

# GREENSTONE OAI IMPLEMENTATION

Wednesday, May 22, 2013  
Kathmandu - Nepal



**Diego J. Spano**

 [dspano@prodigioconsultores.com](mailto:dspano@prodigioconsultores.com)

 [diegospano](#)

 [@diegospano](#)



# **THE OPEN ARCHIVES APPROACH**

# The Open Archives Approach

---

"The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content."

- Facilitates access to heterogeneous web-accessible material
- A low-barrier interoperability solution
- Based on repositories supporting
  - Metadata sharing
  - Publishing
  - Archiving
- Arose out of the e-print community
- Main features
  - Open Archives Initiative
  - OAI Protocol for Metadata Harvesting (OAI-PMH)

# The Open Archives Approach

---

To facilitate metadata harvesting there needs to be agreement on:

- Transport protocol - HTTP
- Metadata format - Dublin Core
- Metadata Quality Assurance - mandatory element set, naming and subject conventions
- Intellectual Property and Usage Rights

# OAI-PMH

---

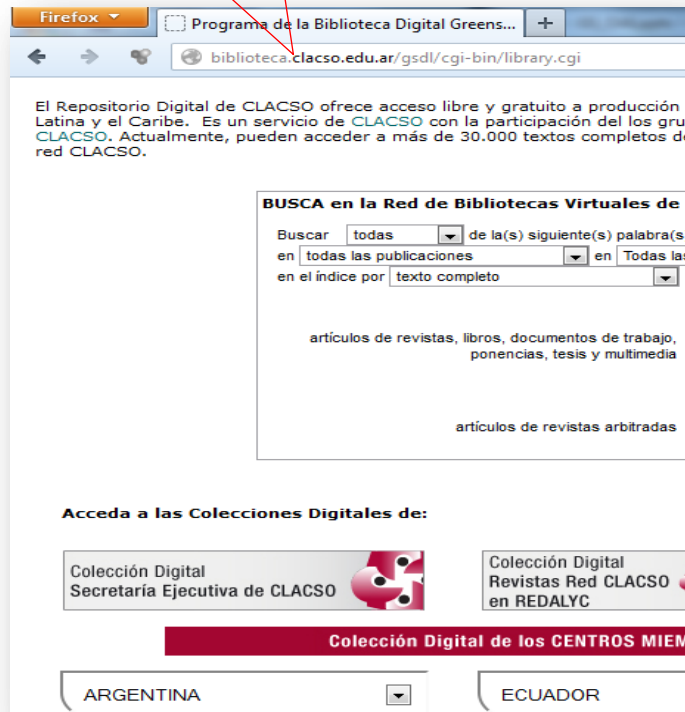
- A mechanism for harvesting
- Data providers make metadata available for harvesting
- Service Providers harvest metadata
- Metadata can be centrally collected or “aggregated”
- That’s all it is: a way to bring metadata together in one place!



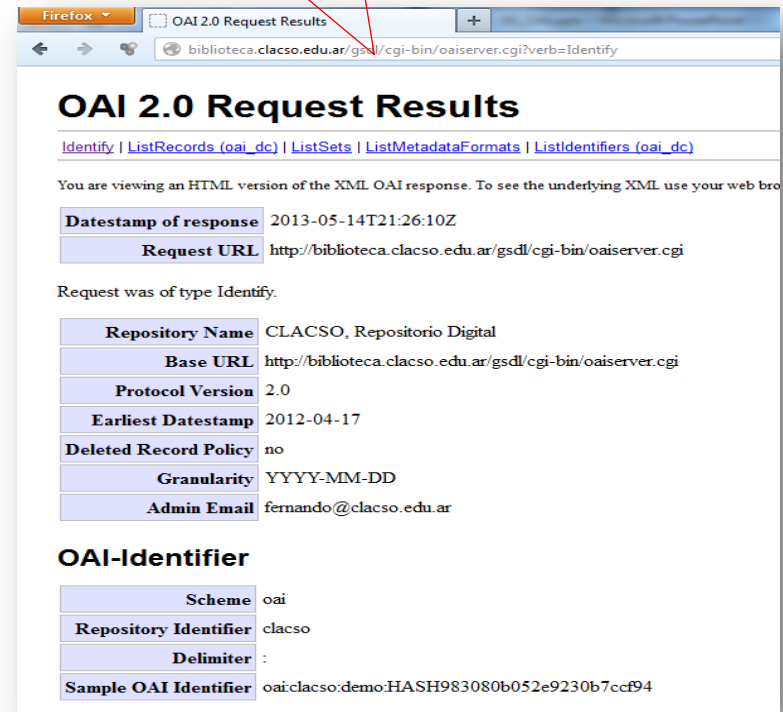
# **GaDP** **(GREENSTONE AS DATA PROVIDER)**

Greenstone comes with a built-in OAI data provider. This runs as a CGI program called "oaiserver.cgi". It can be accessed via the same URL as the Greenstone library (replacing "library.cgi" with "oaiserver.cgi").

