4.1 Required Software

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)

4.2 Optional recommended software

Pico or PFE or other text editor

5 Copy software to Unix server

This can be done via FTP, HTTP or drive mapping from a Windows system. For this project, Samba software was used to facilitate drive mapping from Windows PCs to Unix.

6 Install software

6.1 Unzip the software

Two different directories should be used; one for unzipping the software and one for installation. The `gunzip` command will create one of these automatically. For example, to use

```
/opt to unpack the software and
```

```
/opt/stargate to install the software,
```

issue the commands:

```
cd /opt
mkdir stargate
gunzip -c srepod-2.1.6.tar.gz | tar xvf -
```

This will create a directory called `srepod-2.1.6` to match the zipped file name. It may be useful to rename this directory to something simpler, e.g. `'srepod'`, using the command

```
mv srepod-2.1.6 srepod
```

The examples below assume that the directory has been renamed to `'srepod'`.

6.2 Set the installation directory

Edit the file `'Makefile'` in the `'src'` directory, using `pico` or `PFE` or a similar text file editor, e.g.

```
cd /opt/srepod/src
pico Makefile
```

and change the line

```
INSTDIR=/opt/srepod-2-1-6
```

to

```
INSTDIR=/opt/stargate
```

where `'stargate'` is the directory created earlier. If using `pico`, press `Ctrl X` to exit and `Y` to save the file.

6.3 Install the software

This process is documented in the `INSTALL` file. First ensure you are in the correct directory, e.g. by

```
cd /opt/srepod/src
```

and then issue the following three commands:

```
make clean
make
make install
```

There will be numerous status messages, e.g. as illustrated below:

```

checking dlfcn.h usability... yes
checking dlfcn.h presence... yes
checking for dlfcn.h... yes
checking for ranlib... ranlib
checking for strip... strip
checking for objdir... .libs
checking for gcc option to produce PIC... -fPIC
checking if gcc PIC flag -fPIC works... yes
checking if gcc static flag -static works... yes
checking if gcc supports -c -o file.o... yes
checking if gcc supports -c -o file.lo... yes
checking if gcc supports -fno-rtti -fno-exceptions... yes
checking whether the linker (/usr/bin/ld) supports shared libraries... yes
checking how to hardcode library paths into programs... immediate
checking whether stripping libraries is possible... yes
checking dynamic linker characteristics... GNU/Linux ld.so
checking if libtool supports shared libraries... yes
checking whether to build shared libraries... yes
checking whether to build static libraries... yes
checking whether -lc should be explicitly linked in... no
creating libtool
checking for an ANSI C-conforming const... yes
checking for stdlib.h... (cached) yes
checking for string.h... (cached) yes
checking sys/file.h usability... yes
checking sys/file.h presence... yes

```

Most of these messages indicate successful operation and can be ignored. However, it is possible that errors may occur, probably owing to the absence of a prerequisite item of software. If there are any error messages they need to be investigated before proceeding with the next 'make' command. For example, the following messages include an error:

```

rm -fr .libs/libgdbm.la .libs/libgdbm.* .libs/libgdbm.*
gcc -shared gdbmopen.lo gdbmdelete.lo gdbmfetch.lo gdbmstore.lo gdbmclose.lo gd
bmreorg.lo gdbmseq.lo gdbmsync.lo gdbmerrno.lo gdbmexists.lo gdbmfdesc.lo gdbmse
topt.lo bucket.lo falloc.lo findkey.lo global.lo hash.lo update.lo version.lo
-Wl,-soname -Wl,libgdbm.so.3 -o .libs/libgdbm.so.3.0.0
(cd .libs && rm -f libgdbm.so.3 && ln -s libgdbm.so.3.0.0 libgdbm.so.3)
(cd .libs && rm -f libgdbm.so && ln -s libgdbm.so.3.0.0 libgdbm.so)
ar cru .libs/libgdbm.a gdbmopen.o gdbmdelete.o gdbmfetch.o gdbmstore.o gdbmclose.o
gdbmreorg.o gdbmseq.o gdbmsync.o gdbmerrno.o gdbmexists.o gdbmfdesc.o gdbmse
topt.o bucket.o falloc.o findkey.o global.o hash.o update.o version.o
ranlib .libs/libgdbm.a
creating libgdbm.la
(cd .libs && rm -f libgdbm.la && ln -s ../libgdbm.la libgdbm.la)
makeinfo ./gdbm.texinfo
make[1]: makeinfo: Command not found
make[1]: *** [gdbm.info] Error 127
make[1]: Leaving directory `/opt/srepod/src/gdbm-1.8.3'

```

This particular error has occurred owing to the absence of the makeinfo program from the system. The solution in this case is to ensure that the makeinfo program (which is part of the texinfo package) is installed on the system and then rerun the 'make' commands.

