GeoNetwork User's Guide

Author: **ANDREA CARBONI**

Revision 3 (23-Apr-2007)

Contents

Introduction

The purpose of this document is to explain how to use, configure and administer GeoNetwork using its graphical tools. These tools are subdivided into 2 main categories:

- **web tools** : these are directly accessible from the web interface and allow the user to tune almost all aspects of GeoNetwork.
- **GAST tools** : these are special tools that cannot be executed while GeoNetwork is running or that perform special tasks better done in a separate tool.

Part I

Web interface

Chapter 1

Basic configuration

1.1 System configuration

Most of the GeoNetwork's configuration parameters can be changed using the web interface. Those parameters that cannot be changed through the web interface can be usually changed using the GAST application. To reach this web interface, you must be logged in as an Administrator. The page's link is located into the administration page (see figure 1.1. The link is surrounded with a red rectangle).

Clicking the page's link you will get the set of parameters that you can change (see figure 1.2). Here follows a detailed description of them:

Site General site parameters.

- **name** The name of the GeoNetwork's installation. This name will be used to identify the node in operations like the harvesting.
- **organization** The organization the node belongs to. Just for informative reasons.

Server Here you have to enter the address of your GeoNetwork's node. This address is important because it will be used to access the node.

- **host** The node's address or IP number. If your node is publicly accessible from the Internet, you have to use the machine's domain/address. If your node is hidden into your private network and you have a firewall or web server that redirects incoming calls to the node, you have to enter the public address of the firewall or web server. A typical configuration is to have an Apache web server on address **A** that is publicly accessible and redirects the requests to a Tomcat server on a private address **B**. In this case you have to enter **A** in the host parameter.

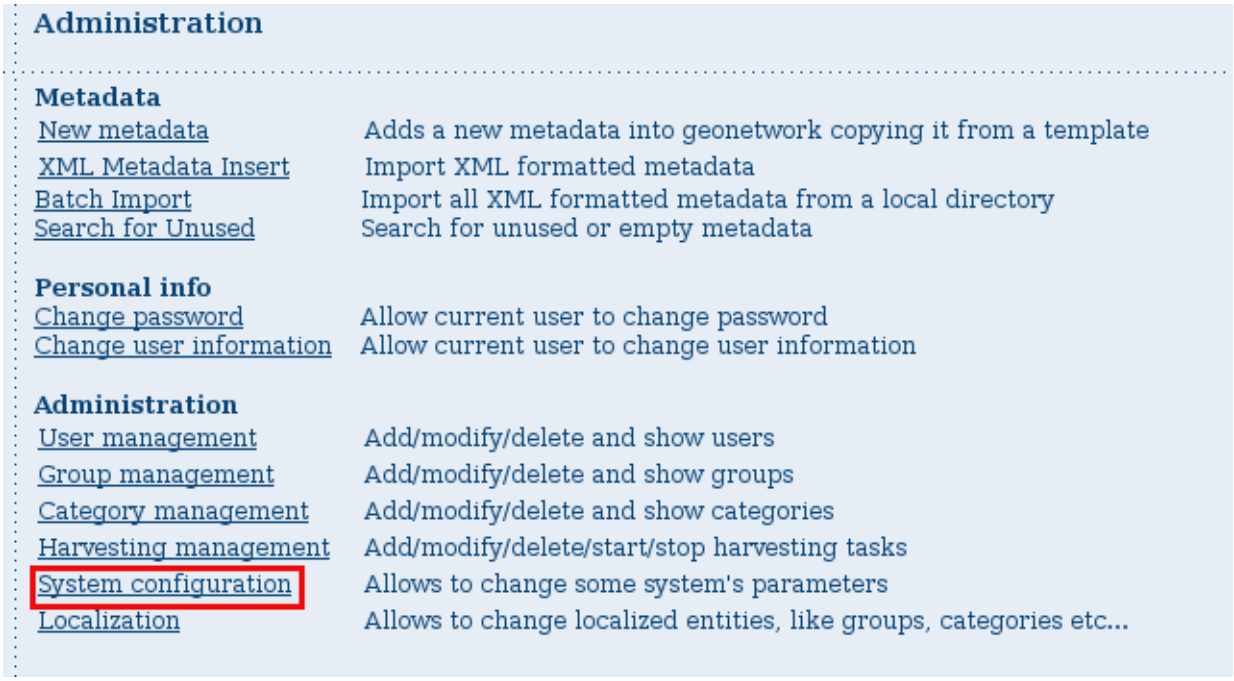

Figure 1.1: How to reach the configuration page

port The node's port (usually 80 or 8080). If the node is hidden, you have to enter the port on the public firewall or web server.

Intranet A common need for an organization is to discriminate between internal anonymous users (users that access the node from within the organization) and external ones (users from the Internet). Node's administrators can specify different privileges for internal and external anonymous users and, in order to do so, they have to specify the parameters of the internal network.

network The internal network's address in IP form.

netmask The network's mask.

Z39.50 GeoNetwork can act as a Z39.50 server, which is an OGC communication protocol to query and retrieve metadata.

enable check this option to start the Z39.50 submodule. Please, notice that GeoNetwork must be restarted in order to make this change active.

port This is the port on which GeoNetwork will be listening for incoming Z39.50 requests. Usually, the value of 2200 is a standard one, but to have multiple GeoNetwork nodes on the same machine you have to change this value in order to avoid port conflicts between the different nodes.