<http://biblioteca.clacso.edu.ar/gsd/cgi-bin/library.cgi>



<http://biblioteca.clacso.edu.ar/gsd/cgi-bin/oaiserver.cgi>



- Configuration of the server is done via the *oai.cfg* file in the Greenstone *etc* directory.
- This file specifies general information about the repository, and lists collections to be made accessible to OAI clients. By default, collections are not accessible.
- Greenstone's OAI server currently supports Dublin Core, Qualified Dublin Core, RFC1807 metadata. For collections that use other metadata sets, including extracted metadata, metadata mapping rules should be provided to map the existing metadata to Dublin Core.
- Next ... Main parameters to configure





- **baseServerURL**: the URL of the web server (the domain and port number if necessary)  
baseServerURL “http://www.mylibrary.org”
- **oaiserverPath**: the relative path of the oaiserver.cgi (will be added to baseServerURL)  
oaiserverPath “/greenstone/cgi-bin/oaiserver.cgi”
- **libraryPath**: the relative path of the Greenstone library.cgi  
libraryPath “/greenstone/cgi-bin/library.cgi”
- **repositoryName**: The OAI repository name - human readable  
repositoryName “University of Waikato - Institutional Repository”
- **repositoryId**: The OAI repository domain name or unique identifier - used for generating the document identifiers, which will be like: oai:reposId:coll:oid  
repositoryId “waikato”
- **maintainer**: The central administrative email address for the repository  
maintainer “george@mylib.edu”
- **resumeafter**: How many records to output before needing a resumption token  
resumeafter 250

- **oaiinfo**: Free text items describing the repository.  
oaiinfo "Purpose" "This Repository has all the papers ..."  
oaiinfo "Size" "More than 5000 articles"
- **oaicollection**: A list of collections that should be made available to OAI clients. By default, collections are not available. There can be more than one line  
oaicollection thesis images videos
- **oaisetname / oaisetdescription**: Optional set name and description for each collection.  
oaisetname thesis "Thesis of Chemistry subjects"  
oaisetdescription thesis "Ph students' papers"
- **oaimetadata**: the list of metadata formats this server supports  
oaimetadata oai\_dc gsdl\_qdc
- Next ... how to map your own metadata fields



- Mapping rules allow you to match your own metadata schema to the right Dublin Core metadata. The format is:

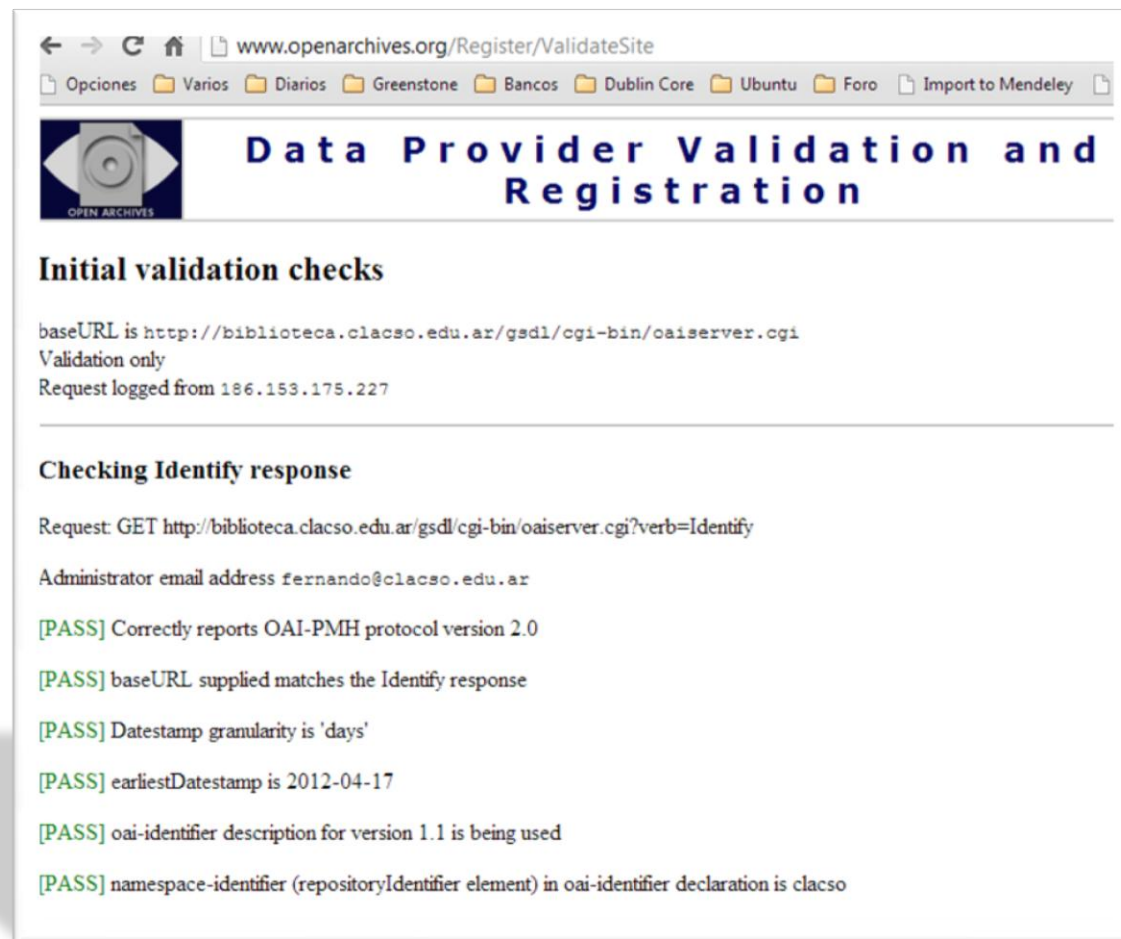
*oaimapping* <greenstone fieldname> <metadataPrefix.fieldname>

oaimapping	dls.Subject	oai_dc.coverage
oaimapping	nz.Region	oai_dc.coverage
oaimapping	nz.Region	gsdl_qdc.spatial

- Next ... Validate your site



- Go to <http://www.openarchives.org/data/registerasprovider.html> and validate your site to be sure it can be a data provider.



- Once your OAI server is operational, to provide maximal international visibility for your open access collections you should register them in at least one (and ideally all) of the following:
  - the ROAR directory (<http://roar.eprints.org/>)
  - the OAI directory (<http://www.openarchives.org/Register/BrowseSites>)
  - the OpenDOAR directory (<http://www.opendoar.org/>)

- Next ... A real case.



Firefox <http://biblioteca.clacso.edu.ar/gsd/cgi-bin/oaiserver.cgi>

## OAI 2.0 Request Results

[Identify](#) | [ListRecords \(oai\\_dc\)](#) | [ListSets](#) | [ListMetadataFormats](#) | [ListIdentifiers \(oai\\_dc\)](#)

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source

<b>Datestamp of response</b>	2013-05-15T16:51:55Z
<b>Request URL</b>	http://biblioteca.clacso.edu.ar/gsd/cgi-bin/oaiserver.cgi

Request was of type Identify.

<b>Repository Name</b>	CLACSO, Repositorio Digital
<b>Base URL</b>	http://biblioteca.clacso.edu.ar/gsd/cgi-bin/oaiserver.cgi
<b>Protocol Version</b>	2.0
<b>Earliest Datestamp</b>	2012-04-17
<b>Deleted Record Policy</b>	no
<b>Granularity</b>	YYYY-MM-DD
<b>Admin Email</b>	fernando@clacso.edu.ar

repositoryName

baseServerURL + oaiserverPath

maintainer

repositoryId

## OAI-Identifier

<b>Scheme</b>	oai
<b>Repository Identifier</b>	clacso
<b>Delimiter</b>	:
<b>Sample OAI Identifier</b>	oai:clacso:demo:HASH983080b052e9230b7ccf94

oaiinfo

## Description

<b>Documentos</b>	mas de 25000 documentos a texto completo
<b>Proposito</b>	Producción académica y científica de la red CLACSO (centros miembros y Secretaria Ejecutiva), trabajo e informes de investigación, ponencias de congresos, tesis y producciones multimedia (audio y video).

Firefox OAI 2.0 Request Results +

← → biblioteca.clacso.edu.ar/gsd/cgi-bin/oaiserver.cgi?verb=ListSets

## OAI 2.0 Request Results

[Identify](#) | [ListRecords \(oai\\_dc\)](#) | [ListSets](#) | [ListMetadataFormats](#) | [ListIdentifiers \(oai\\_dc\)](#)

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view s

<b>Datestamp of response</b>	2013-05-15T17:02:12Z
<b>Request URL</b>	http://biblioteca.clacso.edu.ar/gsd/cgi-bin/oaiserver.cgi

Request was of type ListSets.

