



## D5.3.1 – Europeana OAI-PMH Infrastructure – Documentation and final prototype

### Appendix – REPOX User Manual



co-funded by the European Union

The project is co-funded by the European Union, through the **eContentplus** programme

<http://ec.europa.eu/econtentplus>



EuropeanaConnect is coordinated by the Austrian National Library



## Table of Contents

1	Installation .....	4
1.1	System Requirements .....	4
1.1.1	Operating System .....	4
1.1.2	Java 6 .....	4
1.1.3	Application server.....	4
1.2	Download Source Code .....	4
1.3	Database Management System .....	4
1.4	Compilation .....	5
1.5	Deployment.....	5
1.6	Configuration and Testing the Installation.....	5
2	User Manual .....	7
2.1	Login .....	7
2.2	Homepage .....	7
2.3	Aggregator .....	8
2.3.1	Create an Aggregator.....	8
2.3.2	View Aggregator.....	9
2.3.3	Edit an Aggregator .....	9
2.3.4	Remove an Aggregator .....	10
2.4	Data Provider .....	11
2.4.1	Create a Data Provider.....	11
2.4.2	Edit a Data Provider .....	12
2.4.3	Remove a Data Provider .....	12
2.5	Data Set.....	13
2.5.1	Create a Data Set .....	13
2.5.2	Edit a Data Set.....	19
2.5.3	Remove a Data Set.....	20
2.5.4	Ingest a Data Set .....	22
2.5.5	Export a Data Set.....	23
2.6	Ingesting Tasks.....	25
2.6.1	Schedule Tasks.....	25



2.6.2	Consulting Tasks.....	26
2.6.3	Edit and Delete Tasks .....	27
2.7	Schema Transformation Mapper .....	29
2.7.1	Add Schema Transformation.....	29
2.7.2	Visual Schema Mapper .....	30
2.8	Statistics.....	31
2.9	Search for a specific Data Provider or Data Set .....	32
2.10	OAI-PMH Server .....	32



# 1 Installation

## 1.1 System Requirements

### 1.1.1 Operating System

REPOX was successfully tested in the following operating systems: Windows XP, Windows Vista, Windows 7, Linux (Ubuntu) and Mac OS X 10.5.8.

No other operating systems supporting the Java Virtual Machine were tested.

### 1.1.2 Java 6

REPOX is written in Java and assumes the Java 6 (or higher) virtual machine to run properly.

The URL to download the most recent version of Java is: <http://java.sun.com/javase/downloads/index.jsp>. It may be required to have administrative rights in Windows or root permission in Linux to install it.

### 1.1.3 Application server

REPOX runs as a web application within a J2EE servlet container.

REPOX was tested on Apache Tomcat 6.x and Jetty 6.1.11. However it also might run properly on others J2EE servlet containers, since they were not identified other dependencies.

## 1.2 Download Source Code

The REPOX source code is available in the SVN repository at [http://europeanalabs.eu/svn/contrib/repoX/branches/repoX\\_rhine/](http://europeanalabs.eu/svn/contrib/repoX/branches/repoX_rhine/). The source code is located in the folder "REPOX\_HOME".

## 1.3 Database Management System

The current version of REPOX requires the Derby and the PostgreSQL Database Management System. If necessary, in order to install it follow the link: <http://www.postgresql.org/download/>. After this step it is necessary to create a database to contain the REPOX data.

After downloading the source code, it is necessary to configure the PostgreSQL settings in the REPOX configuration file, in these recommended three steps:

1. Create a copy of this file:

"REPOX\_HOME/src/main/resources/repoX2sip-applicationRepoX-context.xml.template-alternate"

2. Rename it to:

"REPOX\_HOME/src/main/resources/repoX2sip-applicationRepoX-context.xml"

3. Edit the file "repoX2sip-applicationRepoX-context.xml".



```
<bean id="dataSource" class="com.mchange.v2.c3p0.ComboPooledDataSource">
  <property name="driverClass" value="org.postgresql.Driver"/>
  <property name="jdbcUrl" value="jdbc:postgresql://host:port/database"/>
  <property name="user" value="ADMIN"/>
  <property name="password" value="ADMIN"/>
</bean>
```

For the property “jdbcUrl” replace “jdbc:postgresql://host:port/database” for the name of the PostgreSQL database that was previously created and change the values of “user” and “password” to reflect your server database configuration.

## 1.4 Compilation

REPOX is compiled using Maven. Instructions to download and learn how to install it can be found here:

“<http://maven.apache.org/download.html>”

Assuming Maven is installed, type this in the command prompt, in the REPOX\_HOME:

```
mvn clean
mvn install -Dmaven.test.skip=true
```

Maven will then compile REPOX and produce a “.war” file located at “REPOX\_Home/target/repox.war”

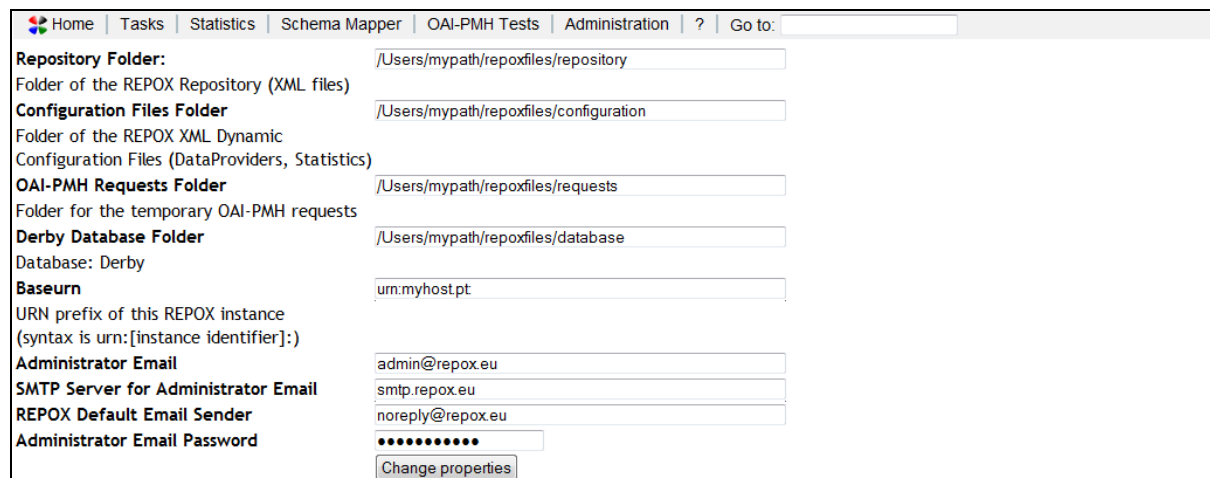
## 1.5 Deployment

After compilation copy the file REPOX\_HOME/target/repox.war to the web application server folder and deploy it using the application server administration tools. Please be sure that the memory of the Java virtual machine in the application server is at least 512 MB (1024 MB recommended) and that PostgreSQL is running.

## 1.6 Configuration and Testing the Installation

To test the installation, connect to <http://server.address:port/repox/>.

The first time the system will redirect the user to the login page. After authentication, REPOX presents the configuration page (Figure 1). Here, the administrator must set up the folder used by REPOX to store data.



The screenshot shows a web interface with a navigation bar at the top containing links for Home, Tasks, Statistics, Schema Mapper, OAI-PMH Tests, Administration, and a search field. Below the navigation bar is a configuration form with the following fields:

- Repository Folder:** /Users/mypath/repoxmlfiles/repository
- Configuration Files Folder:** /Users/mypath/repoxmlfiles/configuration
- OAI-PMH Requests Folder:** /Users/mypath/repoxmlfiles/requests
- Derby Database Folder:** /Users/mypath/repoxmlfiles/database
- Baseurn:** urn.myhost.pt
- Administrator Email:** admin@repor.eu
- SMTP Server for Administrator Email:** smtp.repor.eu
- REPOX Default Email Sender:** noreply@repor.eu
- Administrator Email Password:** [masked]

A "Change properties" button is located at the bottom of the form.

Figure 1 – The configuration page

The form consists on the following fields:

- Repository Folder: where the XML record files will be stored.
- Configuration Files Folder: where the configuration XML will be stored.
- OAI-PMH Requests Folder: folder for the temporary OAI-PMH requests.
- Derby Database Folder: where to store the Derby database. This database will be used to store the information about the harvested metadata records.
- Baseurn: URN prefix of this REPOX instance (the syntax is “urn:[instance identifier:]”).
- Administrator Email: email to notify the administrator when ingestion processes are concluded.
- SMTP server for “Administrator Email”
- REPOX Default Email Sender
- “Administrator Email” Password

After editing the form, click on “Change Settings” to save change settings. Then REPOX will show the main page.

The configuration properties also can be changed at any time using the “Administration” link (Figure 2).



The screenshot shows the REPOX Home Page with a navigation bar at the top. The "Administration" link is circled in red. Below the navigation bar is a table with the following data:

Aggregator ID	Aggregator Name	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records
PT Agg	Portuguese Aggregator	20	Arquivo Municipal de Lisboa	record	ese	OAI-PMH ese			0

Figure 2 - Administration link

## 2 User Manual

### 2.1 Login

When REPOX starts, after the initial configuration, the first screen presented is the login (Figure 3).



Figure 3 – Login Screen

In order to proceed, the user must insert a valid username and password (defined in 1.3).

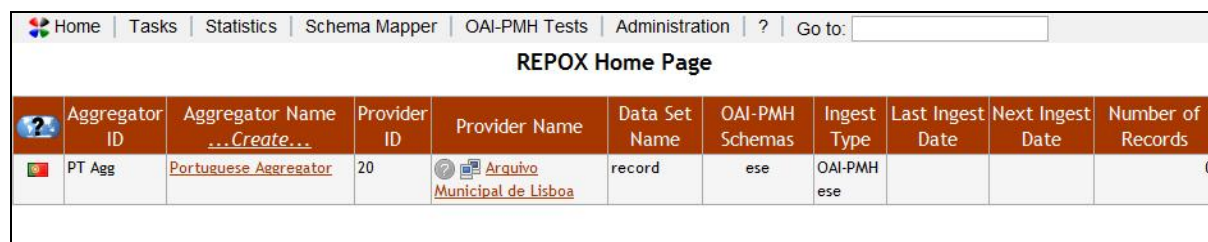
It is also possible to check the “Don't ask for my password for two weeks” checkbox for the system to remember the credentials.

The “Reset” button resets the two fields.

The first time the system is executed, after the login the user will be redirected to the configuration page (see section 1.6 for more details).

### 2.2 Homepage

After the login, REPOX will then present the homepage (Figure 4).





?	Aggregator ID	Aggregator Name ... Create ...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records
	PT Agg	<a href="#">Portuguese Aggregator</a>	20	 <a href="#">Arquivo Municipal de Lisboa</a>	record	ese	OAI-PMH ese			0

Figure 4 – Main list of the REPOX system

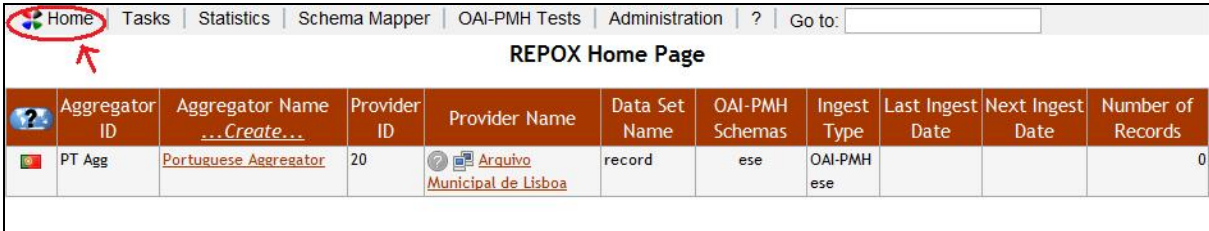
The REPOX homepage presents the registered list of Aggregators, the Data Providers associated to them, and the Data Sources of each Data Provider.

In each column the table shows:

- Aggregator’s country (it is presented the national flag of the country)
- Aggregator Id (the name code of the aggregator)
- Aggregator Name (the name given to the aggregator in the system)
- Provider Id (the name code of the provider)
- Provider Name (the name given to the provider in the system)

- Data Set Name (the name given to the data set in the system)
- OAI-PMH Schemas (the data schemas available for each Data Set)
- Ingest Type (data format type from which the data can be ingested).
- Last Ingest Date (date of the last time the data was ingested)
- Next Ingest Date (date of the next time the data will be ingested)
- Number of Records (number of records for each ingest)

The “Home” link redirects in any moment the user to the homepage.



Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records
PT Agg	Portuguese Aggregator	20	Arquivo Municipal de Lisboa	record	ese	OAI-PMH ese			0

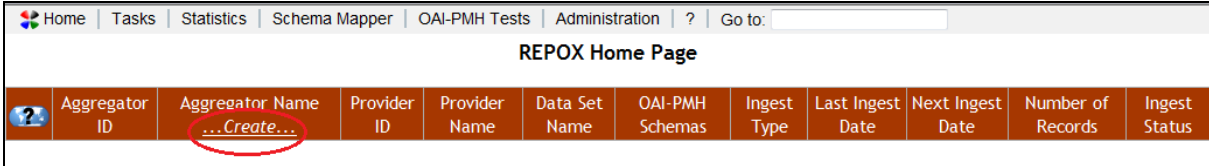
Figure 5 – Link to the homepage

## 2.3 Aggregator

An Aggregator is an entity that aggregates Data Sets from Data Providers, with the purpose of making them available also through the OAI-PMH protocol. The system administrator is allowed to register and manage Aggregators.

### 2.3.1 Create an Aggregator

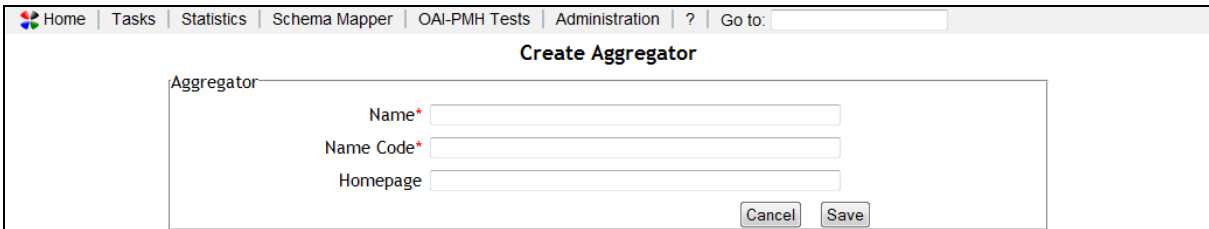
To create an Aggregator the administrator selects the “...Create...” link present in the third column of the home, as showed in the Figure 6.



Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status

Figure 6 – Create Aggregator

In consequence, REPOX presents a related form (Figure 7).



Home | Tasks | Statistics | Schema Mapper | OAI-PMH Tests | Administration | ? | Go to:

### Create Aggregator

Aggregator

Name\*

Name Code\*

Homepage

Figure 7 – Create Aggregator form



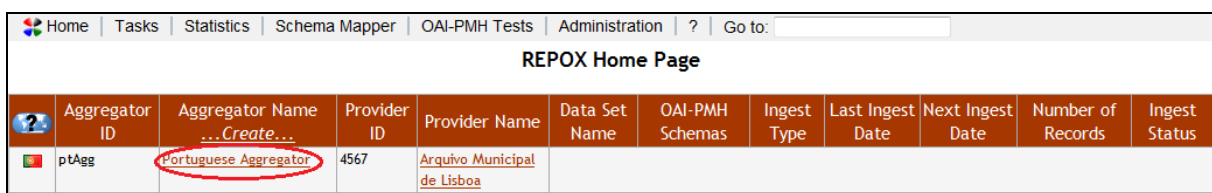
The Aggregator Name field must be filled with the name to be given to the Aggregator, which can be any string or statement.

The Name Code field must be filled with a code assigned from the Ingestion team. This code is used to identify the Aggregator.

The Homepage field should be filled with a URL that represents the Aggregator homepage. The Aggregator Name field and Name Code field are required, the Homepage field is optional.

### 2.3.2 View Aggregator

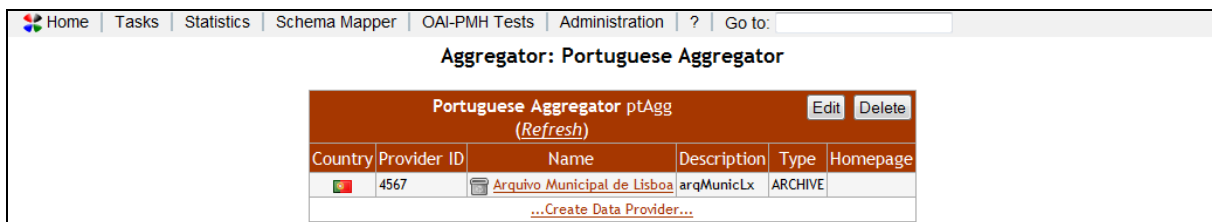
To view information about an Aggregator simply click on its name, as showed in the Figure 8.



Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
ptAgg	Portuguese Aggregator	4567	Arquivo Municipal de Lisboa							

Figure 8 – Selecting the Aggregator form the main list

REPOX will show a screen presenting the information about that Aggregator (Figure 9).



Country	Provider ID	Name	Description	Type	Homepage
	4567	Arquivo Municipal de Lisboa	arqMunicLx	ARCHIVE	

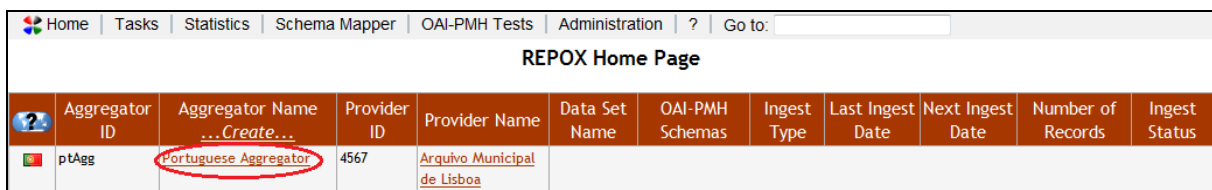
Figure 9 – Screen showing the Aggregator information

This screen contains a table showing the information about all the Data Providers associated with the selected Aggregator (see 2.4 for how to register and manage the detailed information about Data Providers, as also for the meaning of each field in the table):

The garbage icon deletes the current Data Provider.

### 2.3.3 Edit an Aggregator

To edit an Aggregator select it by its name from the homepage (Figure 10).



Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
ptAgg	Portuguese Aggregator	4567	Arquivo Municipal de Lisboa							

Figure 10 – Selecting the Aggregator form the main list

Next, click on the “Edit” button (Figure 11).

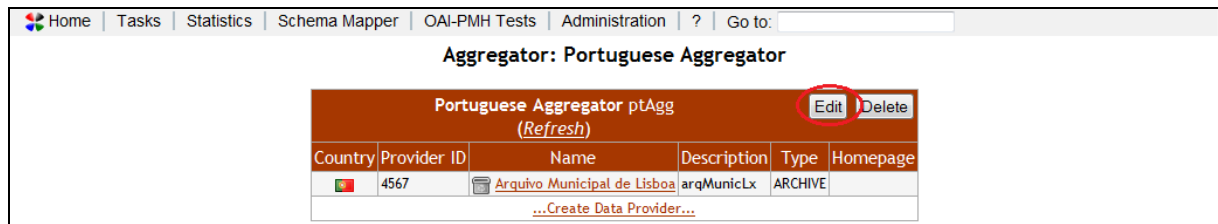


Figure 11 – Aggregator Edit button.

REPOX will show as a result a form to edit the Aggregator, as presented and described in the section 2.3.1).

### 2.3.4 Remove an Aggregator

To remove an Aggregator, select its name from the homepage (Figure 12).

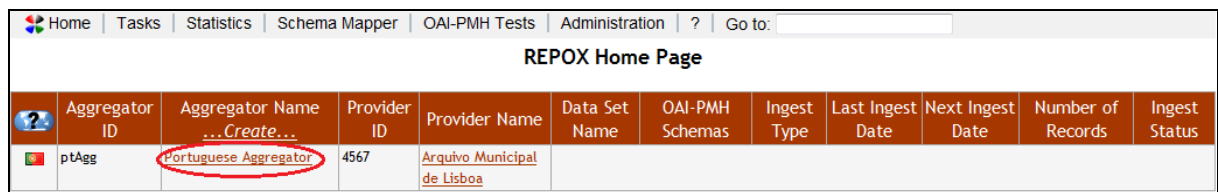


Figure 12 – Selecting the Aggregator form the main list

Next, click on the “Delete” button (Figure 13).

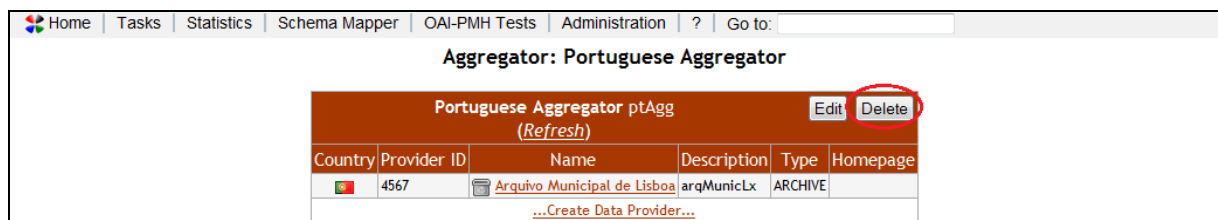


Figure 13 – Aggregator Delete button

A pop-up menu is showed to confirm the deletion of the Aggregator (Figure 14).

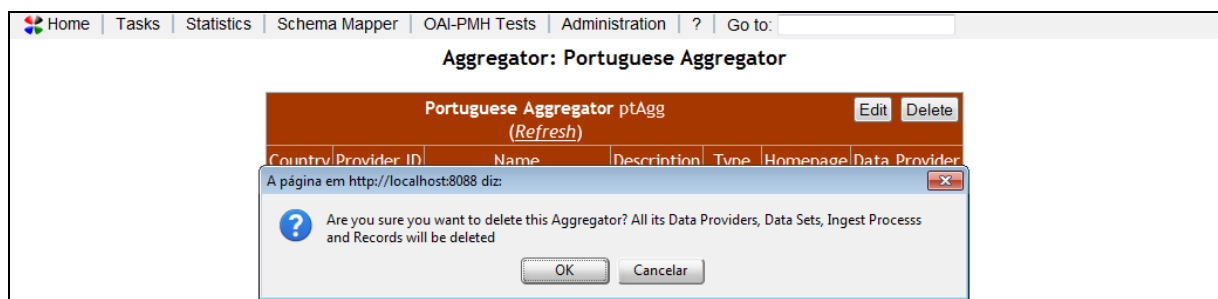


Figure 14 – Pop-up menu to confirm the deleting of the Aggregator

To conclude the operation the administrator should select “Ok”.

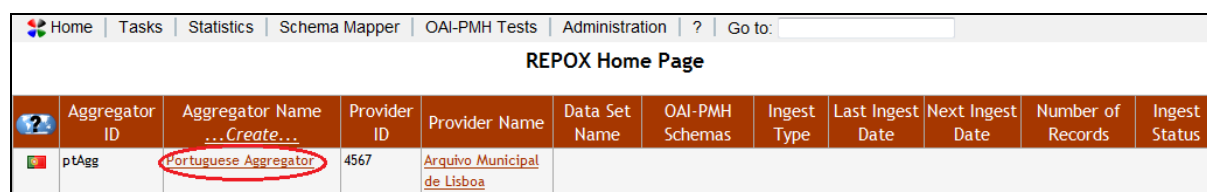
When an Aggregator is deleted all of its Data Providers, Data Sets, ingest processes and Records will be deleted.

## 2.4 Data Provider

A Data Provider is an entity that contributes with descriptive metadata for Europeana.

### 2.4.1 Create a Data Provider

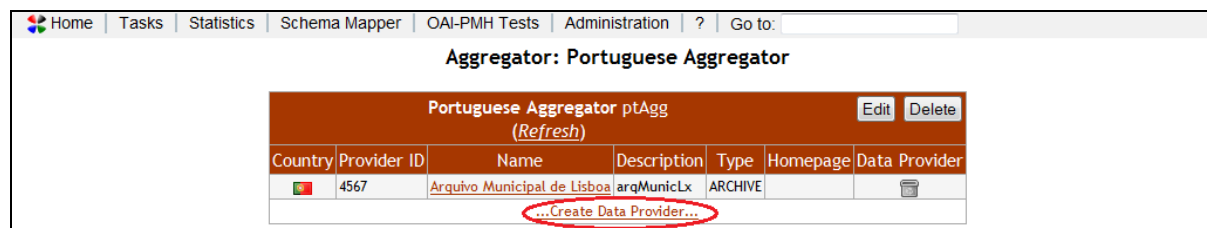
To create a Data Provider, select from the homepage the respective Aggregator.