If other errors occur during installation it may be possible to diagnose and correct them with the help of technical staff. If not then it will probably be necessary to request email support from the software provider.

For example, the error messages below indicated a problem with the validate section of the installation process.

```
make[1]: Leaving directory `/opt/srepod/src/registry'
Making validate
make -C validate install INCDIR=/opt/stargate/include LIBDIR=/opt/stargate/lib
INDIR=/opt/stargate/bin
make[1]: Entering directory `/opt/srepod/src/validate'
gcc -O -I/opt/stargate/include hstrip.c -o hstrip
gcc -O -I/opt/stargate/include hasout.c -o hasout
gcc -O -I/opt/stargate/include validurl.c -o validurl -l expat
/usr/bin/ld: cannot find -lexpat
collect2: ld returned 1 exit status
make[1]: *** [validurl] Error 1
make[1]: Leaving directory `/opt/srepod/src/validate'
make: *** [val] Error 2
```

The only way to resolve this particular error was to contact the software author, who was able to provide a new version of the Makefile program for the validate process. After this new file was copied to the 'validate' directory, the 'make clean' and 'make' commands were run again, successfully, after which the 'make install' command also ran successfully. It should be noted however, that this particular problem has been addressed in the 2.1.6 release of the software.

7 Configure software

This stage is documented in the INSTALL file.

7.1 Establish a link between the web server and the gateway software

This is done via two symbolic link commands, e.g.

```
cd /var/network/www/cgi-bin/  
ln -s /opt/stargate/cgi-bin gateway
```

and

```
cd /var/network/www/htdocs/  
ln -s /opt/stargate/htdocs/ gateway
```

where

```
/var/network/www/htdocs/
```

is the top-level web server directory, and

```
/opt/stargate/
```

is the directory for the static repository gateway installation.

7.2 Edit the main configuration file

The main configuration file for the gateway software is called environ. To edit this, type a command such as:

```
pico /opt/stargate/cgi-bin/environ
```

and then change the lines beginning with the following text:

```
export GWADMIN=  
export GWBASEURL=  
export GWREGISTRY=  
export GWTERMINATOR=  
export GWHOME=
```

e.g.

```
export GWADMIN="cdlr@strath.ac.uk"  
export GWHOME=/opt/stargate
```

The entry for the machine name 'localhost:8080' should be changed to the name of the relevant web server in the other three lines, e.g. change:

```
export BASEURL="http://localhost:8080/cgi-bin/gateway/gateway.cgi"
```

to

```
export BASEURL="http://stargate.cdlr.strath.ac.uk/cgi-  
bin/gateway/gateway.cgi"
```

7.3 Edit the HTML files

The two files registry.html and terminator.html will also need to be checked, and may need to be edited.

These are stored in the htdocs directory. For example:

```
cd /opt/stargate/htdocs
pico registry.html
```

The line

```
<form action="/cgi-bin/gateway/registry.cgi method="GET">
```

should be correct, provided that the cgi-bin directory is accessible via the web server. For example, using a web browser to connect to:

```
http://stargate.cdlr.strath.ac.uk/cgi-bin/
```

should produce the message

```
Forbidden
You don't have permission to access /cgi-bin/ on this server.
```

If instead it gives the message:

```
Not Found
The requested URL /cgi-bin/ was not found on this server.
```

that means that the reference to the cgi-bin directory in the registry.html file will need to be changed.

8 Set permissions

It is important that the access permissions are set correctly, otherwise repository registration will not succeed, even though the software appears to be working.