### Set

<b>setName</b>	ar/ar-001
<b>setDescription</b>	Centro de Estudios Avanzados - CEA/UNC
<b>setSpec</b>	ar/ar-001 <a href="#">Identifiers</a> <a href="#">Records</a>

### Set

<b>setName</b>	ar/ar-003
<b>setDescription</b>	Centro de Estudios Históricos de Estado, Política y Cultura - CEHEPYC-UNCo
<b>setSpec</b>	ar/ar-003 <a href="#">Identifiers</a> <a href="#">Records</a>

### Set

<b>setName</b>	ar/ar-008
<b>setDescription</b>	Centro de Población, Empleo y Desarrollo - CEPED/UBA
<b>setSpec</b>	ar/ar-008 <a href="#">Identifiers</a> <a href="#">Records</a>

oaicollection

oaisetname

oaisetdescription

Firefox OAI 2.0 Request Results

[biblioteca.clacso.edu.ar/gsdll/cgi-bin/oaiserver.cgi?verb=ListRecords&metadataPrefix=oai\\_dc&set=ar/ar-001](#)

## OAI 2.0 Request Results

[Identify](#) | [ListRecords \(oai\\_dc\)](#) | [ListSets](#) | [ListMetadataFormats](#) | [ListIdentifiers \(oai\\_dc\)](#)

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source option. More info

**Datestamp of response** 2013-05-15T17:08:54Z

**Request URL** <http://biblioteca.clacso.edu.ar/gsdll/cgi-bin/oaiserver.cgi>

Request was of type ListRecords.

**OAI Record: oai:clacso:ar/ar-001:D8024**

**OAI Record Header**

<b>OAI Identifier</b>	oai:clacso:ar/ar-001:D8024	<a href="#">oai_dc</a>	<a href="#">formats</a>
<b>Datestamp</b>	2013-04-08		
<b>setSpec</b>	ar/ar-001	<a href="#">Identifiers</a>	<a href="#">Records</a>

**Dublin Core Metadata (oai\_dc)**

<b>Author or Creator</b>	cea-unc
<b>Other Contributor</b>	Castagno, Pablo - Autor/a
<b>Date</b>	2012-12-12
<b>Description</b>	he aim of this work is to analyze the relations between the cultural changes inserted in the Political in constituting a global society. It will begin with an introduction of the globalization processes in or substratum. Subsequently, according to Renato Ortiz, mondialisation will be analyzed as a deeper of the constitution of a global society. Afterwards, the relations between the mondialisation proces trying to give an answer to the following questions: What does the relation between the Representa globalization-mondialisation process?; Are the Representation cultural changes of part of the monc have reservations about the orthodox neoliberal model, do they use in their symbolic construction Representation changes are analyzed according to the cultural guidelines they generate and not dep

oaimapping





# **GaSP**

## **(GREENSTONE AS SERVICE PROVIDER)**

- GLI can serve as an OAI client application: it can connect to a remote OAI server and retrieve metadata and even download documents.

You have three methods for doing this:

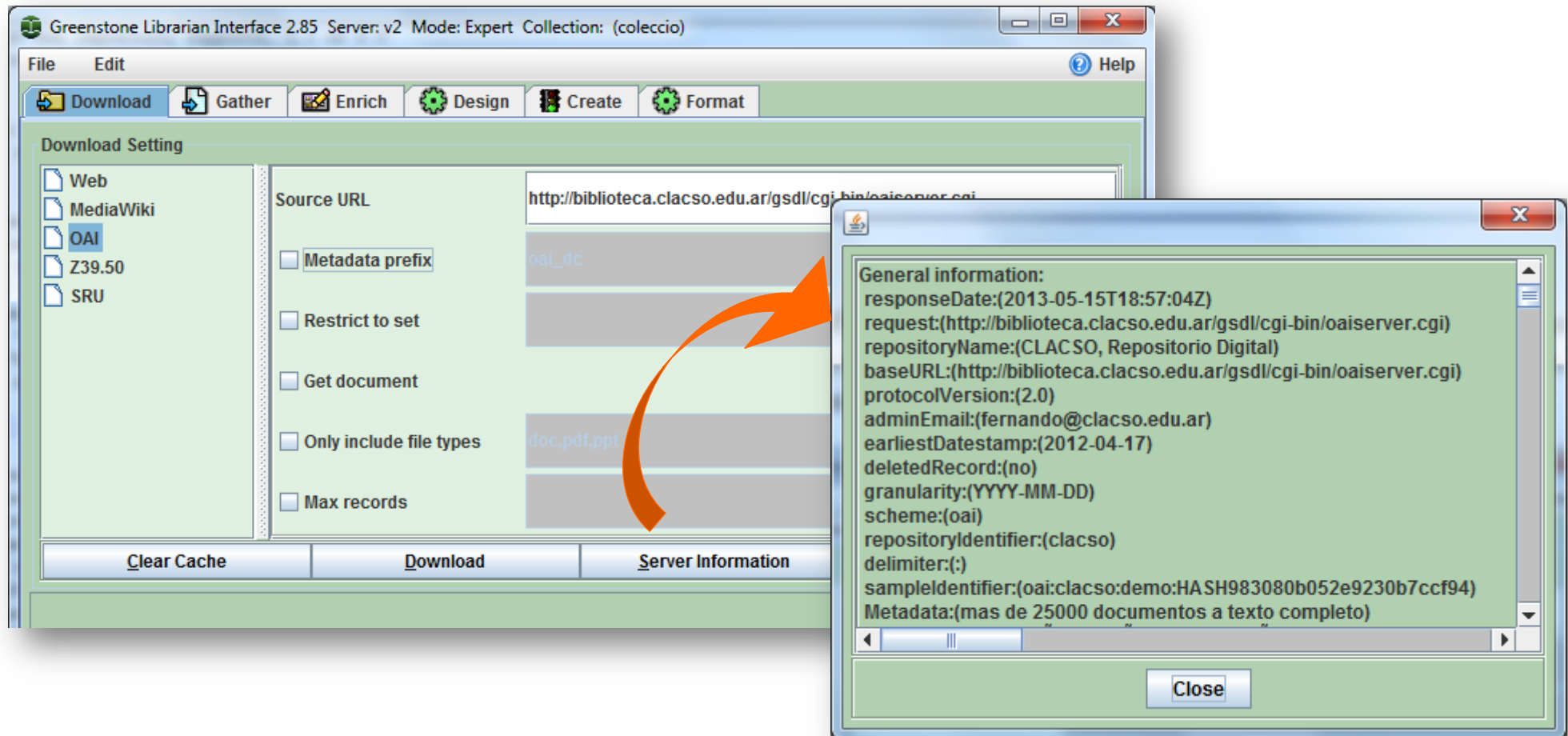
- Using GLI (Greenstone Librarian Interface)
- Using command line from within Greenstone environment
- Using a third party application that allows you to download OAI records.

We'll focus on the first two methods.