Aggregator ID	Aggregator Name	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
ptAgg	Portuguese Aggregator	4567	Arquivo Municipal de Lisboa							

Figure 15 – Selecting the Aggregator form the main list

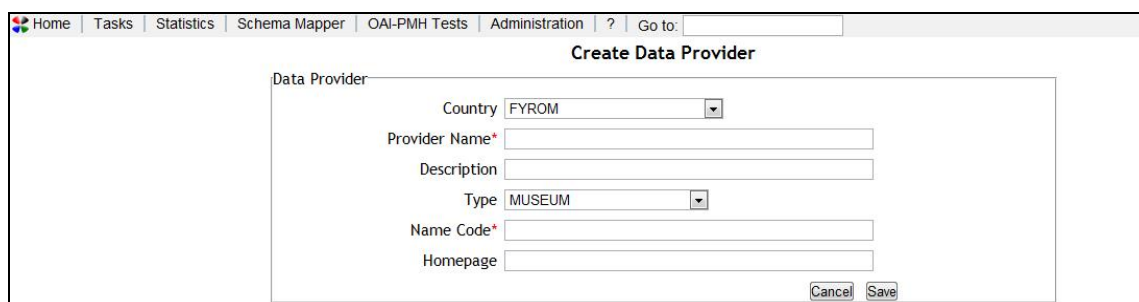
Next, click on “...Create Data Provider...” link (Figure 16).



Country	Provider ID	Name	Description	Type	Homepage	Data Provider
PT	4567	Arquivo Municipal de Lisboa	argMunicLx	ARCHIVE		...Create Data Provider...

Figure 16 – Create Data Provider button

REPOX will present a form with information related to the Data Provider (Figure 17).



Data Provider

Country: FYROM

Provider Name\*:

Description:

Type: MUSEUM

Name Code\*:

Homepage:

Cancel Save

Figure 17 – Create Data Provider form

The form is composed by the following fields:

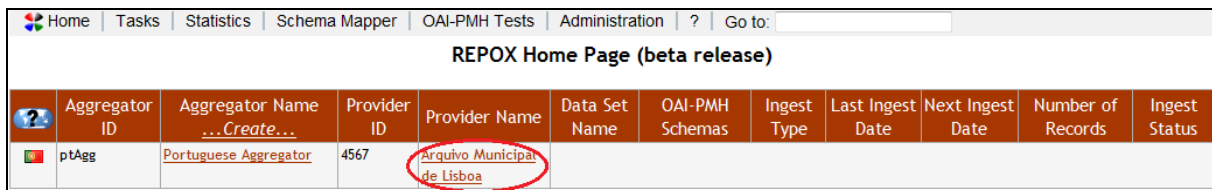
- Country: the country of the Data Provider can be selected from a drop down list.
- Provider Name: the data provider name (a free text).
- Description: a short description of the Data Provider (a free text).

- Type: the type of the Data Provider – the options in the actual version are MUSEUM; ARCHIVE; LIBRARY, AUDIO VISUAL ARCHIVE; RESEARCH EDUCATIONAL; CROSS SECTOR; PUBLISHER; PRIVATE; AGGREGATOR or UNKNKOWN.
- Name Code: the name code value that identifies the Data Provider.
- Homepage: the URL of the homepage of the Data Provider.

The fields Provider Name and Name Code are mandatory; all the others are optional.  
 Edit the fields and click “Save” to save the changes.

### 2.4.2 Edit a Data Provider

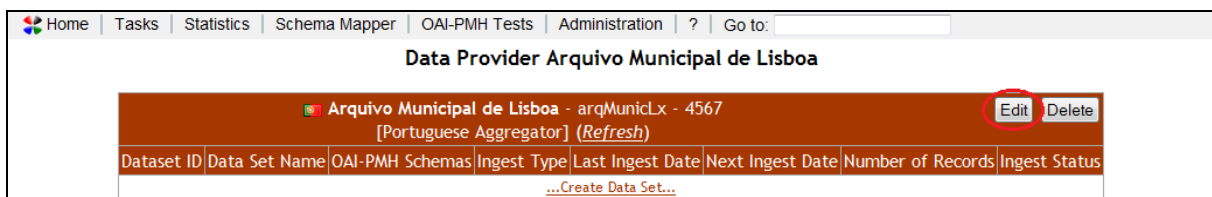
To edit a Data Provider select its name from the homepage (Figure 18).



Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
ptAgg	Portuguese Aggregator	4567	Arquivo Municipal de Lisboa							

Figure 18 – Select Data Provider from the main list

As a consequence, it is presented a new page with all the registered information about that Data Provider. To edit that, click on the “Edit” button (Figure 19).



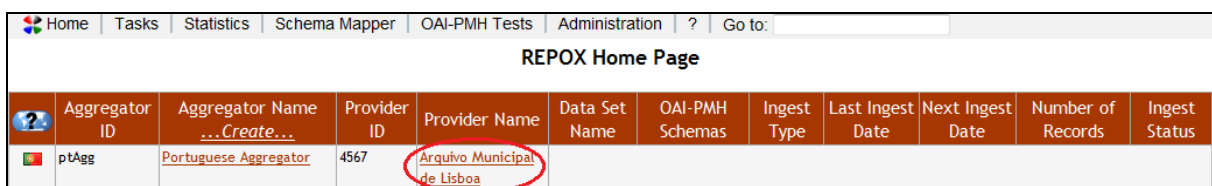
Dataset ID	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
Arquivo Municipal de Lisboa - arqMunicLx - 4567 [Portuguese Aggregator] (Refresh)							
<input type="button" value="Edit"/> <input type="button" value="Delete"/>							
<a href="#">...Create Data Set...</a>							

Figure 19 – Edit Data Provider button

This present the same form as already described in the section 2.4.1.

### 2.4.3 Remove a Data Provider

To remove a Data Provider, select its name from the homepage (Figure 20).



Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
ptAgg	Portuguese Aggregator	4567	Arquivo Municipal de Lisboa							

Figure 20 – Select Data Provider from the main list

Next, click on the “Delete” button in the Data Provider’s details page (Figure 21).

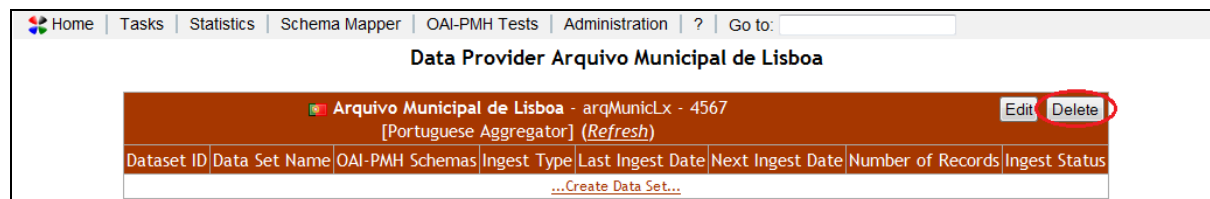


Figure 21 – Delete Data Provider button

A pop-up menu is showed to confirm the deletion of the Aggregator (Figure 22).

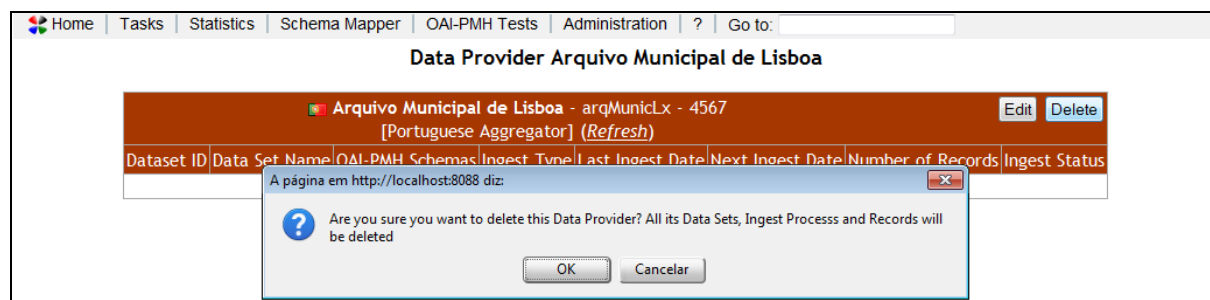


Figure 22 – Pop-up menu to confirm the deleting of the Data Provider

To conclude the operation the administrator should select “Ok”.

When a Data Provider is deleted all of its Data Sets, ingest processes and Records will be deleted.

## 2.5 Data Set

A Data Set is a set of metadata records provided by a Data provider.

A Data Set can be initially stored in an OAI-PMH remote server, in a folder of the file system of the server where the REPOX is running, or available on-line through a URL.

REPOX also makes it possible to harvest a data set from a Z39.50 server, even if that is not its main purpose (since this protocol was not designed for this purpose, REPOX uses an expedite process that never can assure all the records will be harvested).

### 2.5.1 Create a Data Set

To create a Data Set, select, from the homepage (Figure 23) the Data Provider that will provide it (if the Data Provider is not created yet, see 2.4.1 about how to do that).

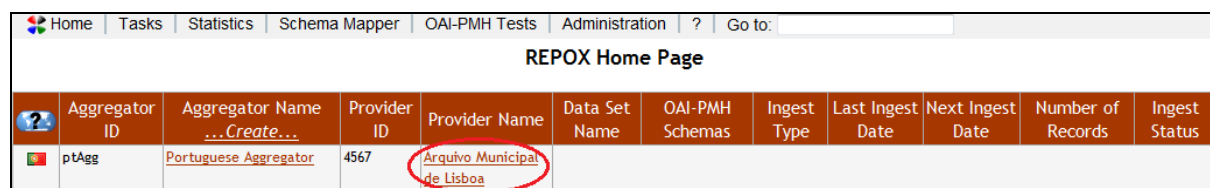


Figure 23 – Select the Data Set

Next, click on “...Create Data Set...” (Figure 24).

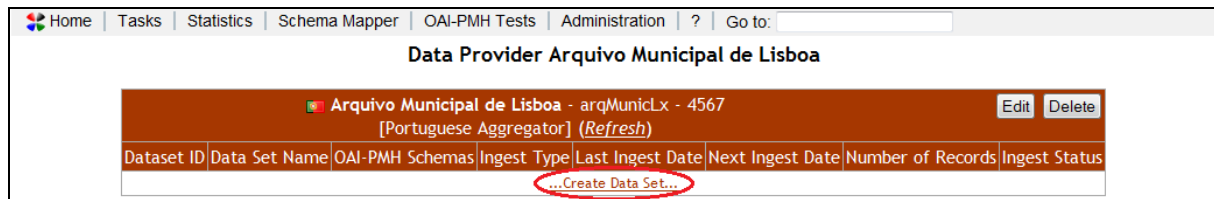


Figure 24 – Create Data Set button

In the next screen, choose one of the options for the source of the Data Set (“Folder”, “OAI-PMH” or “Z39.50”), and fill the respective form.

### 2.5.1.1 Import From a File System Folder or URL

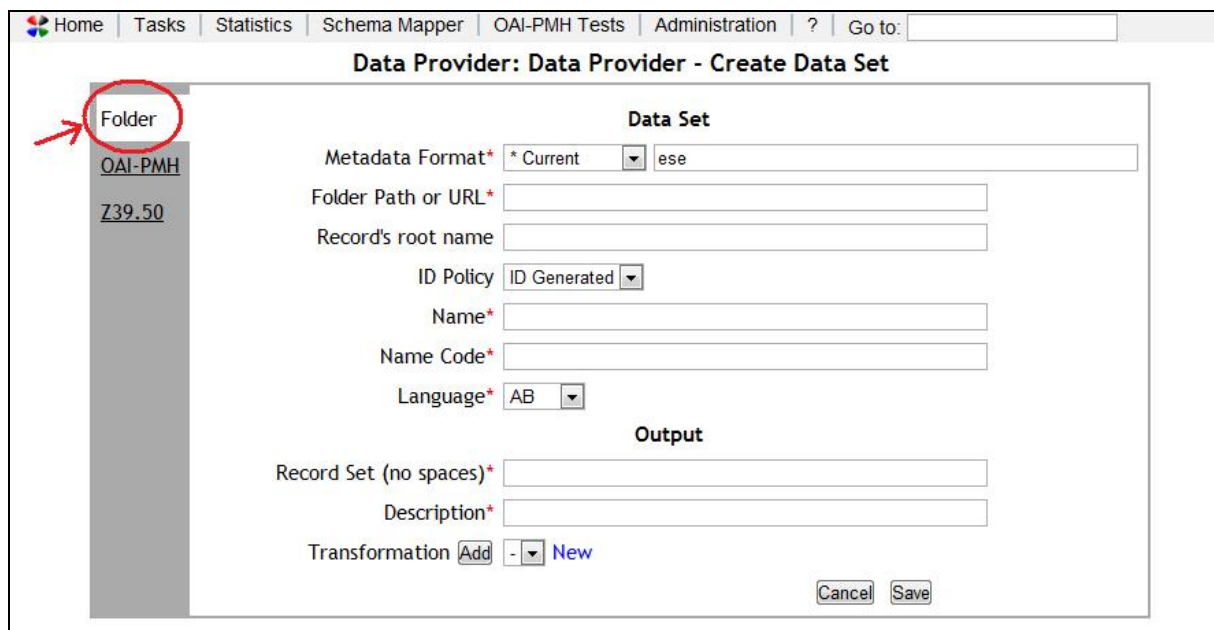


Figure 25 – Form to create a Data Set from a folder or URL

The importing process “Folder” ingests a Data Set from files already existing in a specified folder. The file formats that REPOX recognise for this purpose are ISO2709 and XML. Each Data Set is expected to be in a file with one of these formats, which can also be compressed according to the ZIP format.

