

# Open Geoportal **Harvester**

Patrick Florance, Tufts University  
Jeroen Ticheler, GeoCat

# What is it?

Open source web application that provides the automation of customized harvesting from partner metadata nodes as well as direct harvesting of XML metadata files within a web directory.

- Organization can create a custom OGP or Geonetwork instance that includes only selected records.
- Supports a variety of metadata formats
- Supports a variety of options to batch process and edit metadata
- Provides additional quality control options when uploading institutional metadata.

[Upload Metadata](#)[Ingest External Records](#)[Manage Ingests](#)[Process/Edit Metadata](#)[Delete Records](#)[Admin Page \(admin only\)](#)

## Ingest External Records

Use this form to ingest from chosen source directly into the local Solr instance

### Source Repository

Type of Instance:

Remote OGP Instance ▼

Catalog of Services:

Service name ▼

Url:

[Search Criteria](#) (optional) ? ▼

[Required Fields](#) (optional) ? ▼

[Schedule ingest ▶](#)

- Upload Metadata
- Ingest External Records
- Manage Ingests
- Process/Edit Metadata
- Delete Records
- Admin Page

Ingest External Records

Use this form to ingest from chosen source directly into the local Solr instance

Source Repository

Type of Instance:

Remote OGP Instance

Catalog of Services:

Service name

Custom OGP Instance Uri:

Search Criteria (optional)

Geographic Extent:

 ?


Theme Keyword:

Place Keyword:


Topic:

Select an ISO topic

Date range of content:



 to 



Originator:

Data Type:


Select Data Types

Data Repository:


Select Repositories

☐ Exclude Restricted Data

Filter records by Solr timestamp:



 to 



Custom Solr Query:

Need sample label text

OGP Required Fields (optional)

- ☐ Geographic Extent
- ☐ Theme Keyword
- ☐ Place Keyword
- ☐ Web Services
- ☐ Topic
- ☐ Date of Content
- ☐ Originator
- ☐ Data Type
- ☐ Data Repository

Schedule ingest

[Upload Metadata](#)[Ingest External Records](#)[Manage Ingests](#)[Process/Edit Metadata](#)[Delete Records](#)[Admin Page](#)

## Ingest External Records

Use this form to ingest from a chosen source directly into the local Solr instance

### Source Repository

Type of Instance:

WebDAV/WAF

WebDAV/WAF Uri:

**Note:** Service must be registered first. Register via the [admin page](#)

Name OGP Repository ?

Select Data Repositories

### Search Criteria

Filter records by last modified date: ?



to



### OGP Required Fields

- |  |  |  |
|--|--|--|
| <input type="checkbox"/> Geographic Extent | <input type="checkbox"/> Topic           | <input type="checkbox"/> Data Type       |
| <input type="checkbox"/> Theme Keyword     | <input type="checkbox"/> Date of Content | <input type="checkbox"/> Data Repository |
| <input type="checkbox"/> Place Keyword     | <input type="checkbox"/> Originator      |  |

Schedule ingest ▶

[Upload Metadata](#)
[Ingest External Records](#)
[Manage Ingests](#)
[Process/Edit Metadata](#)
[Delete Records](#)
[Admin Page](#)

## Ingest External Records

Use this form to ingest from a chosen source directly into the local Solr instance

### Source Repository

Type of Instance:

Geonetwork

Geonetwork Url:

Name OGP Repository

Select Data Repositories

### Search Criteria (optional) ?

Title: ?

Keyword: ?

Abstract text: ?

Free text: ?

Geonetwork Sources: ?

Select Repositories

### OGP Required Fields (optional) ?

- |  |  |   |
|--|--|---|
| <input type="checkbox"/> Geographic Extent | <input type="checkbox"/> Topic           | <input type="checkbox"/> Data Type          |
| <input type="checkbox"/> Theme Keyword     | <input type="checkbox"/> Date of Content | <input type="checkbox"/> Data Repository    |
| <input type="checkbox"/> Place Keyword     | <input type="checkbox"/> Originator      | <input type="checkbox"/> Creation/Edit date |

Schedule ingest ▶

[Upload Metadata](#)[Ingest External Records](#)[Manage Ingests](#)[Process/Edit Metadata](#)[Delete Records](#)[Admin Page](#)

## Ingest External Records

Use this form to ingest from a chosen source directly into the local Solr instance

### Source Repository

Type of Instance:

CSW ▾

CSW Url:

Name OGP Repository

Select Data Repositories ▾

### Search Criteria

Location: ?