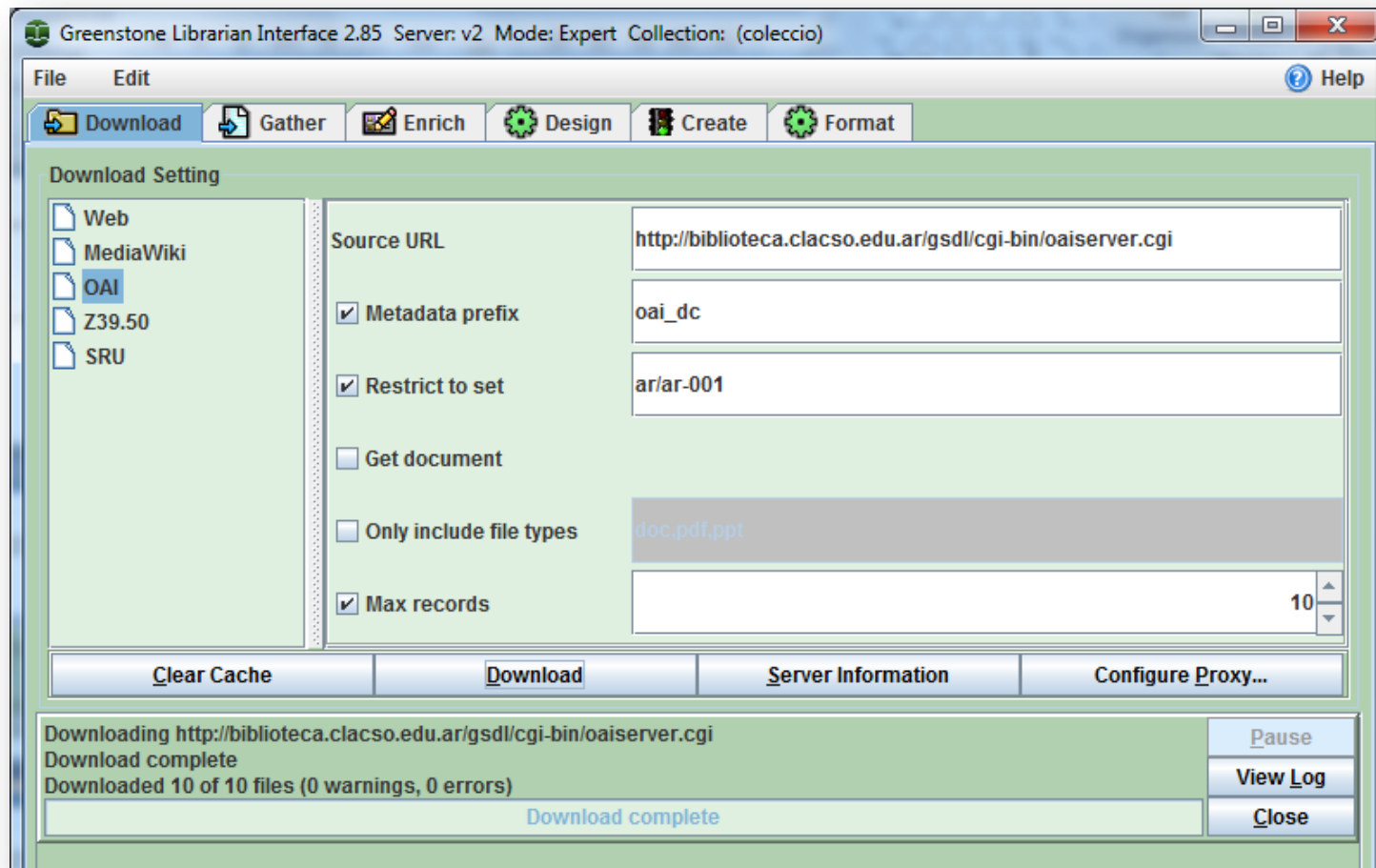
- Next ... Using GLI to download OAI records



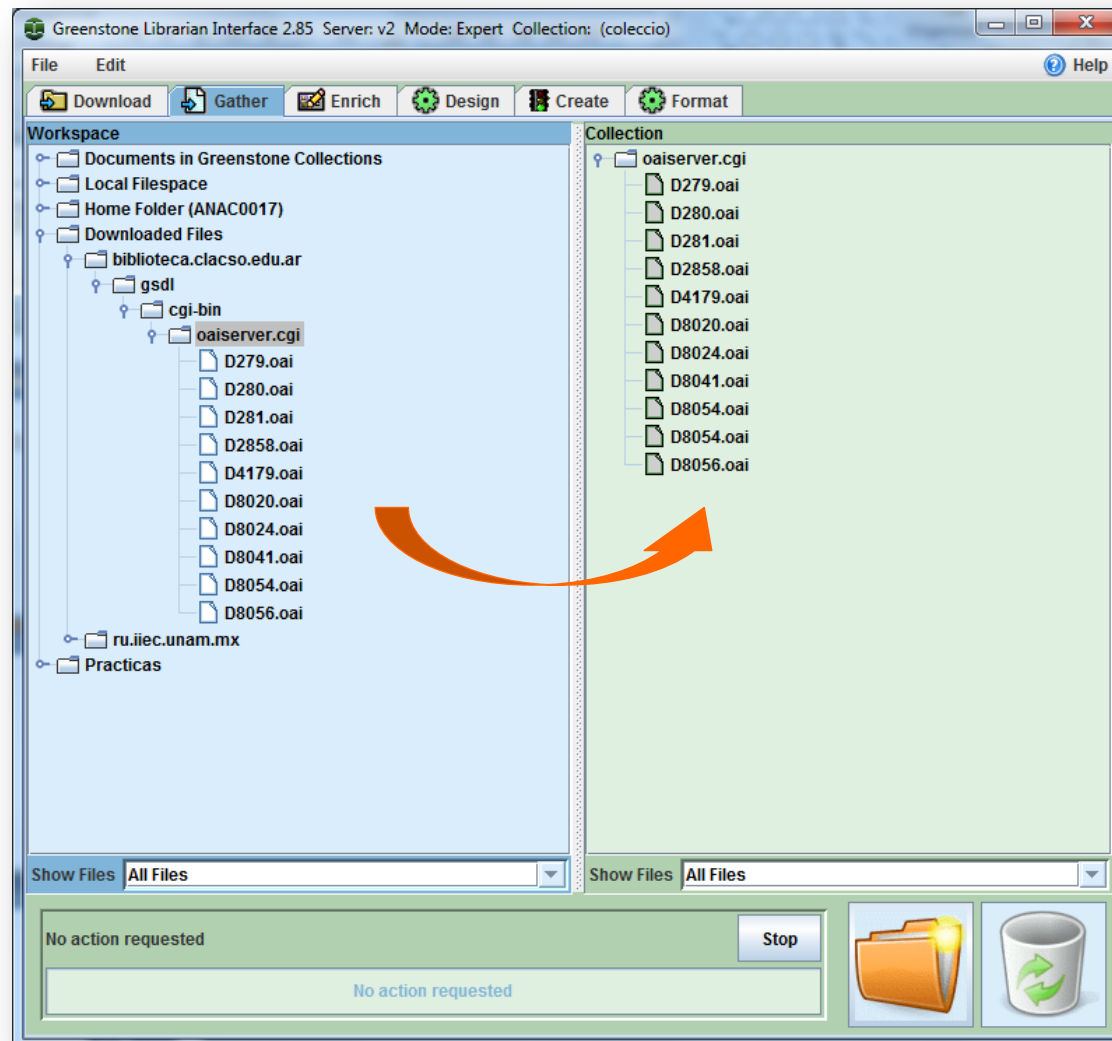
- GLI offers a “Download” tab for downloading documents and/or records. Assuming that you know the URL for the external repository, you can know the rest of the parameters by clicking “Server Information”, i.e.. the name of the sets.