For the ISO2709 format is possible to select the variant that is used (standard; variant from Albania; variant from Ukraine) and also the character encoding of the file.

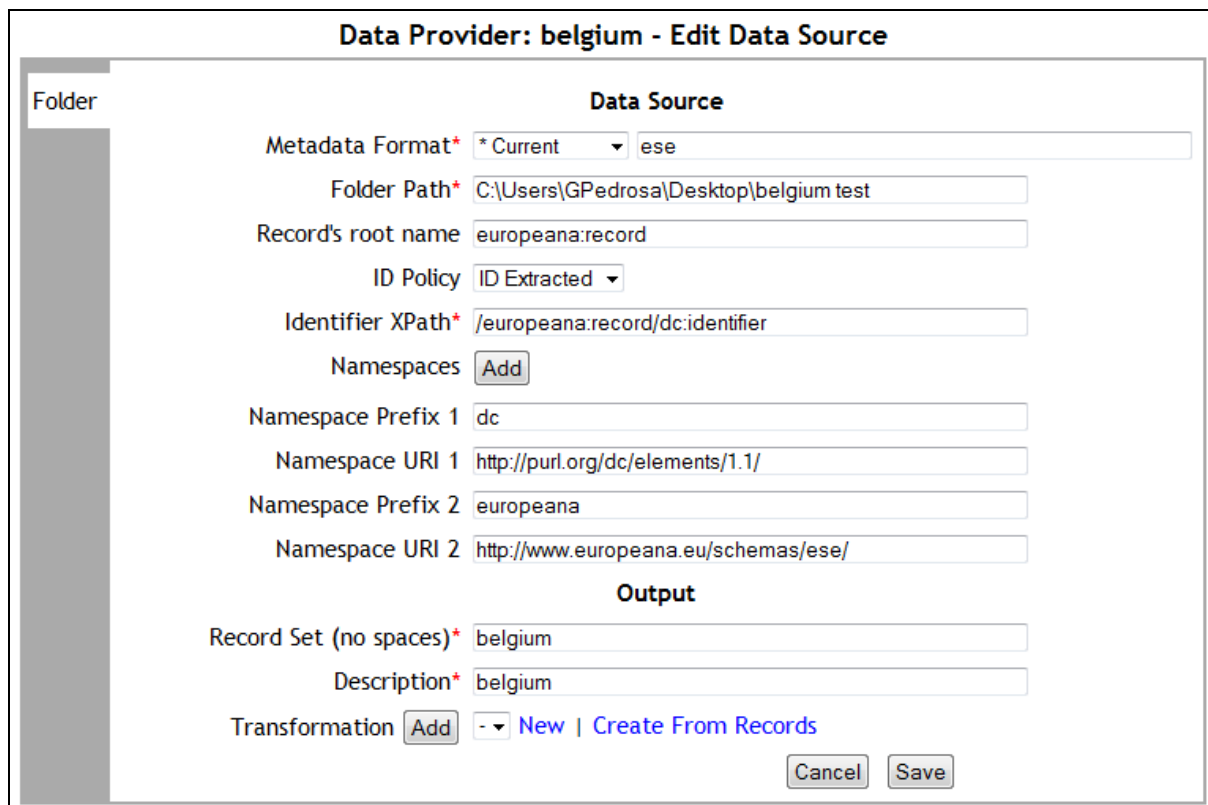
The Figure 25 shows the form to create a new Data Set for this importing process:

- Metadata Format: the format of the input metadata (ese, ISO2709, MarcXchange, tel, oai-dc, other).
- Folder Path or URL: the folder path that contains the files, or a URL to a specific file. If an URL is given, REPOX will download the file to a local folder.
- Record’s root name: the XML element name (the tag name) that holds the record in the file. This field is optional, as it only applies to XML files, and not to ISO2709.



- ID policy: defines how the record identifiers are assigned (to apply only for data sets in XML files). When importing records from a XML file, it can happen that the records already have an element that is a unique identifier or not. For those cases it is possible to indicate the element to use as identifier, or to ask REPOX to generate and assign it. The options are:
  - ID generated – in this case it is assumed that the records to import do not have a unique identifier. A unique identifier is generated by REPOX and assigned to each record.
  - ID extracted – in this case it is assumed that the records to import already have an element with a unique identifier that can be extracted from the record using an XPath expression (ex: /record/id). This is a common scenario, as many bibliographic and other metadata descriptive formats have elements for the identifier. When this option is chosen, new fields appear in the form, as shown on Figure 26. In this case it is necessary to fill the “Identifier XPath” field with the XPath that points to the ID. It is also necessary to add all the namespaces that appear in the XPath, indicating their prefix and respective URI.
- Name: the name that identifies the Data Set.
- Name Code: the name code is to be filled with a code assigned from the Ingestion team. This code is used to identify the Data Set.
- Language: the language of the data values in the Data Set.
- Record Set: the name of the record set. A set is an optional construct for grouping items for the purpose of selective harvesting. Repositories may organize items into sets.
- Description: a description of the Data Set.
- Transformation: A schema transformation that can be used to transform the original data set from its initial format to a new one, to be exposed through the OAI-PMH REPOX server. The transformation rules must be stored in an XSLT file. To add a new transformation, click on the “New” button (see Section 2.7 for more details on adding schema transformations).

The fields “Metadata Format” and “Folder Path” are mandatory, all the others are optional. If the “ID Extracted” is selected for “ID Policy”, then the “Identifier XPath” field is also mandatory, as it is the XML attribute from where the identifier will be extracted.



**Data Provider: belgium - Edit Data Source**

**Folder**

**Data Source**

Metadata Format\*

Folder Path\*

Record's root name

ID Policy

Identifier XPath\*

Namespaces

Namespace Prefix 1

Namespace URI 1

Namespace Prefix 2

Namespace URI 2

**Output**

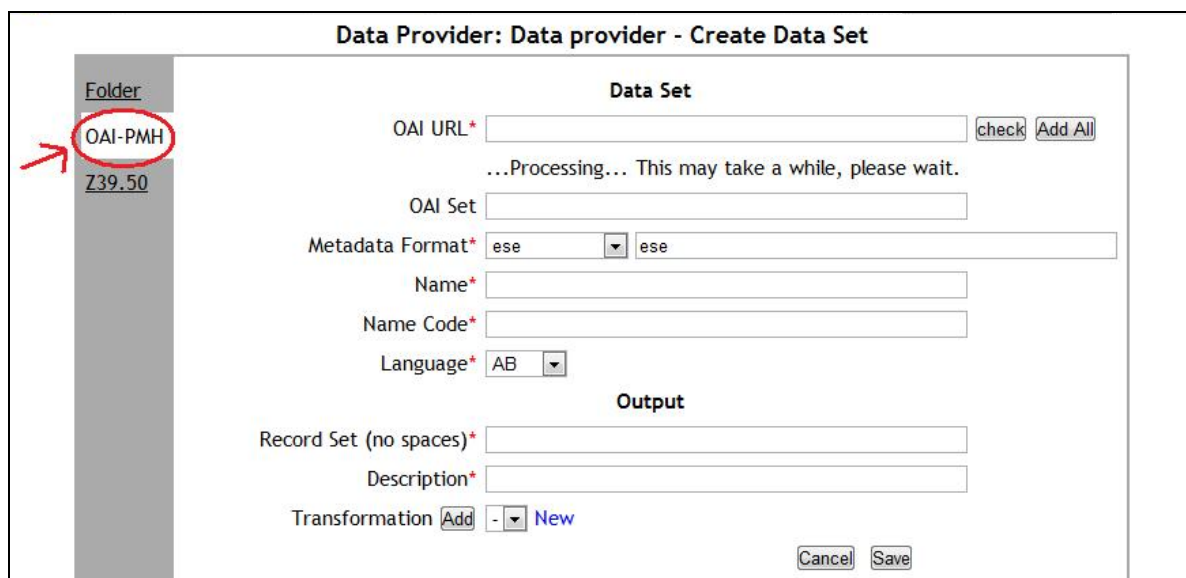
Record Set (no spaces)\*

Description\*

Transformation   [New](#) | [Create From Records](#)

Figure 26 – Creating a Data Source using ID Extracted

### 2.5.1.2 Harvest by OAI-PMH



**Data Provider: Data provider - Create Data Set**

**Folder**

**Data Set**

OAI URL\*

...Processing... This may take a while, please wait.

OAI Set

Metadata Format\*

Name\*

Name Code\*

Language\*

**Output**

Record Set (no spaces)\*

Description\*

Transformation   [New](#)

Figure 27 – Form to create a Data Set from OAI-PMH server

The importing process “OAI-PMH” imports a Data Set from a OAI-PMH server. The Figure 27 shows the form to create a new Data Set for this process, where the fields are:





- OAI URL: the URL of the OAI server from which the data will be obtained. The check button verifies the validity of the URL
- OAI Set: name of the Data Set on the OAI server. A set is an optional construct for grouping items for the purpose of selective harvesting. Repositories may organize items into sets. If this field is not provided, REPOX will ingest all records that exist in the OAI-PMH server.
- Metadata Format: the format of the Data Set to be harvested.
- Name: the name that identifies the Data Set.
- Name Code: the name code that identifies the Data Set.
- Language: the language of the data values in the Data Set.
- Record Set: the new name of the Record Set in the REPOX system.
- Description: a description of the Data Set.
- Transformation: A schema transformation that can be used to transform the original data set from its initial format to a new one, to be exposed through the OAI-PMH REPOX server. The transformation rules must be stored in an XSLT file. To add a new transformation, click on the “New” button (see Section 2.7 for more details on adding schema transformations).

“OAI Set” and “Transformation” fields are optional, the others are mandatory.

### 2.5.1.3 Harvest by Z39.50

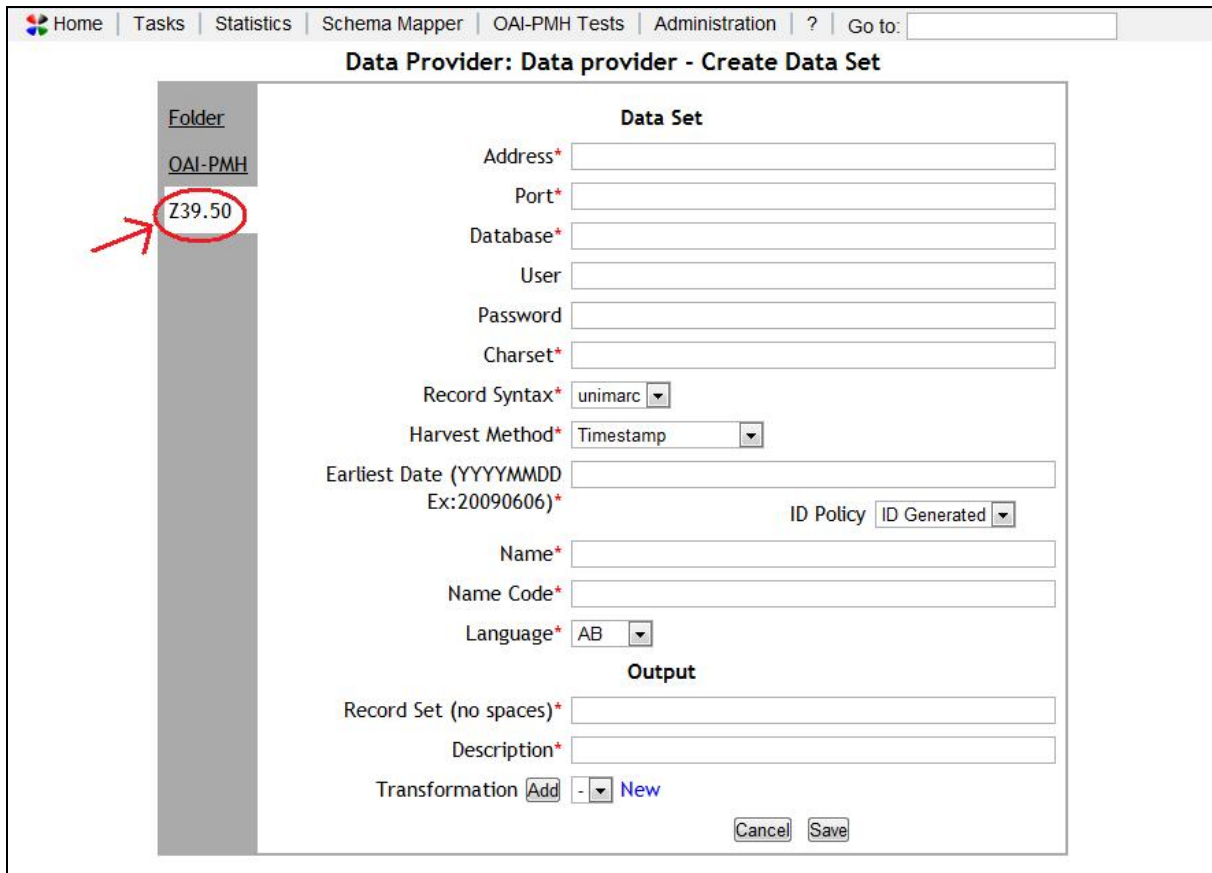


Figure 28 – Form to create a Data Set from Z39.50 server

The Figure 28 shows the importing process for Z39.50. The form to create a new Data Set for this process comprises the following fields:

- Address: address of the Z39.50 server
- Port: service port
- Database: name of the database that contains the records in the remote system
- User: username to access the Z39.50 service
- Password: password for the username
- Charset: The character encoding set of the data (for example UTF-8 or ISO 8859-1)
- Record Syntax: The original syntax of the records. It can be UNIMARC or MARC21.
- Harvest Method: the harvest method, which can be Timestamp, File (one Id per file) or Id Sequence.
- Earliest Date: REPOX will harvest the records from the Z39.50 server, until it reaches the records with the date specified on this field.
- ID policy: defines how the record identifiers are assigned. The options are:



- ID generated – in this case it is assumed that the records to import do not have a unique identifier. A unique identifier is generated by REPOX and assigned to each record.
- ID extracted - the imported records have a unique identifier, to be extracted from the Record using an XPath expression (ex: /record/id). This is a common scenario, as bibliographic metadata formats have tags for the identifier.
- Name: name of the Data Set.
- Name Code: the name code that identifies the data set.
- Language: the language of the data values in the Data Set.
- Record Set: the new name of the record set on the REPOX system.
- Description: a description of the Data Set.
- Transformation: A schema transformation that can be used to transform the original data set from its initial format to a new one, to be exposed through the OAI-PMH REPOX server. The transformation rules must be stored in an XSLT file. To add a new transformation, click on the “New” button (see Section 2.7 for more details on adding schema transformations).

The “User” and “Password” fields are optional, all the others are mandatory.

### 2.5.2 Edit a Data Set

To edit a Data Provider select it by its name from the homepage (Figure 29).

Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
ptAgg	Portuguese Aggregator	4567	Arquivo Municipal de Lisboa	arquivocmbx	ese	OAI-PMH ese			0	

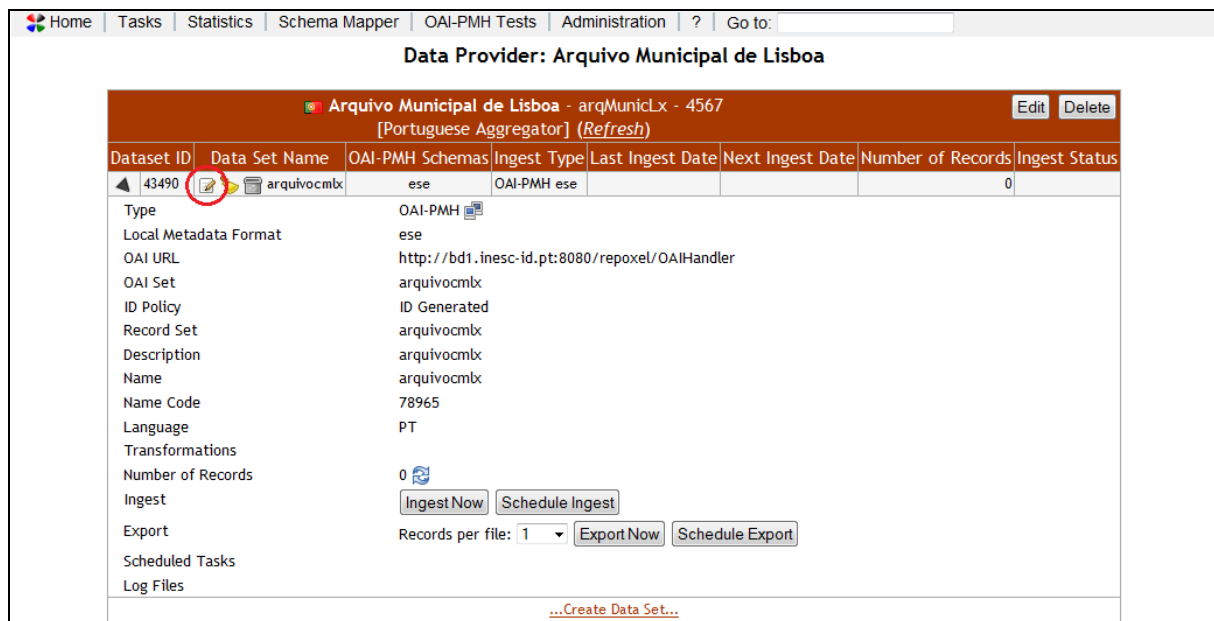
Figure 29 – Select the Data Provider from the main list

Then click on the arrow button to see the Data Set details (Figure 30).

Dataset ID	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
▶ 3490	arquivocmbx	ese	OAI-PMH ese			0	

Figure 30 – Arrow button to show the Data Set details

Finally, click on the pencil icon as showed in Figure 31.



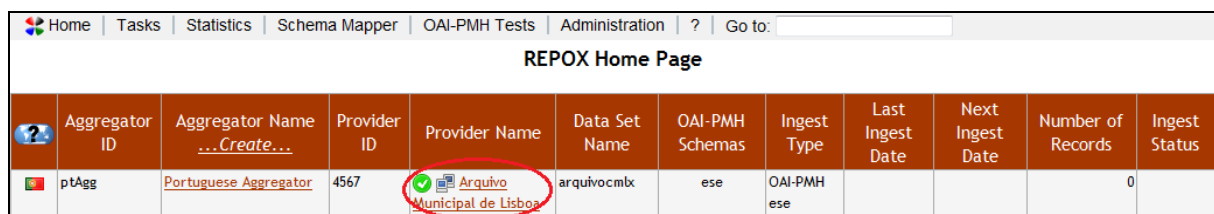
The screenshot shows the 'Data Provider: Arquivo Municipal de Lisboa' page. At the top, there is a header for 'Arquivo Municipal de Lisboa - arqMunicLx - 4567' with 'Edit' and 'Delete' buttons. Below this is a table with columns: Dataset ID, Data Set Name, OAI-PMH Schemas, Ingest Type, Last Ingest Date, Next Ingest Date, Number of Records, and Ingest Status. The first row shows Dataset ID 43490, Data Set Name 'arquivocmbx', OAI-PMH Schemas 'ese', Ingest Type 'OAI-PMH ese', and Number of Records '0'. A pencil icon is circled in red in the 'Data Set Name' column. Below the table, there are various metadata fields like Type, Local Metadata Format, OAI URL, OAI Set, ID Policy, Record Set, Description, Name, Name Code, Language, Transformations, Number of Records, Ingest, Export, and Scheduled Tasks. At the bottom, there is a link to '...Create Data Set...'.

Figure 31 – Edit Data Set button

This action calls the form already described in the section 2.5.2.

### 2.5.3 Remove a Data Set

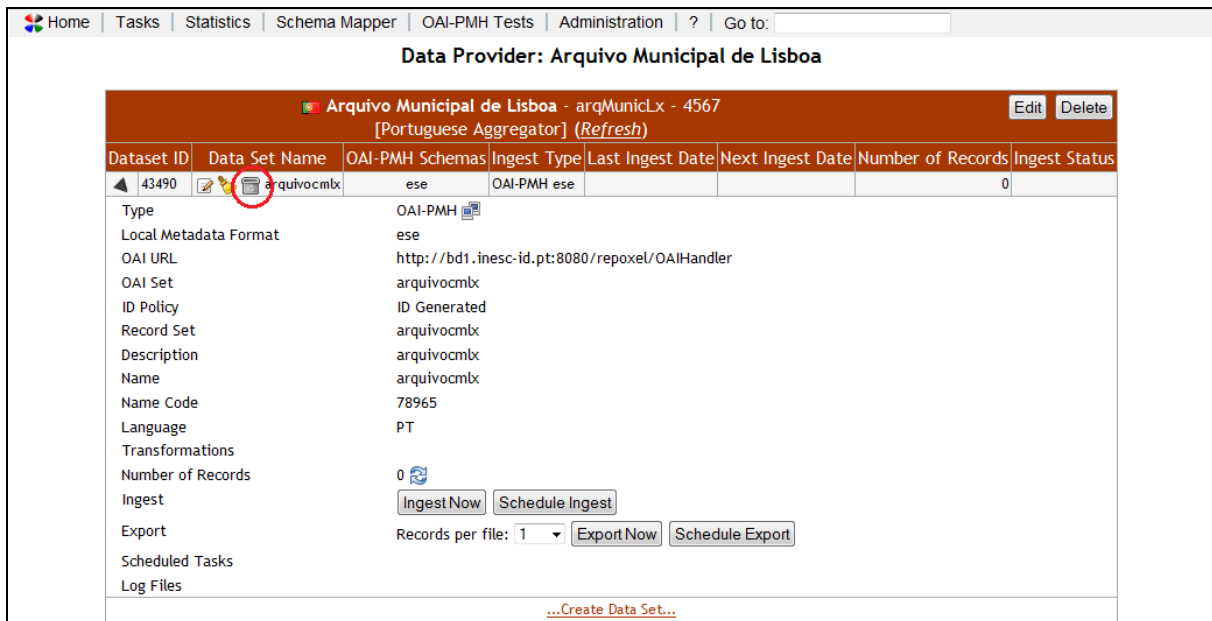
To remove a Data Set select in the homepage the Data Provider to which it is associated (Figure 32).



The screenshot shows the 'REPOX Home Page' with a table of aggregators. The table has columns: Aggregator ID, Aggregator Name, Provider ID, Provider Name, Data Set Name, OAI-PMH Schemas, Ingest Type, Last Ingest Date, Next Ingest Date, Number of Records, and Ingest Status. The first row shows Aggregator ID 'ptAgg', Aggregator Name 'Portuguese Aggregator', Provider ID '4567', Provider Name 'Arquivo Municipal de Lisboa' (circled in red), Data Set Name 'arquivocmbx', OAI-PMH Schemas 'ese', Ingest Type 'OAI-PMH ese', and Number of Records '0'.

Figure 32 – Select the Data Provider from the homepage

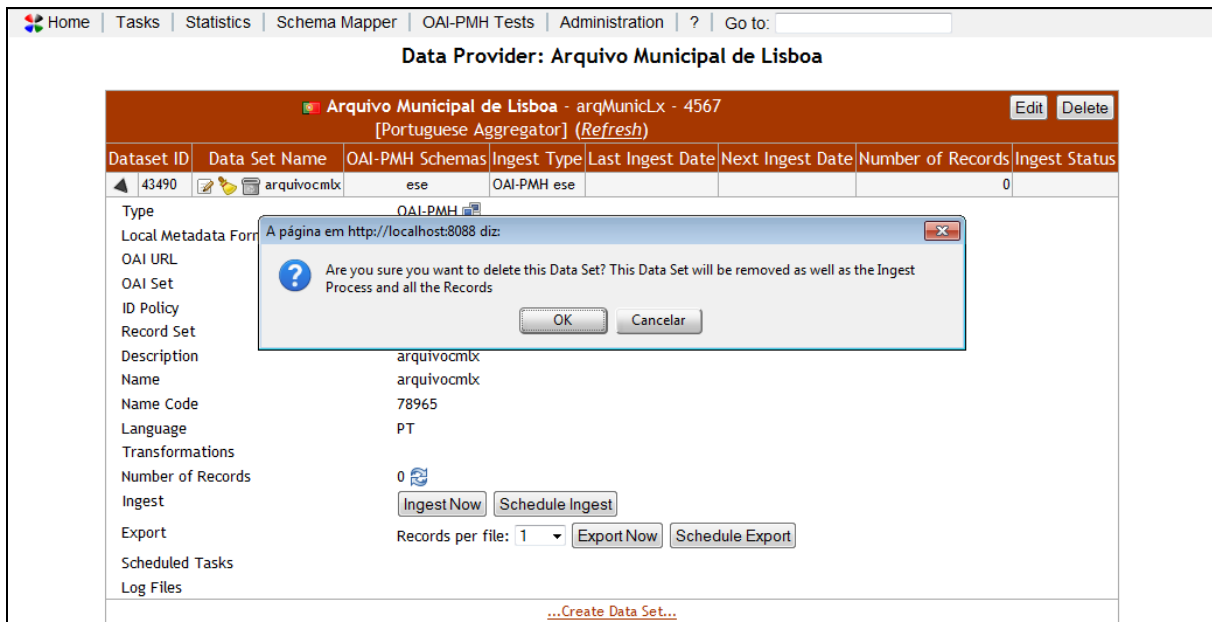
Then, click on the recycle bin icon showed on Figure 33.