A Unix shell script called `chmod.sh` is provided with the gateway software to make it easy to set permissions, but this requires root access to run successfully and securely, e.g.

```
cd /opt/stargate
su
sh chmod.sh
```

This will produce the following two prompts:

```
Enter the user name or id of the process running cgi scripts: [nobody]
Enter the group name or id of the process running cgi scripts: [nobody]
```

The user and group name 'nobody' are suggested, but these may not be available or appropriate on every Unix system (on some systems, 'nobody' is the user name and 'nogroup' is the group name). It is best to use the same user and group name as the Apache web server. The system administrator should be able to advise on the user and group name to use.

9 Start the gateway software

The gateway software is called srepod, and it is activated using the daemonctl program in the bin folder, e.g.

```
cd /opt/stargate/bin
./daemonctl start
```

Depending on the system configuration, this may require privileges of the root user. Once started the gateway can simply be left running. However, if necessary it can be stopped with the commands:

```
cd /opt/stargate/bin
./daemonctl stop
```

And then restarted as above.

10 Register a static repository

Once the gateway software has been successfully installed and started, a static repository can be registered. The easiest way to do this is by completing the form on the registration page, e.g. at

```
http://stargate.cdrl.strath.ac.uk/gateway/
```

or equivalent address. All that is needed to complete registration are the URL of the static repository XML file and the email address of the static repository administrator.

It is also possible to register a static repository by issuing a URL from a web browser. For example, if the gateway is at <http://stargate.cdrl.strath.ac.uk/> and the repository XML file is at

<http://cdlr.strath.ac.uk/stargate/repositories/sr2.xml> then the repository could be registered via the following URL:

```
http://stargate.cdrl.strath.ac.uk/cgi-bin/gateway/gateway.cgi
?initiate=http://cdlr.strath.ac.uk/stargate/repositories/sr2.xml
```

Subsequent message screens will indicate whether or not the registration has been successful.

11 Testing repository registration

Once a static repository has been successfully registered, it should function as any valid OAI-PMH repository and be able to respond to OAI-PMH requests. The easiest way to test this is by issuing the OAI Identify command from a web browser. For example, the URL:

```
http://stargate.cdrl.strath.ac.uk/cgi-
bin/gateway/gateway.cgi/cdrl.strath.ac.uk/stargate/repositories/sr3.xml?verb=Identif
y
```

would issue the Identify command to the static repository called `sr3.xml` that is available at <http://cdlr.strath.ac.uk/stargate/repositories/sr3.xml>

and had been registered with the gateway at

```
http://stargate.cdrl.strath.ac.uk/
```

Note that the 'http://' prefix in the URL of the XML file is omitted from the resulting URL.

If the Identify command is successful then details of the repository should be displayed in XML format in the web browser. The repository is then available for harvesting via OAI-PMH.

12 Troubleshooting

Once the gateway software has been successfully installed, most subsequent problems are likely to be due to the content of the static repository itself.

12.1 XML errors

All static repositories should be well-formed XML files. If metadata held in a database or other well-structured form then creating an XML file in OAI-PMH static repository format is relatively easy. However, it is common for XML files to contain data errors unless they have been validated. The easiest way to test whether an XML file will be accepted for the static repository gateway software is to open it in Internet Explorer, which will indicate the position of any character set errors. Such errors are usually due to the presence of smart quotes, long hyphens, accented letters or other non-standard characters. While these can easily be corrected by manually editing the XML file, it is better to set up a procedure to correct them in the source metadata or automatically as the XML file is created.

12.2 BaseURL validation

Part of the specification of a static repository is that: *'The value of the baseURL element of the Static Repository must match the Static Repository base URL at which the Static Repository is available through intermediation from a Static Repository Gateway.'* For example, if the static repository gateway is at

```
http://stargate.cdrl.strath.ac.uk/cgi-bin/gateway/gateway.cgi
```

and a static repository XML file is at:

```
http://cdlr.strath.ac.uk/stargate/repositories/sr3.xml
```

then the XML file should contain the following entry:

```
<oai:baseURL>  
http://stargate.cdrl.strath.ac.uk/cgi-  
bin/gateway/gateway.cgi/cdrl.strath.ac.uk/stargate/repositories/sr3.xml  
</oai:baseURL>
```

If the XML file does not contain such a matching entry, it is still possible to register the XML file with the gateway provided that URL validation is switched off. This is done by editing the 'environ' file and setting the value of the relevant variable to 0, i.e.

```
export GWVALIDBASEURL=0
```

Switching off validation in this way can be useful for testing, but is not advisable for a gateway service, as URL validation is important for the termination process, as well as being required for compliance with the specification.