- Once you filled the required fields just press the Download button.



- The downloaded records will appear in the “Gather” tab under “Downloaded Files” section. Drag them to the right panel to get the records ready for building the collection.



- OAI records imported into a Greenstone collection are saved as Greenstone archive format, similar to XML format but with some extra data

```
1 <?xml version="1.0" encoding="utf-8" standalone="no"?>
2 <!DOCTYPE Archive SYSTEM "http://greenstone.org/dtd/Archive/1.0,
3 <Archive>
4 <Section>
5 <Description>
6   <Metadata name="gsdlsourcefilename">import\oaiserver.cgi\D8(
7   <Metadata name="gsdldoctype">indexed_doc</Metadata>
8   <Metadata name="Source">D8054.oai</Metadata>
9   <Metadata name="SourceFile">D8054.oai</Metadata>
10  <Metadata name="Language">en</Metadata>
11  <Metadata name="Encoding">utf8</Metadata>
12  <Metadata name="Plugin">OAIPugin</Metadata>
13  <Metadata name="ex.dc.Identifier">http://biblioteca.clacso.
14  </Metadata>
15  <Metadata name="ex.dc.Date">2012-12-13</Metadata>
16  <Metadata name="ex.dc.Publisher">CLACSO, Consejo Latinoameri
17  <Metadata name="ex.dc.Publisher">CEA-UNC, Centro de Estudio
18  <Metadata name="ex.dc.Publisher">Córdoba ; Buenos Aires</Me
19  <Metadata name="ex.dc.Publisher">Argentina</Metadata>
20  <Metadata name="ex.dc.Subject">Enseñanza</Metadata>
21  <Metadata name="ex.dc.Subject">Investigación</Metadata>
22  <Metadata name="ex.dc.Subject">Libros de textos</Metadata>
23  <Metadata name="ex.dc.Creator">cea-unc</Metadata>
24  <Metadata name="prettymd">&lt;table width=100% cellpadding=
25  </Metadata>
26  </Description>
27  <Content>&lt;OAI-PMH xmlns=http://www.openarchives.org/OAI
28  xsi:schemaLocation=http://www.openarchives.org/OAI/2.0/
29  ....
30  </Content>
31  </Section>
32  </Archive>
```

Dublin Core metadata extracted as “ex” Schema

“prettymd” metadata: its value is a preformatted table (with style) Containing the full OAI record

- format DocumentText "[prettymd]" is all you need to display the OAI record...