The screenshot shows the REPOX interface for a data provider named 'Arquivo Municipal de Lisboa'. The main content area displays details for a specific data set with ID 43490. A red circle highlights the 'Delete' button (represented by a trash can icon) in the 'Data Set Name' column of the table. Below the table, various metadata fields are listed, including 'Type', 'Local Metadata Format', 'OAI URL', 'OAI Set', 'ID Policy', 'Record Set', 'Description', 'Name', 'Name Code', 'Language', 'Transformations', 'Number of Records', 'Ingest', 'Export', 'Scheduled Tasks', and 'Log Files'. The 'Ingest' section includes buttons for 'Ingest Now' and 'Schedule Ingest'. The 'Export' section includes a dropdown for 'Records per file' and buttons for 'Export Now' and 'Schedule Export'. At the bottom, there is a link to '...Create Data Set...'. The top navigation bar includes links for Home, Tasks, Statistics, Schema Mapper, OAI-PMH Tests, Administration, and a search field.

Figure 33 – Delete Data Set button

A pop-up menu is showed to confirm the deletion of the Aggregator (Figure 34).



This screenshot is similar to Figure 33, but it shows a confirmation dialog box overlaid on the interface. The dialog box has a question mark icon and contains the text: 'A página em http://localhost:8088 diz: Are you sure you want to delete this Data Set? This Data Set will be removed as well as the Ingest Process and all the Records'. Below the text are two buttons: 'OK' and 'Cancelar'. The background interface is dimmed, showing the same data set details as in Figure 33.

Figure 34 – Pop-up menu to confirm the deleting of the Data Set

To conclude the operation the administrator should select “Ok”.

When a Data Set is deleted all related ingest processes and Records will be deleted.

### 2.5.4 Ingest a Data Set

To start an ingesting process, select in the homepage the Data Provider to which the Data Set is associated.

REPOX Home Page											
Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status	
ptAgg	Portuguese Agregator	4567	Arquivo Municipal de Lisboa	arquivocmbx	ese	OAI-PMH ese			0		

Figure 35 – Select the Data Provider in the homepage.

Click on the “Ingest Now” button to start ingesting the data immediately (Figure 36).

Data Provider: Arquivo Municipal de Lisboa							
Arquivo Municipal de Lisboa - arqMunicLx - 4567 [Portuguese Agregator] (Refresh)							
Dataset ID	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
43490	arquivocmbx	ese	OAI-PMH ese			0	
Type		OAI-PMH					
Local Metadata Format		ese					
OAI URL		http://bd1.inesc-id.pt:8080/repoxl/OAIHandler					
OAI Set		arquivocmbx					
ID Policy		ID Generated					
Record Set		arquivocmbx					
Description		arquivocmbx					
Name		arquivocmbx					
Name Code		78965					
Language		PT					
Transformations							
Number of Records		0					
Ingest		<input type="button" value="Ingest Now"/> <input type="button" value="Schedule Ingest"/>					
Export		Records per file: 1 <input type="button" value="Export Now"/> <input type="button" value="Schedule Export"/>					
Scheduled Tasks							
Log Files							

Figure 36 – “Ingest Now” button




REPOX will show a message to confirm that the harvesting will start a few seconds (Figure 37).

Data Provider Arquivo Municipal de Lisboa							
Harvest will start in a few seconds.							
Arquivo Municipal de Lisboa - arquimLX <input type="button" value="Edit"/> <input type="button" value="Delet"/>							

Figure 37 – Starting harvest message

During the ingestion process, the ingest status will display an icon showing activity (Figure 38). Clicking on the refresh button will update the number of harvested records.

At the end of the ingestion process, the icon will change according to the result of process. There are three different possibilities:

- successful  (the ingest process has successfully completed);
- error  (an error occurred during the ingest process);
- warning  (the ingest process has successfully completed, but the system could not send an email to the administrator).

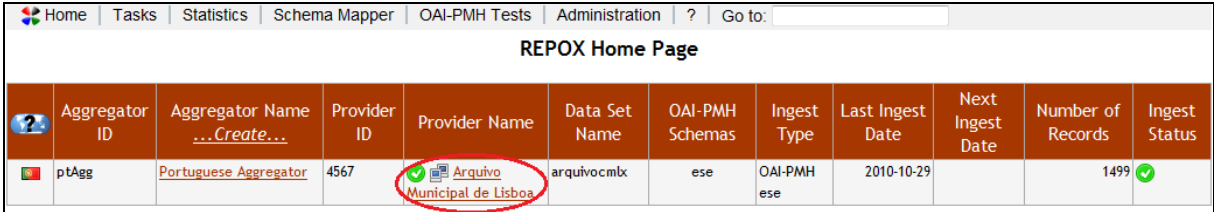


The screenshot shows the REPOX interface for a specific data provider. The main heading is "Data Provider: Arquivo Municipal de Lisboa". Below this, there is a table with the following columns: Dataset ID, Data Set Name, OAI-PMH Schemas, Ingest Type, Last Ingest Date, Next Ingest Date, Number of Records, and Ingest Status. The first row of data shows Dataset ID 43490, Data Set Name arquivocmbx, OAI-PMH Schemas ese, Ingest Type OAI-PMH ese, and Number of Records 500. The Ingest Status column for this row contains a refresh icon. Below the table, there are several sections: "Type" (OAI-PMH), "Local Metadata Format" (ese), "OAI URL" (http://bd1.inesc-id.pt:8080/repoxl/OAIHandler), "OAI Set" (arquivocmbx), "ID Policy" (ID Generated), "Record Set" (arquivocmbx), "Description" (arquivocmbx), "Name" (arquivocmbx), "Name Code" (78965), "Language" (PT), "Transformations", "Number of Records" (500 with a refresh icon), "Ingest" (Ingest Now, Schedule Ingest), "Export" (Records per file: 1, Export Now, Schedule Export), "Scheduled Tasks", and "Log Files" (2010-10\1\_20101029\_174119.log). There are also "Edit" and "Delete" buttons at the top right of the table header.

Figure 38 – Refresh number of records button

## 2.5.5 Export a Data Set

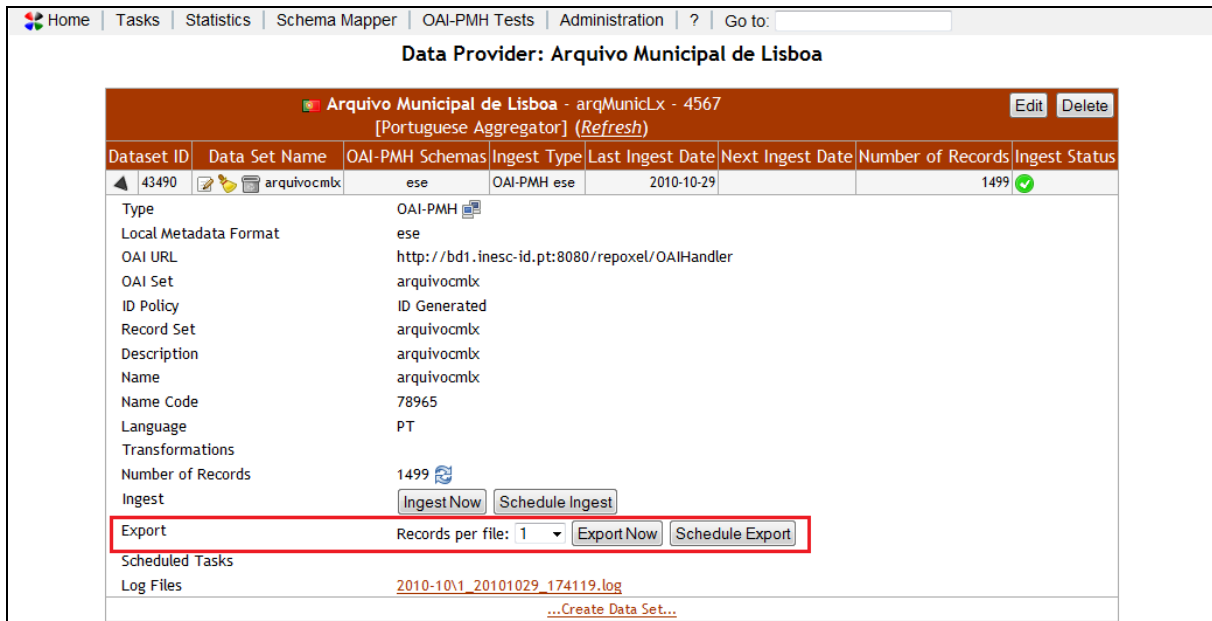
REPOX allows the user to export all the records from a Data Set to a compressed ZIP file. To execute this functionality, select the Data Set from the main list.



The screenshot shows the REPOX Home Page with a table of data providers. The table has the following columns: Aggregator ID, Aggregator Name, Provider ID, Provider Name, Data Set Name, OAI-PMH Schemas, Ingest Type, Last Ingest Date, Next Ingest Date, Number of Records, and Ingest Status. The first row of data shows Aggregator ID ptAgg, Aggregator Name Portuguese Aggregator, Provider ID 4567, Provider Name Arquivo Municipal de Lisboa, Data Set Name arquivocmbx, OAI-PMH Schemas ese, Ingest Type OAI-PMH ese, Last Ingest Date 2010-10-29, Next Ingest Date, Number of Records 1499, and Ingest Status (green checkmark). The 'Arquivo Municipal de Lisboa' provider name is circled in red.

Figure 39 – Select Data Provider from the main list

The Export Data configuration options are showed in the Figure 40.



The screenshot shows the 'Data Provider: Arquivo Municipal de Lisboa' interface. A table lists dataset information, and below it, a detailed view for dataset ID 43490. The 'Export' section is highlighted with a red box, showing 'Records per file: 1' and buttons for 'Export Now' and 'Schedule Export'.

Dataset ID	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
43490	arquivocmbx	ese	OAI-PMH ese	2010-10-29		1499	✓

Export options: Records per file: 1, Export Now, Schedule Export.

Figure 40 – Export data options

Set the number of records that each file will contain using “Records per File” drop down list.

It is possible to export the records immediately, clicking on the “Export Now” button. In this case the result will be stored in a compressed ZIP file and the pathname presented to the user.

It is also possible to schedule the export process clicking on the “Schedule Export” button. In this case REPOX will show a calendar window with the related options (Figure 41).



The screenshot shows the 'Schedule' window for the export process. It includes a calendar for November and December 2010, and configuration options for 'First Run', 'Records per file', and 'Frequency'.

First Run: 02-11-2010 at 16 h 0 min  
 Records per file: 1  
 Frequency:  Once,  Daily,  Weekly,  Every 1 Months

Figure 41 – Window to schedule the export data process





Select from the calendar the day to start the exporting process. Set the hour and the minutes from the two drop down lists. Set the number of records per file and the full path of the folder that will contain the records. Set the frequency of the process, which can be once, daily, weekly or monthly. If monthly is chosen the number interval months can be set. Finally, click on the “Schedule” button to save the task.

The result of an export will be stored in a compressed ZIP file and the pathname presented to the user.

## 2.6 Ingesting Tasks

The Tasks allows scheduling the ingesting tasks for each Data Set.

### 2.6.1 Schedule Tasks

To schedule an ingesting task for a Data Set select, from the homepage, the Data Set to which the Data Set is associated.

REPOX Home Page											
Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status	
ptAgg	Portuguese Agregator	4567	Arquivo Municipal de Lisboa	arquivocmbx	ese	OAI-PMH ese	2010-10-29		1499	✓	

Figure 42 – Select Data Provider from the homepage

Then, click on the “Schedule Ingest” button as showed in Figure 43.

Home | Tasks | Statistics | Schema Mapper | OAI-PMH Tests | Administration | ? | Go to: \_\_\_\_\_

**Data Provider: Arquivo Municipal de Lisboa**

Arquivo Municipal de Lisboa - arqMunicLx - 4567							
[Portuguese Agregator] (Refresh)							
Dataset ID	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
43490	arquivocmbx	ese	OAI-PMH ese	2010-10-29		1499	✓
Type	OAI-PMH						
Local Metadata Format	ese						
OAI URL	http://bd1.inesc-id.pt:8080/repoxml/OAIHandler						
OAI Set	arquivocmbx						
ID Policy	ID Generated						
Record Set	arquivocmbx						
Description	arquivocmbx						
Name	arquivocmbx						
Name Code	78965						
Language	PT						
Transformations							
Number of Records	1499						
Ingest	<input type="button" value="Ingest Now"/> <input type="button" value="Schedule Ingest"/>						
Export	Records per file: 1 <input type="button" value="Export Now"/> <input type="button" value="Schedule Export"/>						
Scheduled Tasks							
Log Files	<a href="#">2010-10\1_20101029_174119.log</a> ...Create Data Set...						