Proxy In some occasions (like harvesting) GeoNetwork must be able to connect to remote sites and this may be denied if an organization uses proxy servers. In this cases, GeoNetwork must be configured to use the proxy server in order to route outgoing requests.

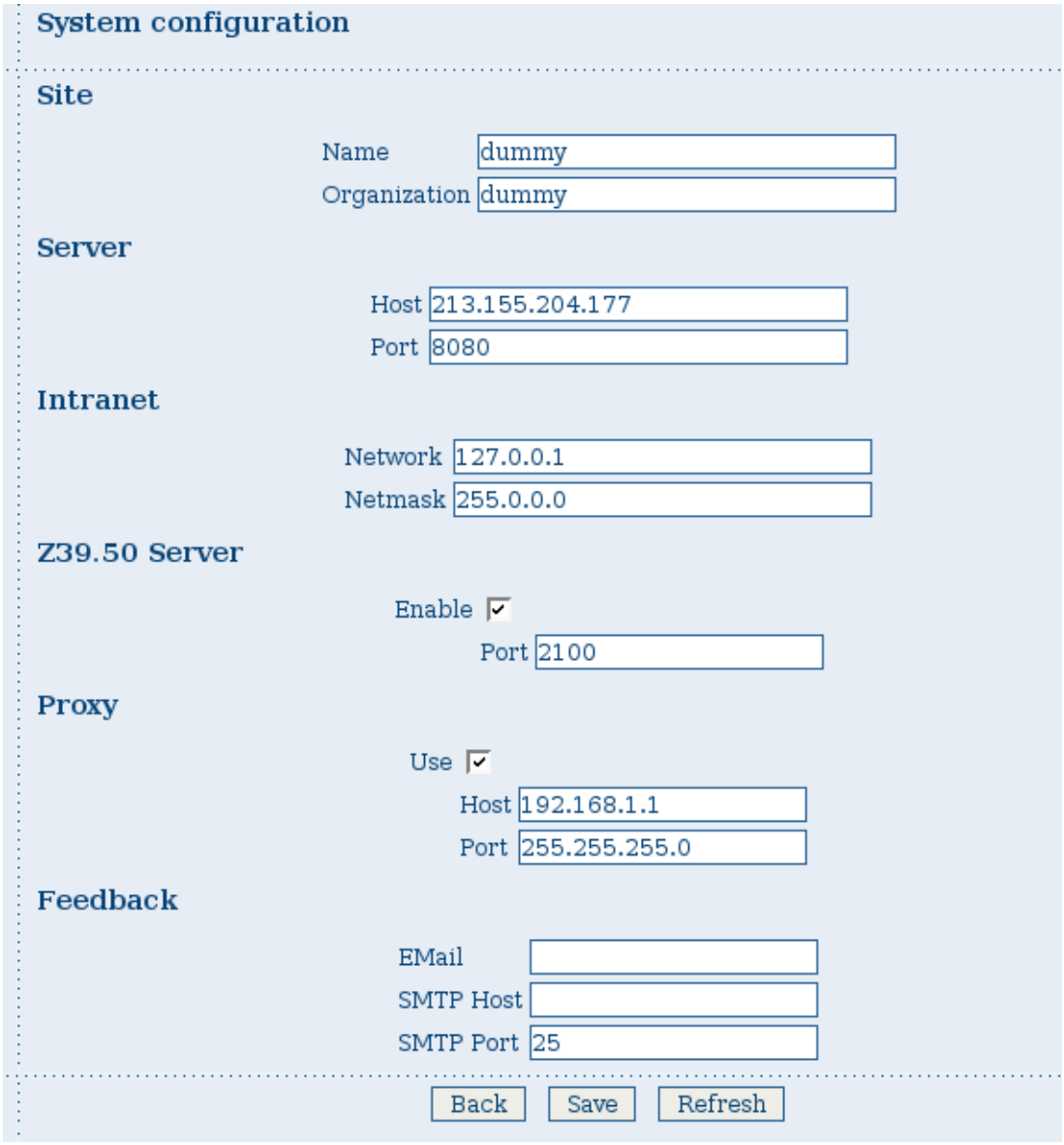

Figure 1.2: The configuration options

host The proxy's name or address to use (usually an IP address).

port The proxy's port to use.

Feedback GeoNetwork can provide some feedback to the administrator if some events happen, for example a metadata is downloaded or some feedback is provided using the online form. You have to configure GeoNetwork with proper parameters to send an email.

email This is the email address that will be used to send feedback to the administrator.

SMTP host Organization's mail server to use when sending emails

SMTP port Mail server's SMTP port (usually 25).

At the bottom of the page there are some buttons with the following purpose:

Back Simply returns to the main administration page.

Save Saves the current options. If some options are invalid, the system will show a dialog with the wrong parameter and will focus its textfield on the page. Once the configuration is saved a success dialog will be shown.

Refresh This button simply refreshes the displayed options taking the new values from the server. This can be usefull if some options get changed dynamically (for example by another user).

Public host and port usage

Up to now, the server's host and port are used in these cases:

- During an editing session, when adding data links to a metadata. The host and port will be used to build download links to store inside the metadata.
- During CSW requests. The GetCapabilities operation returns an XML document with HTTP links to the CSW services. These links are dynamically built using the host and port values.