Title: ?

Subject: ?

Free text: ?

Date range of content: ?

  to 

Custom CSW Query: ?

Need sample label text

### OGP Required Fields

- |  |  |  |
|--|--|--|
| <input type="checkbox"/> Geographic Extent | <input type="checkbox"/> Topic           | <input type="checkbox"/> Data Type       |
| <input type="checkbox"/> Theme Keyword     | <input type="checkbox"/> Date of Content | <input type="checkbox"/> Data Repository |
| <input type="checkbox"/> Place Keyword     | <input type="checkbox"/> Originator      |  |

[Schedule Ingest ▸](#)



Upload Metadata

**Ingest External Records**

Manage Ingests

Process/Edit Metadata

Delete Records

Admin Page

## Schedule Ingest

Use this form to schedule either single or repeating ingests

**Ingest Name:**

**Begin Ingest On:**



**Frequency:**

Once



◀ Ingest Query

Start ingest





Upload Metadata

Ingest External Records

**Manage Ingests**

Process/Edit Metadata

Delete Records

Admin Page

## Ingest Report, Name - 7/29



**Passed: 1645 Records**



**Warning: 545 Records**



**Failed: 923 Records**





Upload Metadata

Ingest External Records



**Manage Ingests**

Process/Edit Metadata

Delete Records

Admin Page

## Ingest Report, Name - 7/29



 **Passed:** 1645 Records 


Restricted Records: 745


Public Records: 120



Vector Records: 850

Raster Records: 845

 **Warning:** 545 Records 

Unrequired Fields Warnings: 345 

Web Service Warnings: 200 

 **Failed:** 923 Records 

☐ Required Fields Errors: 745 



☐ Web Service Errors: 120 

☐ System Errors: 58 

**Download Metadata**

[Upload Metadata](#)[Ingest External Records](#)[Manage Ingests](#)[Process/Edit Metadata](#)[Delete Records](#)[Admin Page](#)

## Ingest Report, Name - 7/29



 **Passed:** 1645 Records 


Restricted Records: 745


Public Records: 120

Vector Records: 850

Raster Records: 845

 **Warning:** 545 Records 

Unrequired Fields Warnings: 345 

Web Service Warnings: 200 

 **Failed:** 923 Records 



Required Fields Errors: 745 

☐ Geographic Extent: 245

☐ Theme Keyword: 105

☐ Place Keyword: 200

☐ Topic: 125

☐ Date of Content: 400

☐ Originator: 32

☐ Data Type: 400


☐ Data Repository: 320

☐ Creation/Edit date: 126



Web Service Errors: 120 



System Errors: 58 

[Download Metadata](#)

[Upload Metadata](#)[Ingest External Records](#)[Manage Ingests](#)[Process/Edit Metadata](#)[Delete Records](#)[Admin Page](#)

## Manage Ingests

Name ▼	Last Ingest ▼	Next Ingest ▼	Repositories	
<a href="#">New York Subway</a>	6/19/13	none	Tufts, Harvard, MIT, Berkley...	<a href="#">View Report</a>
<a href="#">Urban Development</a>	6/21/13	8/21/13 <a href="#">cancel</a>	Tufts, Harvard, MIT, Berkley...	<a href="#">View Report</a>
<a href="#">Road Systems</a>	6/22/13	in progress <a href="#">cancel</a>	Tufts, Harvard, MIT, Berkley...	<a href="#">View Report</a>
<a href="#">New York Subway</a>	6/28/13	none	Yale, MIT, Berkley...	<a href="#">View Report</a>
<a href="#">Urban Development</a>	7/9/13	none	Yale, MIT, Berkley...	<a href="#">View Report</a>
<a href="#">Road Systems</a>	7/13/13	8/13/13 <a href="#">cancel</a>	Tufts, Harvard, MIT, Berkley...	<a href="#">View Report</a>
<a href="#">New York Subway</a>	7/19/13	8/19/13 <a href="#">cancel</a>	Tufts, Harvard, MIT, Berkley...	<a href="#">View Report</a>
<a href="#">Urban Development</a>	8/1/13	9/1/13 <a href="#">cancel</a>	Yale, MIT, Berkley...	<a href="#">View Report</a>
<a href="#">Road Systems</a>	8/2/13	none	Yale, MIT, Berkley...	<a href="#">View Report</a>

- Upload Metadata
- Ingest External Records
- Manage Ingests
- Process/Edit Metadata
- Delete Records

Admin Page

Admin Page

Use this form to register either existing or new repositories

Register Existing Repository ?