Figure 43 – Schedule Ingest button

REPOX will show a window with a calendar (Figure 44).

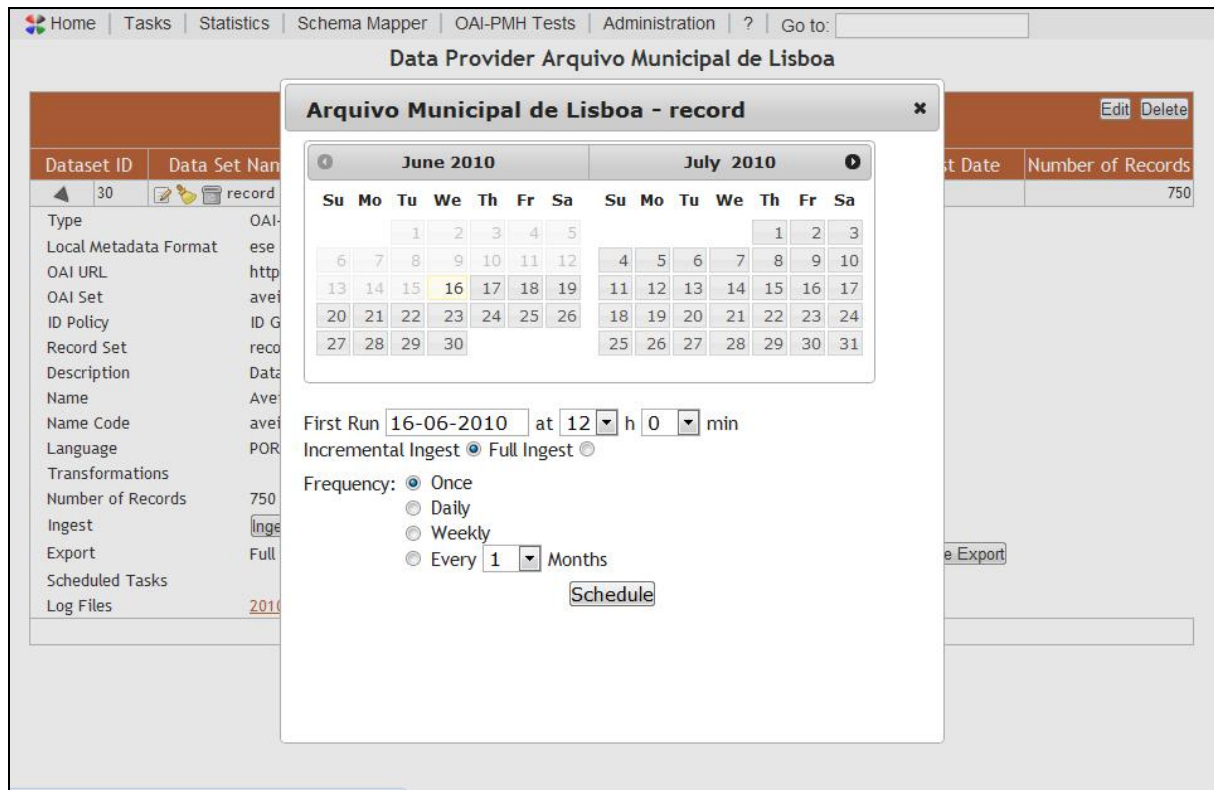


Figure 44 – Window to program the task

Next, select from the calendar the day, hour and the minutes from the two drop down lists.

Choose also Incremental or Full Ingest. If Full Ingest is chosen, REPOX will get all the records that exist in the source. If Incremental Ingest is chosen, REPOX will get only the new or modified records since the last ingest.

Set also the frequency of the process, which can be once, daily, weekly or monthly. If monthly is chosen a number of months can be set to apply between two runs.

Finally, click on the “Schedule” button to save the schedule task.

### 2.6.2 Consulting Tasks

The list of scheduled tasks could be consulted by clicking on the “Tasks” button on the top menu in the homepage (Figure 45).

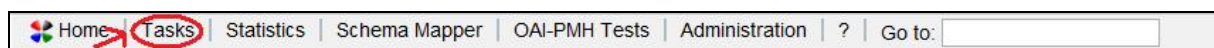


Figure 45 – Tasks button

REPOX will then present a calendar showing all the programmed tasks. Navigate through the months clicking on the arrows on the top of the calendar, bounding the month and year.

The Figure 46 presents the calendar showing all the programmed tasks.

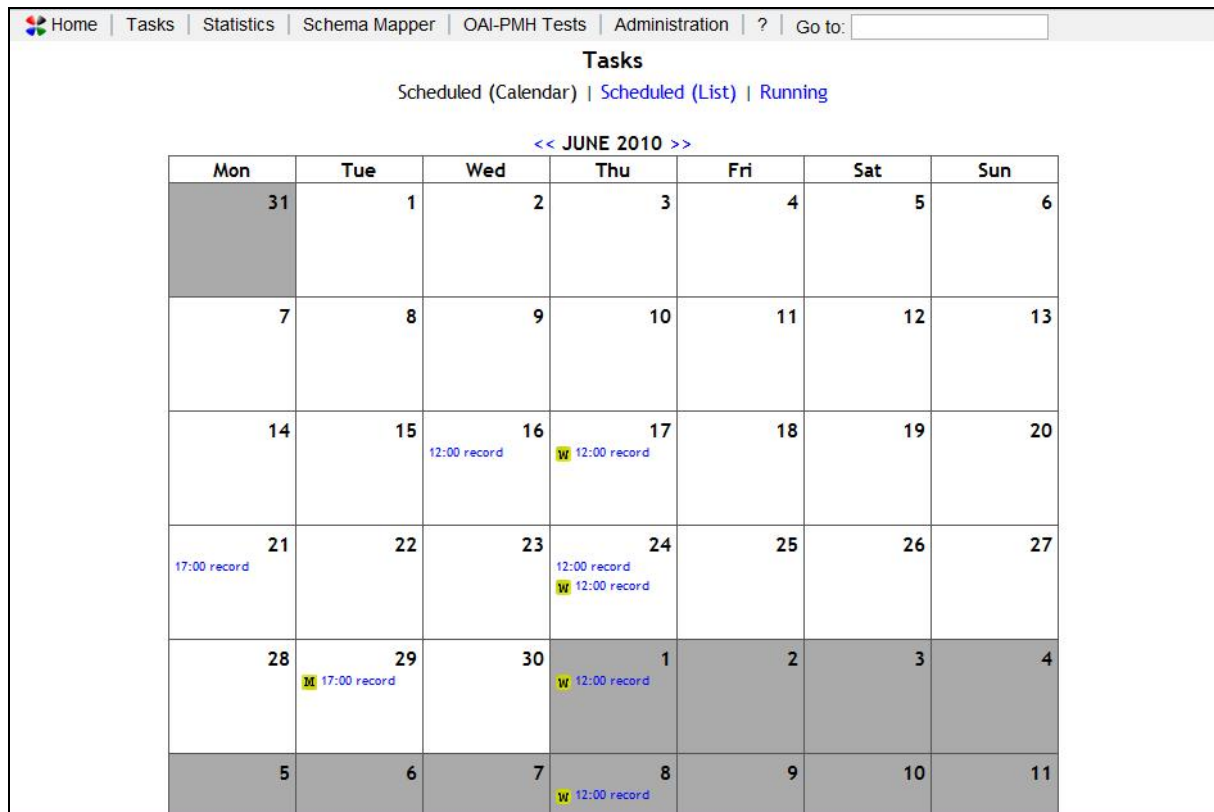


Figure 46 – Calendar showing the scheduled tasks

Clicking on the “Scheduled (List)” button shows a list of the programmed tasks.

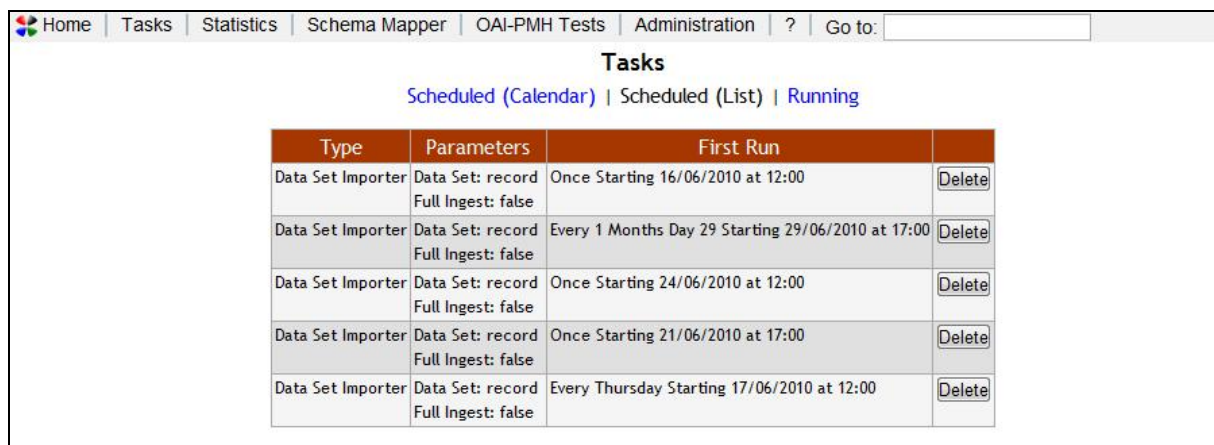


Figure 47 – Scheduled tasks list

The “Running” button shows a list of the tasks that are running in that moment.

### 2.6.3 Edit and Delete Tasks

To edit or delete a task, click on the “Tasks” button on the top menu (Figure 48).



Figure 48 – Tasks button

Clicking on a task, a window with the task scheduling information is presented (Figure 49).

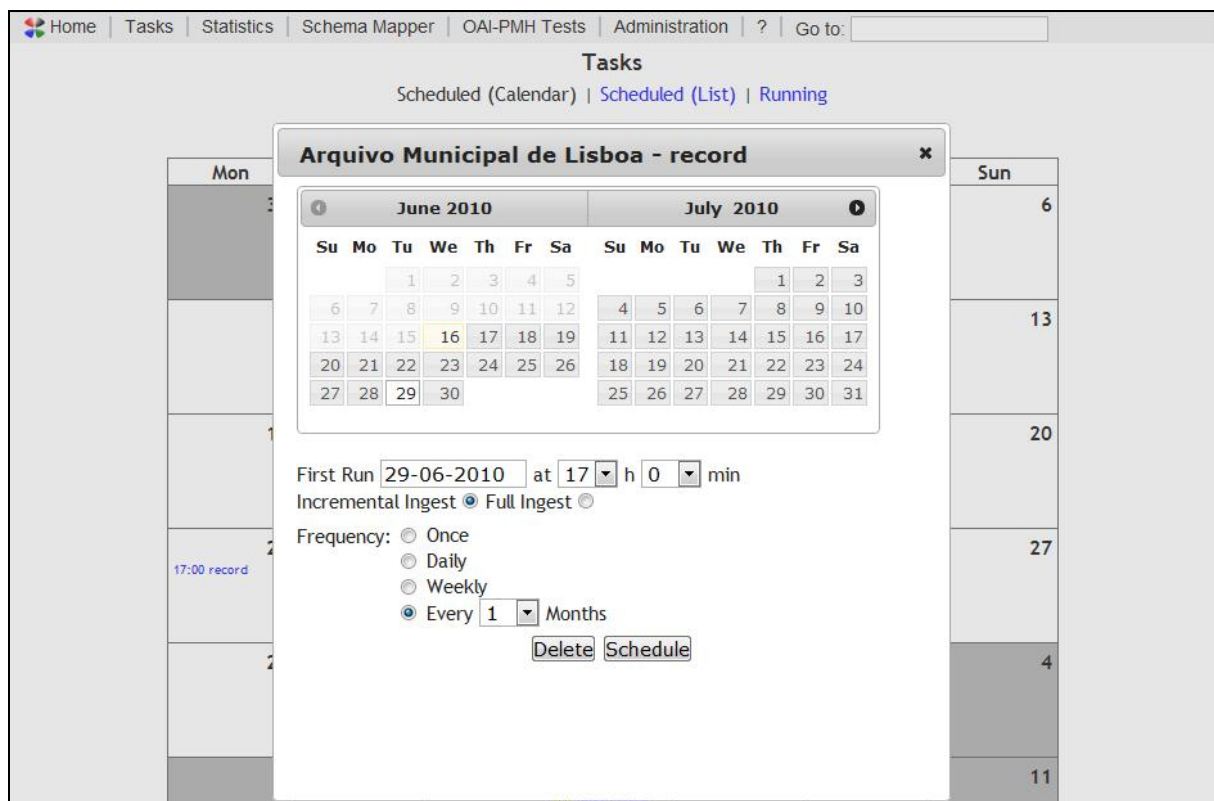


Figure 49 – Window to edit or delete a task

Change the scheduling information and click on “Schedule” button to save the task information.