1.2 Localization

The GeoNetwork's user interface can be easily localized into several languages just adding a few XML files. Anyway, beside static labels on the web page there are some dinamic entities that can be created and changed for which a simple XML file is not enough. These entities are stored inside the database and their localization can be performed using a specialized web form which can be reached from the administration page (see figure 1.3).

This form allows you to localize all entities whose structure is similar: a short non localized name and several localized labels, one for each supported language. Up to now, you can localize the following entities: *groups*, *categories*, *operations* and *regions*. The form is shown in figure 1.4 and is subdivided into a left and a right panel.

The left panel allows you to choose which entity you want to edit. On the top, a dropdown let you choose which entity type to edit. All entities of a choosen type are shown in a list on the bottom part of the panel (actually, the short names are listed).

The right panel allows a translator to read the label in a source language and translate it into a target language. The user can change both the source and the target languages using the provided dropdown

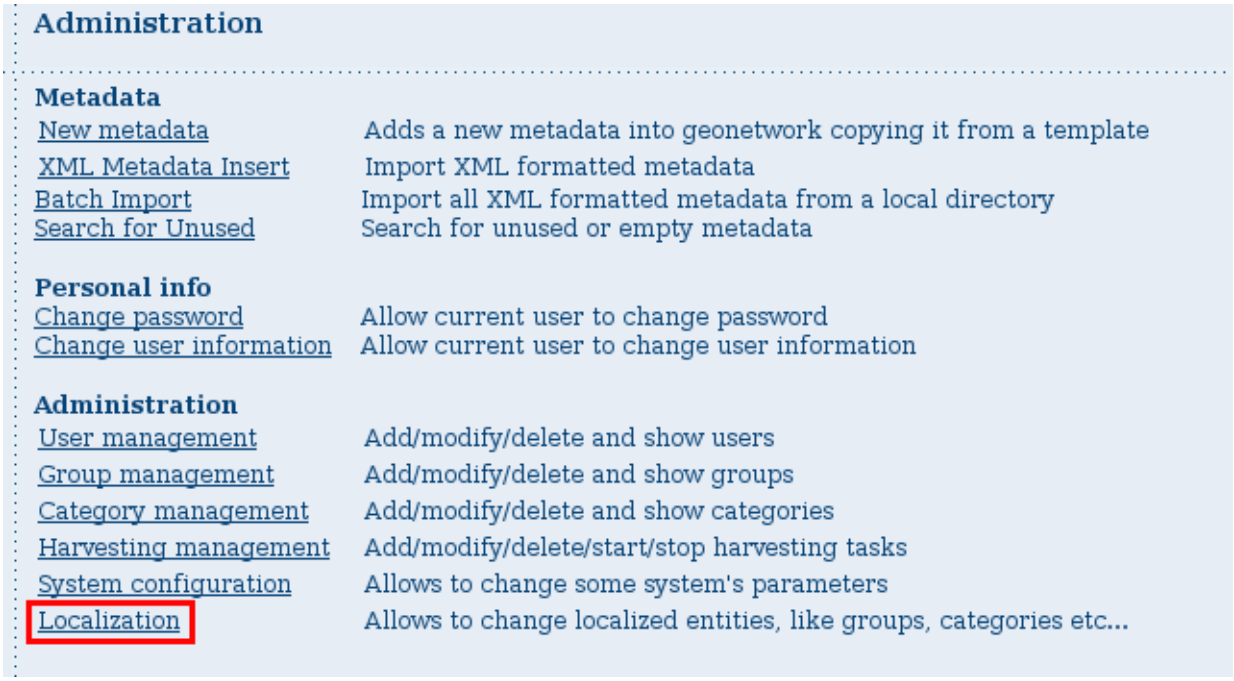

Figure 1.3: How to reach the localization page

controls. The label in the source language is read only while the one in the target language is the one that needs to be localized. When the user changes the target label, the textarea becomes yellow to indicate a change. When the label becomes equal to the initial one again, the textarea becomes white.

On the bottom of the form there are some buttons:

Back Simply returns to the administration page.

Save Saves the current target label to the server and advances to the next one. The normal workflow is to select first both a source and a target language. Then, the user goes through all entities localizing on at time. It is important to notice that if the user changes a label and chooses another target language, the label change is lost so it is better to localize labels in one language at time.

Refresh Reloads all entities of the selected type and any previous cache of other loaded types is invalidated. This button is useful when the number of entities changes (added or removed by other users).