Greenstone demo

HOMEHELPPREFERENCES

SearchDatesTitles

EXPAND TEXT

EXPAND CONTENTS

NO HIGHLIGHTING

DETACH

PRINT

dc.Creator	cea-unc
dc.Contributor	Mora Queipo, Ernesto - Autor/a
dc.Contributor	Ayala, Mario - Autor/a
dc.Coverage	Venezuela
dc.Date	2010-03-22
dc.Description	Venezuelan black organizations have advanced in a profound process of identity reconstruction, structuring new social movements in the context of hegemonic struggle and political-institutional change in the country since the late 1990 decade. This process is beyond the traditional defense of the right to cultural diversity to set a new ethnopolitical discourse aimed at the decolonization of the state and nation project under the Bolivarian revolutionary process. Our paper attempts to make an approach to the analysis of the main factors which have influenced in the identity construction of African descent in the period 1998-2008 in Venezuela. It is based on interviews with leaders and members of Afro Venezuelan organizations, and reviewing literature and various documents which were produced and disseminated by their activists.
dc.Format	application/pdf
dc.Format	24 p.
dc.Relation	277
dc.Identifier	ISSN : 1669-953X
dc.Identifier	URL : <a href="http://bibliotecavirtual.clacso.org.ar/ar/libros/argentina/cea/contra/4.5.6/ayala.pdf">http://bibliotecavirtual.clacso.org.ar/ar/libros/argentina/cea/contra/4.5.6/ayala.pdf</a>
dc.Identifier	Contra Relatos desde el Sur nos. 5-6 (diciembre 2008). Córdoba : CEA-UNC, 2008
dc.Language	Spa
dc.Publisher	CEA, Centro de Estudios Avanzados, UNC, Universidad Nacional de Córdoba
dc.Publisher	Córdoba
dc.Publisher	Argentina
dc.Rights	Reconocimiento-No comercial-Sin obras derivadas 2.0 Genérica
dc.Rights	<a href="http://creativecommons.org/licenses/by-nc-nd/2.0/deed.es">http://creativecommons.org/licenses/by-nc-nd/2.0/deed.es</a>
dc.Subject	Afro-venezolanos

## Using command line to download OAI records

Just a few step to do this task:

1. Open a terminal or console window
2. Go to Greenstone root folder
3. Run `setup.bat/setup.bash`
4. Move to the import folder of the target collection for the records
5. Run `perl -S downloadfrom.pl` (with the following options)
  - `download_mode OAI`
  - `url http://biblioteca.clacso.edu.ar/gsdh/cgi-bin/oaiserver.cgi`
  - `set ar/ar-001`
  - `max_record 10`
6. Build the collection as usual

The command “downloadfrom” will have the same results as the previous task executed through GLI



# GLOSSARY

---

**Harvester:** client application issuing OAI-PMH requests

**Repository:** network accessible server, able to process OAI-PMH requests correctly

**Resource:** object the metadata is “about”, nature of resources is not defined in the OAI-PMH

**Item:** component of a repository from which metadata about a resource can be disseminated. Has an unique identifier

**Record:** metadata in a specific metadata format

**Identifier:** unique key for an item in a repository

**Set:** optional construct for grouping items in a repository

## + INFO...

---

- For GS2, there are 3 tutorials:

[http://wiki.greenstone.org/wiki/index.php/Tutorial exercises#Setting up your Greenstone OAI Server](http://wiki.greenstone.org/wiki/index.php/Tutorial_exercises#Setting_up_your_Greenstone_OAI_Server)

[http://wiki.greenstone.org/wiki/index.php/Tutorial exercises#Connecting to an OAI server from GLI](http://wiki.greenstone.org/wiki/index.php/Tutorial_exercises#Connecting_to_an_OAI_server_from_GLI)

[http://wiki.greenstone.org/wiki/index.php/Tutorial exercises#Connecting to the Greenstone OAI server from the outside world](http://wiki.greenstone.org/wiki/index.php/Tutorial_exercises#Connecting_to_the_Greenstone_OAI_server_from_the_outside_world)

[http://wiki.greenstone.org/wiki/index.php/2.85 Release Notes#OAI downloading](http://wiki.greenstone.org/wiki/index.php/2.85_Release_Notes#OAI_downloading)

- For GS3, the same approximate information is merged into the following 2 tutorials:

[http://wiki.greenstone.org/wiki/index.php/GS3 Tutorial exercises#Setting up your Greenstone OAI Server](http://wiki.greenstone.org/wiki/index.php/GS3_Tutorial_exercises#Setting_up_your_Greenstone_OAI_Server)

[http://wiki.greenstone.org/wiki/gsdoc/tutorial/g3-current/en/OAI downloading.htm](http://wiki.greenstone.org/wiki/gsdoc/tutorial/g3-current/en/OAI_downloading.htm)

**QUESTIONS**





**THANKS !!!**

**Diego Spano**

**[dspano@prodigioconsultores.com](mailto:dspano@prodigioconsultores.com)**

**[www.prodigioconsultores.com](http://www.prodigioconsultores.com)**