12.3 Email notifications

The gateway software web interface asks for the email address of the static repository administrator.

However, at present this feature is not activated, so no email messages will be sent by the gateway software.

12.4 Log files

If problems arise when using the gateway software, the log files can be useful sources of information for diagnosing problems. The log files are stored in the spool directory. Although these files can get quite large once a gateway has been in use for several months, it is easy to see the most recent entries by looking at the end of the files, e.g.

```
cd /opt/stargate/spool
tail srepod.log
tail registry.log
tail -50 gateway.log
```

12.5 Restarting the gateway

The gateway software can be stopped with the commands:

```
cd /opt/stargate/bin
./daemonctl stop
```

And then restarted with the command:

```
./daemonctl start
```

Inspection of the log files should indicate the nature of any errors.

If for any reason the stop command does not work, try deleting the srepod.pid file in the spool directory. A new one will be created when it restarts.

13 Harvesting a static repository

Once a static repository has been registered with a gateway then it is available for harvesting by aggregation services in the same way as any other repository that can be accessed via OAI-PMH. For example, to register the sr3.xml static repository for harvesting from the Stargate gateway via the ARC harvester, the following URL would be entered in the baseurl field of the dplist table in the ARC database:

```
http://stargate.cdlr.strath.ac.uk/cgi-  
bin/gateway/gateway.cgi/cdlr.strath.ac.uk/stargate/repositories/sr3.xml
```

As static repositories are by their nature fairly static, regular harvesting is unnecessary. However, if harvester software is set to carry out regular harvesting this should not matter, as correct operation of the OAI-PMH protocol should ensure that only new or updated records are harvested from each repository.

14 Updating a static repository

A static repository does not have to be entirely static. It is quite reasonable and permissible to update a static repository, e.g. to correct metadata errors or to include additional records. All that is needed to update a static repository is to copy a new version of the XML file to the same web server, with exactly the same file name, so that the URL is unchanged. It is not necessary to re-register the updated file with the gateway. However, when the updated static repository is accessed via an OAI-PMH request, e.g. for harvesting, then the OAI-PMH protocol should ensure that the new or updated records are retrieved. When this happens there will be a short delay in accessing the repository via the gateway, which is likely to display the following message in response to a manual OAI-PMH request:

```
Error 503 (Service unavailable): the Static Repository Gateway is busy
processing an update of the Static Repository
```

As there is a recommended size limit of 5000 records for a static repository, it is possible that such a limit will be reached for some collections. If this happens then the easiest solution is to create a second static repository and make it available on the same web server but with a different file name. This should be registered with the gateway as a new repository. It can then be harvested alongside the metadata from the first repository, so that all records can be retrieved together via an aggregating service.

15 Terminating a static repository

The simplest way of terminating registration of a static repository is to enter the URL of the repository XML file into the termination page of the gateway interface, e.g. that at:

```
http://stargate.cd1r.strath.ac.uk/gateway/terminate.html
```

In addition, the XML file should be renamed or removed from the web server, or the baseURL entry in the XML file should be changed so that it no longer matches that of the gateway.

It is also possible to terminate registration via an HTTP instruction, or for the gateway administrator to run the `unregister.sh` command. Details are provided in the implementation guidelines referred to below.

16 Further information

Original documentation on the gateway software is available along with the software, from:

<http://srepod.sourceforge.net/>

The main source of information is the INSTALL file. After installation the srepod log files can provide useful additional information.

Implementation guidelines are available via the Open Archives web site:

<http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>

There is likely to be a prominent link to the guidelines from any active gateway.

Email support may be available from the author of the gateway software: Patrick.Hochstenbach@ugent.be or from the author of this document: alan.dawson@strath.ac.uk but this is not guaranteed.

Metadata records harvested from multiple sources can be retrieved via various aggregating services, such as:

OAISTER: <http://oaister.umdl.umich.edu/>

Eprints UK: <http://eprints-uk.rdn.ac.uk/>

IRIScotland: <http://cdlr.strath.ac.uk/iriscotland/>

Both Eprints UK and IRIScotland are pilot services, not fully operational services.

Google Scholar also uses OAI-PMH to harvest metadata records from numerous repositories along with other sources of information.