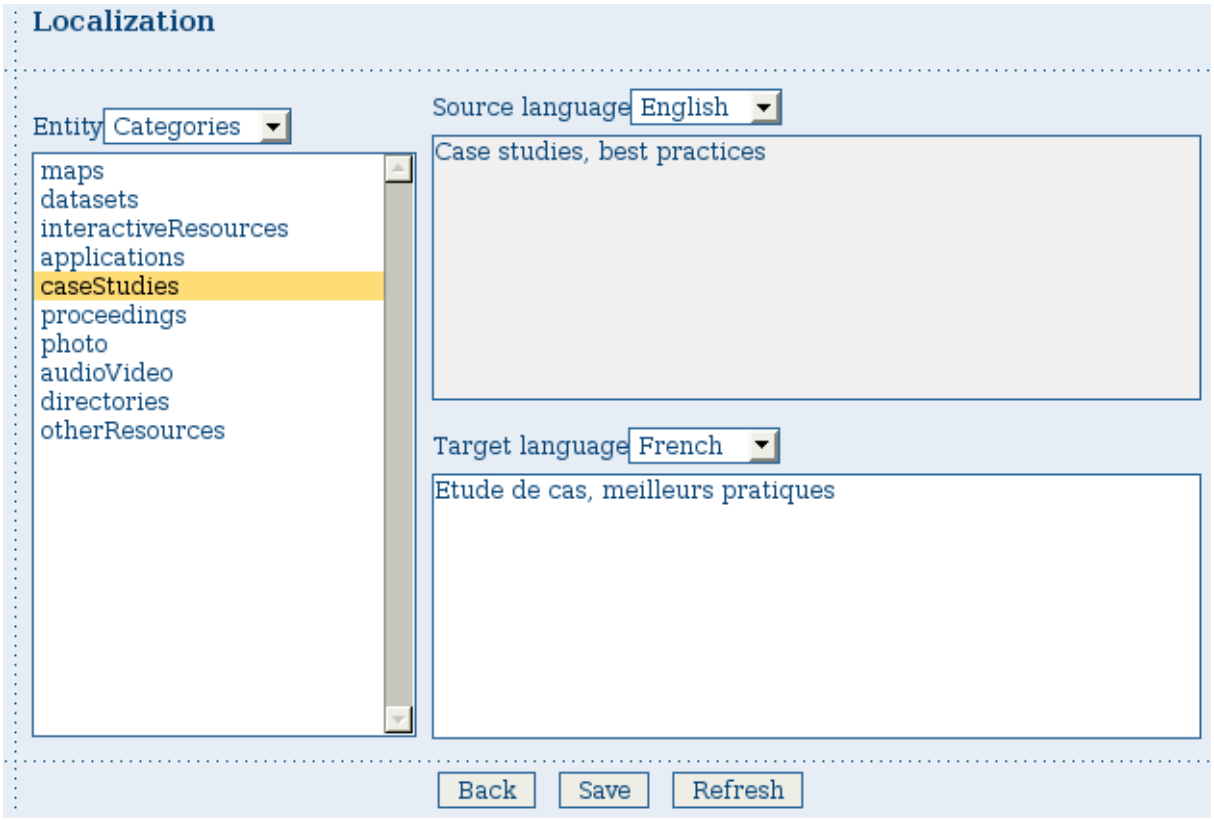

Figure 1.4: The localization form

Chapter 2

Import facilities

2.1 Batch import

The batch import facility allows you to import a set of metadata into the system all at once. In order to use this facility, you have to be logged in as an administrator. After the login step, go to the administration page and select the batch import's link (see figure 2.1. The link is surrounded with a red rectangle).

Clicking the link, you will reach the batch import's page as illustrated in figure 2.2. You have to specify a set of parameters to make the import working. They are:

Directory This is the full path on the server's file system of the directory to scan. GeoNetwork will look for and try to import all XML files present into this directory. It is important to notice that this is the directory on the *server* machine and *not* on the client of the user that is doing the import. All metadata files present into the import directory *must* have the same schema format.

Schema GeoNetwork supports only some metadata formats so you have to specify the schema of the metadata you want to import. If a metadata does not belong to the selected schema, the entire operation will be aborted.

Validate This is a simple validation step that you can choose to perform. The metadata is validated against its schema.

Group You have to select a group to associate to the imported metadata. Usually the group is the creator of the metadata set.

Category You can specify one category to associate to your metadata in order to simplify the search.