Repository Name:

Harvard ▼

Add Institution

Register Custom Repository ?

Repository URL:

Repository Name:

Choose Service Type:

Remote OGP Instance ▼

Add Institution

My Repositories

	Remove
Harvard	X
MIT	X
Custom Repository 1	X
Custom Repository 2	X
Berkley	X

**Upload Metadata**[Ingest External Records](#)[Manage Ingests](#)[Process/Edit Metadata](#)[Delete Records](#)[Admin Page](#)**Metadata Ingest**

Use this page to ingest FGDC metadata files into GeoServer (Local layers only) and Solr.

**Institution (select one)**

Tufts ▼

☒ Ingest to GeoServer and Solr ☐ Ingest to GeoServer only ☐ Ingest to Solr only

Note: This option only applies to local layers. Ingest will not attempt to configure GeoServer for remote layers.

**Required fields**

☒ Title ☒ Publisher ☒ Originator ☒ Abstract ☒ Data Type

☒ Theme Keywords ☒ Place Keywords ☒ Content Date ☒ Bounds ☒ Access

Note: If you require a field, Solr ingest will fail for a data layer if the metadata does not contain a valid value. If you do not require the field, a warning will still be produced for invalid values.

Select XML metadata file(s) or zipped directory of XML metadata files

[+ Add files...](#)[Start ingest](#)

## • Metadata Formats

- FGDC, ISO 19139, OSGEO extended Dublin Core MEF (Metadata Exchange Format) – Geonetwork, MARC

## • Support for ArcGIS Server

## • Additional Error Checking

- Online link required
- Theme keyword - must have ISO theme keyword
- Improved error list
- Validate but not ingest
- Duplicate datasets

[Upload Metadata](#)[Ingest External Records](#)[Manage Ingests](#)[Process/Edit Metadata](#)[Delete Records](#)[Admin Page](#)

- Batch Edit Metadata
  - Key fields
  - Access constraints
  - Use constraints
- Search other Solr instances and be able to batch download metadata
- Employ a repository management system such as Fedora
- Add some indexing capabilities to Solr schema

**Preprocess Metadata**

Use this page to preprocess FGDC metadata files. The returned zip file will contain metadata with updated contact info, cleaned up onlink tags, updated use constraints (if provided), and access constraints. Additionally, it will perform a metadata audit to alert you to potential problems in the ingest process.

**Access Constraints File(.txt)**[+ Add file...](#)**Use Constraints File(.txt)**[+ Add file...](#)**Select XML metadata file(s) or zipped directory of XML metadata fil**

Access

**Metadata File(s)**[+ Add files...](#)[ⓘ Process Metadata](#)



Upload Metadata

---

Ingest External Records

---

**Manage Ingests**

---

Process/Edit Metadata

---

Delete Records

---

Admin Page

- Delete records needs to indicate when it fails
- Add more search criteria
- Ability to share out cart

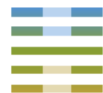
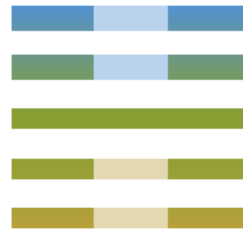


# Harvester Technology

- Java/JavaScript
- Spring MVC/Spring Security
- Angular JS
- PostgreSQL
- Initial support for: solr-to-solr, GeoNetwork, WebDAV/WAF, CSW harvesting

# Harvester Schedule

- GeoCat application development
- January 1, 2014
  - Ingest External Records
  - Manage Ingests
  - Admin Page
- March 2014
  - Upload Metadata
  - Process/Edit Metadata
  - Delete Records



# Open Geoportal **Harvester**

Patrick Florance, Tufts University  
Jeroen Ticheler, GeoCat