Click on the “Delete” button, if you want to delete the task. A pop-up menu appears in this case. Click on the “Ok” button to confirm the task deletion or on “Cancel” button to cancel the deletion operation.

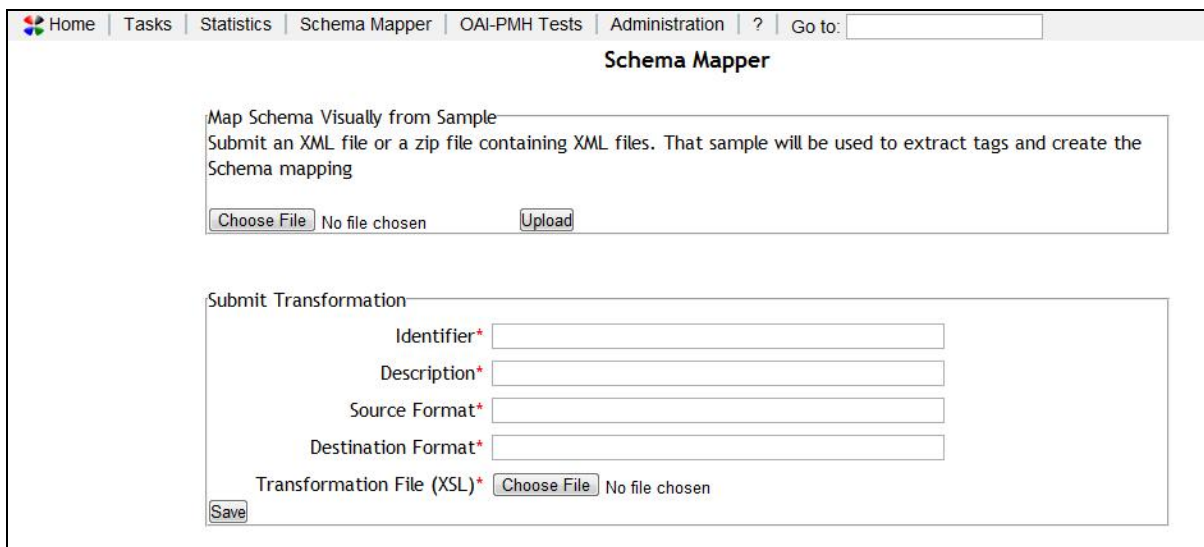
## 2.7 Schema Transformation Mapper

### 2.7.1 Add Schema Transformation

To add a XSLT transformation file (transform a Data Set from its original schema to another) click on the Schema Mapper button on the top menu in the homepage (Figure 50)

Figure 50 – Schema Mapper button

Fill the Submit Transformation form presented in the Figure 51.



The screenshot shows a web browser window with a navigation menu at the top: Home | Tasks | Statistics | Schema Mapper | OAI-PMH Tests | Administration | ? | Go to: [input field]. The main content area is titled "Schema Mapper" and contains two sections. The first section, "Map Schema Visually from Sample", includes the instruction "Submit an XML file or a zip file containing XML files. That sample will be used to extract tags and create the Schema mapping" and a file upload area with a "Choose File" button (showing "No file chosen") and an "Upload" button. The second section, "Submit Transformation", contains four required text input fields: "Identifier\*", "Description\*", "Source Format\*", and "Destination Format\*". Below these is a "Transformation File (XSL)\*" field with a "Choose File" button (showing "No file chosen") and a "Save" button.

**Figure 51 – Form to submit a schema transformation**

The form has the following fields:

- Identifier: the identifier that uniquely identifies the transformation.
- Description: the description of the transformation.
- Source Format: the source format of the transformation. In the actual version REPOX is not controlling these terms, so it is a free text.
- Destination Format: the destination format of the transformation. In the actual version REPOX is not controlling these terms, so it is a free text.
- Transformation File (XSL): click on the "Choose file Button to upload the XSLT file.

## 2.7.2 Visual Schema Mapper

REPOX allows creating a schema transformation to OAI\_DC using a visual interface. This is a limited functionality in this version, but it can prove to be useful for many scenarios.

To define a schemas' mapping click on the “Choose File” button showed on Figure 52.

The process assumes that a file with at least one record in the original format is available, to extract the XPath information. To select that click on “sample file” and press “Upload”.

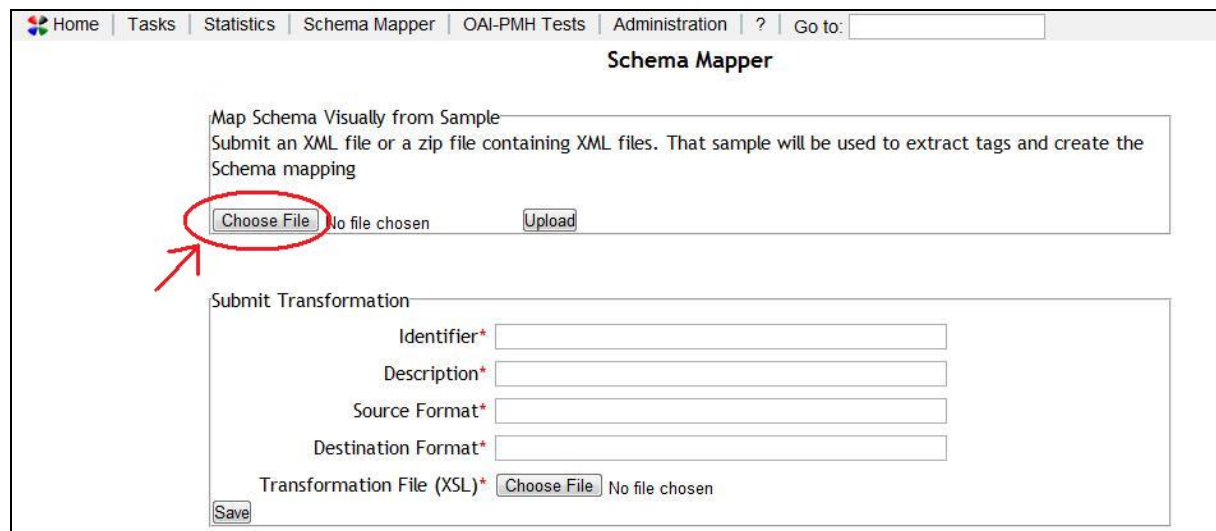


Figure 52 – “Choose file” button

REPOX, then presents an interface that allows the matching the OAI tags with the tags of the given sample file (Figure 53).

The column on the left shows the tags from the OAI\_DC schema, while the right column shows the tags and the respective XPath of the source format. To match the tags simply connect the tags from one column to the other column.



Figure 53 – Connected tags



This interface also allows the user to add new tags, using the “New Tag” buttons present on the top of each column. This can be useful for tags not existing in the sample, but that are expected to be found in new ingested data sets of this schema.

After connecting the tags, click on “Submit” to create and register the new transformation.

## 2.8 Statistics

To view statistics of the current REPOX installation click on the Statistics button on the top menu in the homepage, as showed in Figure 54.

Aggregator ID	Aggregator Name ...Create...	Provider ID	Provider Name	Data Set Name	OAI-PMH Schemas	Ingest Type	Last Ingest Date	Next Ingest Date	Number of Records	Ingest Status
ptAgg	Portuguese Aggregator	4567	Arquivo Municipal de Lisboa	arquivocmbx	ese	OAI-PMH ese	2010-10-29		1499	✓

Figure 54 – Statistics button

REPOX will show a table presenting statistics (Figure 55).

REPOX Statistics	
Aggregators	1
Data Providers	1
Data Sets	1
Type	
OAI-PMH	1
Folder	0
Metadata Format	
ese	1
ID Policy	
ID Extracted	0
ID Generated	1
Records	999
Averages	
Average records/Data Provider	999
Average records/Data Set	999
Countries	
PORTUGAL	999
Statistics Generation Date: Wed Jun 16 14:58:31 BST 2010	

Figure 55 – Statistics table

The table has several rows showing the data contained on the current REPOX installation:

- Aggregators: indicates the total number of Aggregators.
- Data Providers: indicates the total number of Data Providers.
- Data Sets: indicates the total number of Data Sets.
- Type: indicates the number of Data Sets of each import type (OAI-PHM, Folder or Z39.50).



- Metadata Format: indicates the number of Data Sets by metadata format.
- ID Policy: indicates the total number of Data Sets by the ID Policy. The policy can be ID Extracted or ID Generated.
- Records: indicates the total number of harvested records.
- Averages: indicates the average number of records by Data Provider and the average number of records by Data Set.
- Countries: indicates the total number of records per country (according to their Data Providers).

## 2.9 Search for a specific Data Provider or Data Set

The “Go to” field on the top menu of the homepage allows the administrator to search for a specific Data Provider or Data Set (Figure 56).

Simply type the name to search for and press the return button. If it is found, REPOX presents the page showing the information of that entity.

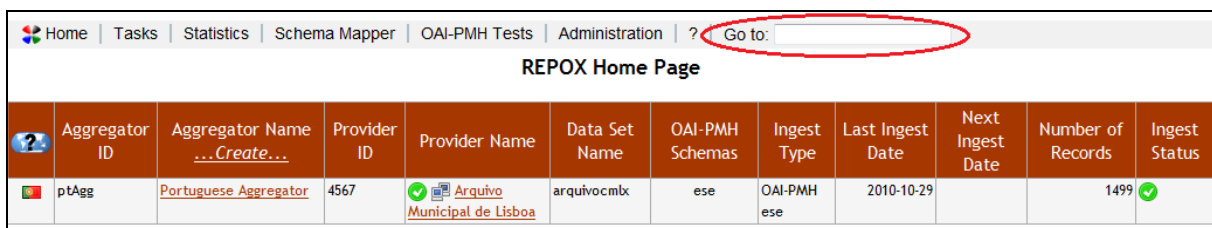


Figure 56 – Go to field

## 2.10 OAI-PMH Server

The OAI-PMH server of the current REPOX installation can be tested. For that, click in “OAI-PMH Tests” in the homepage.



Figure 57 – OAI-PMH Tests link



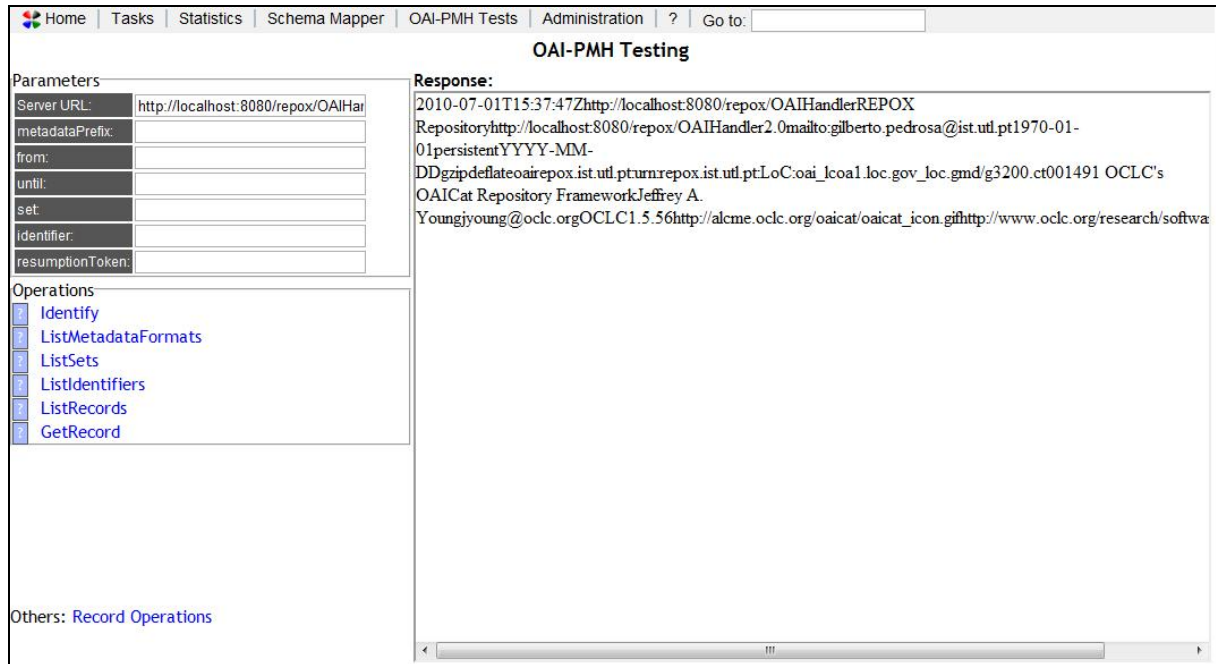


Figure 58 – OAI-PMH testing page

The OAI-PMH testing page allows the administrator to test several operations of the OAI-PMH server of the current REPOX installation (Figure 58). These operations are listed in the “Operations” box. It is also possible to specify the parameters for these operations using the fields in the “Parameters” box.