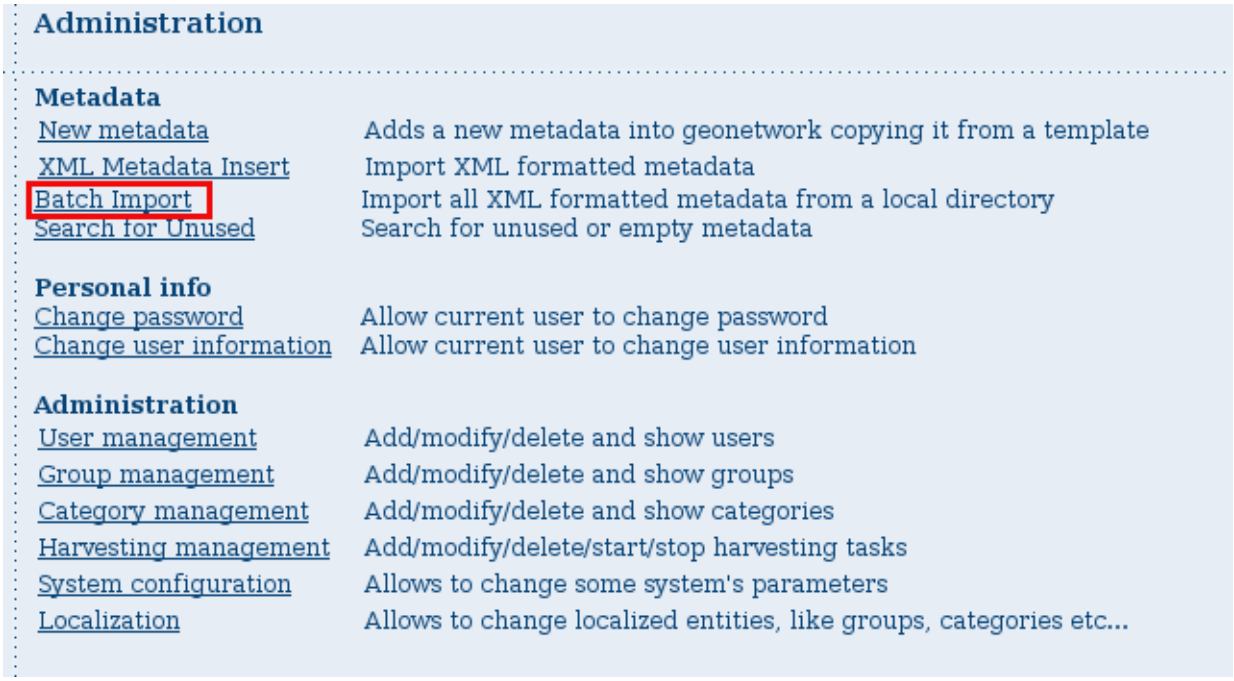

Figure 2.1: How to reach the batch import page

Stylesheet This is a powerfull option because allows you to specify a stylesheet for an XSL transformation. The drop down control is filled with files taken from the web/xsl/conversion/import folder : all XSL files you put there will be made available. This is a dymanic process so you don't have to restart GeoNetwork. The purpose of this option is to allow the conversion of a metadata into a suitable format that is supported by GeoNetwork. Therefore, it is important that the result of the transformation matches the schema format selected above.

Below the page, there are the following buttons:

Back Goes back to the administration form.

Upload Starts the import process. When the process ends, the total count of imported metadata will be shown. Please notice that the import is transactional: the metadata set will be fully imported or fully discarded (there are no partial imports). Files that starts with '.' or that do not end with '.xml' are ignored.

2.1.1 Structured import

An hidden feature of the batch import is the possibility to specify some import parameters in more detail. This feature is triggered when the specified folder contains the **import-config.xml** file. When this happen, this file is read and the standard import switches to the structured one.

The import-config.xml file has a **config** root element with the following children:

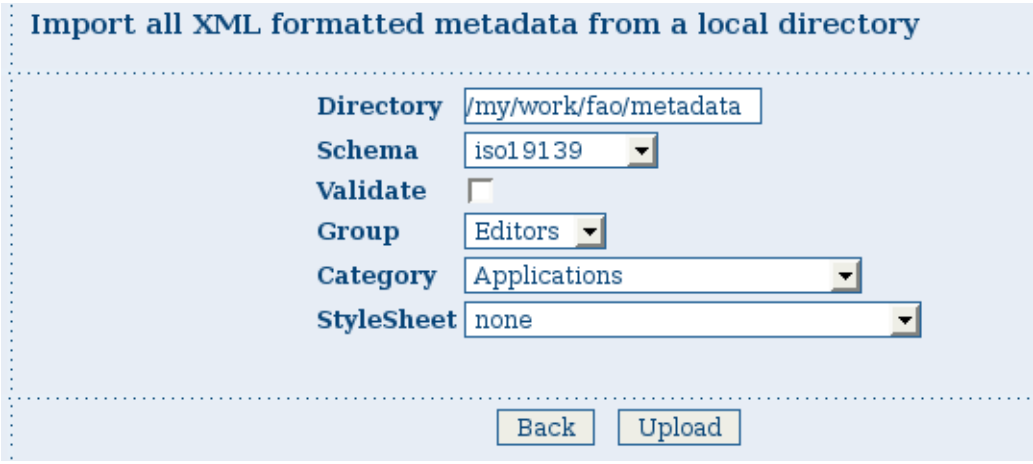

Figure 2.2: The batch import options

- **categoryMapping** [1] : this element specifies the mapping of directories to categories.
  - **mapping** [0..n] : This element can appear 0 or more times and maps one directory name to a category name. It must have a **dir** attribute that indicates the directory and a **to** attribute that indicates the category name.
  - **default** [1] : This element specifies a default mapping of categories for all directories that do not match the other mapping elements. It must have only the **to** attribute.
- **schemaMapping** [1] : this element specifies the mapping of directories to metadata schemas.
  - **mapping** [0..n] : This element can appear 0 or more times and maps one directory to the schema name that must be used when importing. The provided schema must match the one used by the metadata contained into the specified directory, which must all have the same schema. It must have a **dir** attribute that indicates the directory and a **to** attribute that indicates the schema name.
  - **default** [1] : default behaviour to use when all other mapping elements do not match. It must have only the **to** attribute.

Here is an example of the **import-config.xml** file:

```
<config>
   <categoryMapping>
      <mapping dir="1" to="maps" />
      <mapping dir="3" to="datasets" />
      <mapping dir="6" to="interactiveResources" />
      <mapping dir="30" to="photo" />
      <default to="maps" />
   </categoryMapping>
   <schemaMapping>
      <mapping dir="3" to="fgdc-std" />
      <default to="dublin-core" />
   </schemaMapping>
</config>
```

The import procedure starts by scanning the provided directory. This can contain, beside the **import-config.xml** file, only subdirectories which name will be ignored but used only as a container. Inside each directory, there is another level made only by directories that represent a metadata grouping for categories. Each directory name will be used as the **dir** attribute in the mapping scheme previously described.

Chapter 3

Harvesting

3.1 Introduction

Since the beginning of the project, there has been the need to share metadata among several GeoNetwork nodes. Usually, each node takes care of a region of interest so it is important to be able to perform a search over all these nodes at the same time. This is called distributed search and exploits the Internet connectivity. In our cases, this distributed search can be heavy to perform if there are many maps with associated thumbnails Furthermore, GeoNetwork is usually employed in countries (like Africa, Brazil) where the connectivity is bad, making the use of distributed search not feasible.

The harvesting is the process of collecting remote metadata and storing them locally for a faster access. This is a periodic process to do, for example, once a week. Harvesting is not a simple import: local and remote metadata are kept aligned. Using some *magic*, GeoNetwork is capable of discovering metadata that have been added, removed or updated.

GeoNetwork is able to harvest from the following sources:

- Another GeoNetwork node
- A web accessible folder from the Internet. This is a simple page with links to xml files that represent the metadata.

3.2 Mechanism overview

The harvesting mechanism is based on the concept of *universally unique identifier* (uuid). This is a special id because it is not only unique locally to the node that generated it but it is unique across all the world. It is a combination of the network interface's MAC address, the current date/time and a random number. Every time you create a new metadata in GeoNetwork, a new uuid is generated and assigned to it.

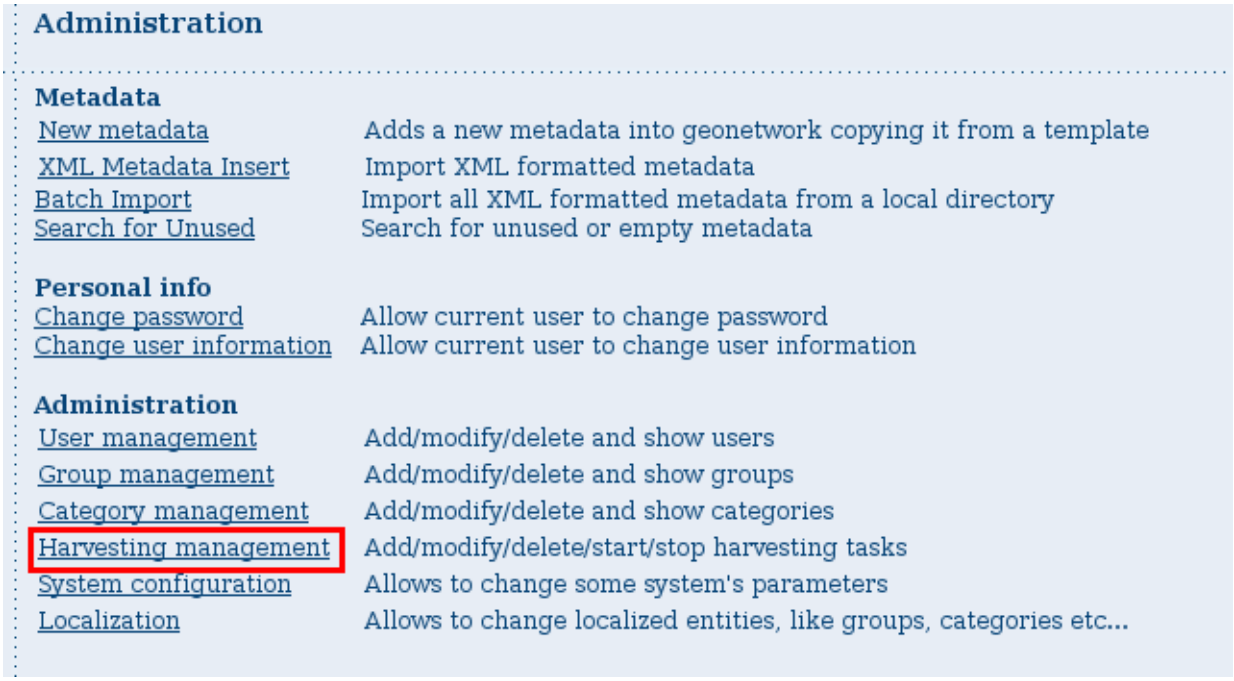

Figure 3.1: How to reach the harvesting main page

Another important concept behind the harvesting is the *last change date*. Every time you change a metadata, its last change date is updated. Just storing this parameter and comparing it with a new one allows any system to find out if the metadata has been modified since last update.

These two concepts allow GeoNetwork to fetch a remote metadata, to check if it has been updated, to remove it locally if it has been removed remotely. Furthermore, thanks to uuids, a hierarchy of harvesting nodes can be built where B harvests from C and A harvests from B. Even loops can be created because harvested metadata cannot be modified.

3.3 Tha main page

To reach the harvesting main page you have to be logged in as an administrator. From the administration page, click the link shown in figure 3.1 with a red rectangle.

Figure 3.2 shows the harvesting main page. The page shows a list of harvesting nodes that have been created so far. On the bottom side there is a set of buttons to manage these nodes. The meaning of each column is as follows:

Select This is just a checkbox to select one or more nodes. The selected nodes will be affected by the first row of buttons (start, stop, run, remove). For example, if you select 3 nodes and press the **Remove** button, these 3 nodes will be removed.

Name This is the node's name provided by the administrator.

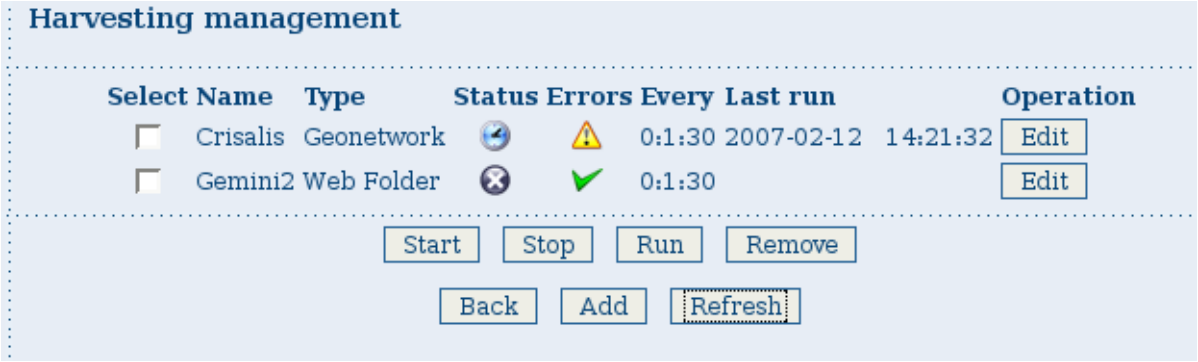

Figure 3.2: The harvesting main page

Type The node's harvesting type choosed when the node was created (geonetwork, web folder etc...).

Status This is an icon that reflects the node's current status. See figure 3.3 for all different icons and status description.

Errors This column reflects the status of the last harvesting run, which could have succeeded or not. The result is reflected on this icon and a tooltip will show detailed information. See figure 3.4 for all different icons.

Every Time (in days, hours, minutes) between two consecutive harvesting from this node.

Last run The date, in ISO 8601 format, of the most recent harvesting run.

Operation A list of buttons for all possible operations on a node. At least, an **Edit** button is present. Pressing this button you will reach the same page used to create the node. From this page you can change the desired parameters.

The bottom side of the page contains two rows of buttons. The first row contains buttons that can operate on a set of nodes. You can select the nodes using the checkbox on the first column and then press the proper button. When the button finishes its action, the checkboxes are cleared. The second row contains general purpose buttons. Here is the meaning of each button:

Start When a new harvesting node is created, it's status is *inactive*. Use this button to make it *active* and thus to start harvesting from the remote node.

Stop Stop harvesting from a node. Please notice that this does not mean that a currently running harvesting will be stopped but it means that this node will be ignored in the future.

Run This button simply tells the harvesting engine to start harvesting immediately. This is usefull for testing during the harvesting setup.

Icon	Status	Description
	Inactive	The harvesting from this node is stopped.
	Active	The harvesting engine is waiting for the timeout specified for this node. When the timeout is reached, the harvesting starts.
	Running	The harvesting engine is currently running, fetching metadata from remote nodes. When the process will be finished, the status will be switched to active.

Figure 3.3: Possible status icons

Icon	Description
	The harvesting was ok, no errors were found. In this case, a tooltip will show some harvesting results (like the number of harvested metadata etc...)
	The harvesting was aborted due to an unexpected condition. In this case, a tooltip will show some information about the error.

Figure 3.4: Possible error icons

Remove Remove all currently selected nodes. A dialog will ask the user to confirm the action.

Back Simply returns to the main administration page.

Add This button allows to create new harvesting nodes. See section 3.4 for more information.

Refresh Refreshes the current list of nodes quering the server. This can be usefull to see if the harvesting list has been altered by someone else or if any harvesting process started.

3.4 Adding new nodes

The **Add** button in the main page allows you to add new harvesting nodes. After pressing it, you will reach the page shown in figure 3.5. When creating a new node, you have to choose the harvesting protocol implemented by the remote server. The supported protocols are:

Geonetwork This is the standard and most powerfull harvesting protocol used in GeoNetwork. It is able to log in into the remote node, to perform a standard search using the common queryable fields and to import all matching metadata. Furthermore, the protocol will try to keep both remote privileges and categories of the harvested metadata if they exist locally. Please notice that since GeoNetwork 2.1 the harvesting protocol has been improved. This means that it is not possible to use this protocol to harvest from version 2.0 or below.

Figure 3.5: Adding a new harvesting node

Web folder This is a weak protocol because the remote server does not offer harvesting facilities. Nevertheless, it can be usefull to users that want to publish their metadata through a web server. The protocol exploits the file list page returned by many web server when trying to access a page without an index file. Because this page contains file names, change dates and links, GeoNetwork is able to download them and keep them updated. Figure 3.6 shows an example of web accessible folder generated by the Apache web server.

The drop down list shows all available protocols. Pressing the **Add** button you will reach an edit page whose content depends on the choosed protocol. The **Back** button will go back to the main page.

Adding a GeoNetwork node

This type of harvesting allows you to connect to a GeoNetwork node, perform a simple search as in the main page and retrieve all matched metadata. The search is usefull because it allows to focus only on metadata of interest. Once, you add a node of this type, you will get a page like the one shown in figure 3.7. The meaning of the options is the following:

Site Here you put information about the GeoNetwork's node you want harvest from (host, port and servlet). If you want to search protected metadata you have to specify an account. The **name** parameter is just a short description that will be shown in the main page beside each node.

Search In this section you can specify search parameters: they are the same present in the main page. Before doing that, it is important to remember that the GeoNetwork's harvesting can be hierarchical so a remote node can contain both its metadata and metadata harvested from other nodes. Therefore, any search is limited to a site id. At the beginning, the **Site name** drop down is empty and you have to use the **Retrieve** button to fill it. The purpose of this button is to query GeoNetwork about all sites which it is currently harvesting from.

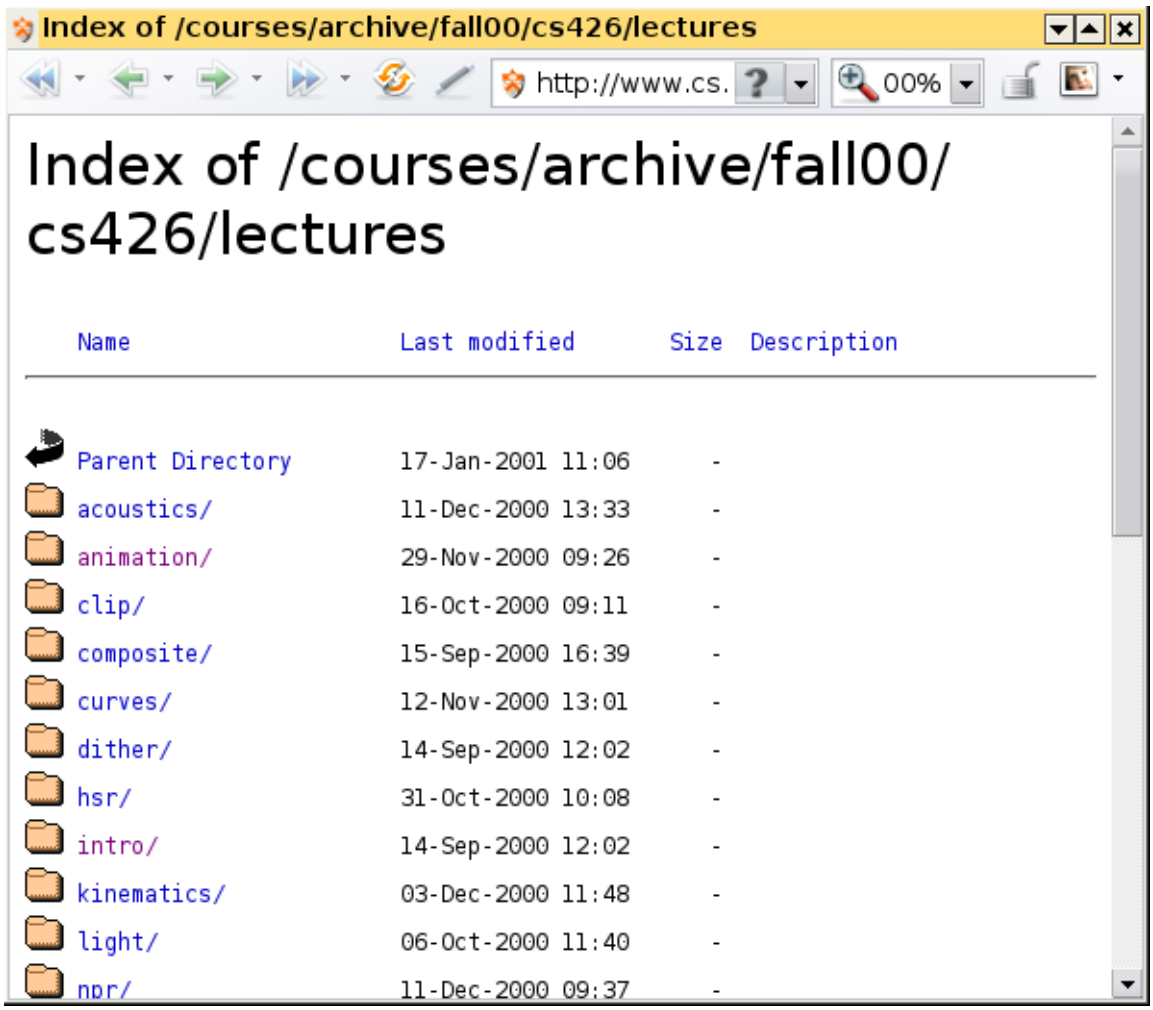

Figure 3.6: An example of a WAF page generated by the Apache web server.

Each entry has a number in parenthesis which is the number of metadata that the remote note has harvested from the other node. Once you get the drop down filled, you can choose a site name and press the **Add** button to create a new search entry that searches on that site. You can add only one search box for each site. Each search box can be removed pressing the small button on the left of the site's name.

Options This is just a container for general options.

Every This is the harvesting period. The smallest value is 1 minute while the greatest value is 100 days.

Create groups When importing metadata, privileges for the Intranet and Internet groups are kept. The other privileges are kept only if a mapping is possible. For each remote group, if there is a local group with the same name then the privilege is kept. If this option is checked, local groups will be created to keep privileges of the remote ones.

Create categories If a remote category exists locally, the metadata-category association is kept, otherwise is discarded. If this option is checked, GeoNetwork will

Figure 3.7: Adding a GeoNetwork node

create local categories to match the remote ones in order to keep the association.

One run only If this option is checked, the harvesting will do only one run after which it will become inactive.

On the bottom side of the page there are some buttons:

Back Simply return to the main harvesting page.

Save Saves the current setting and returns to the main harvesting page. When creating a new node, the node will be actually created only when you press this button.

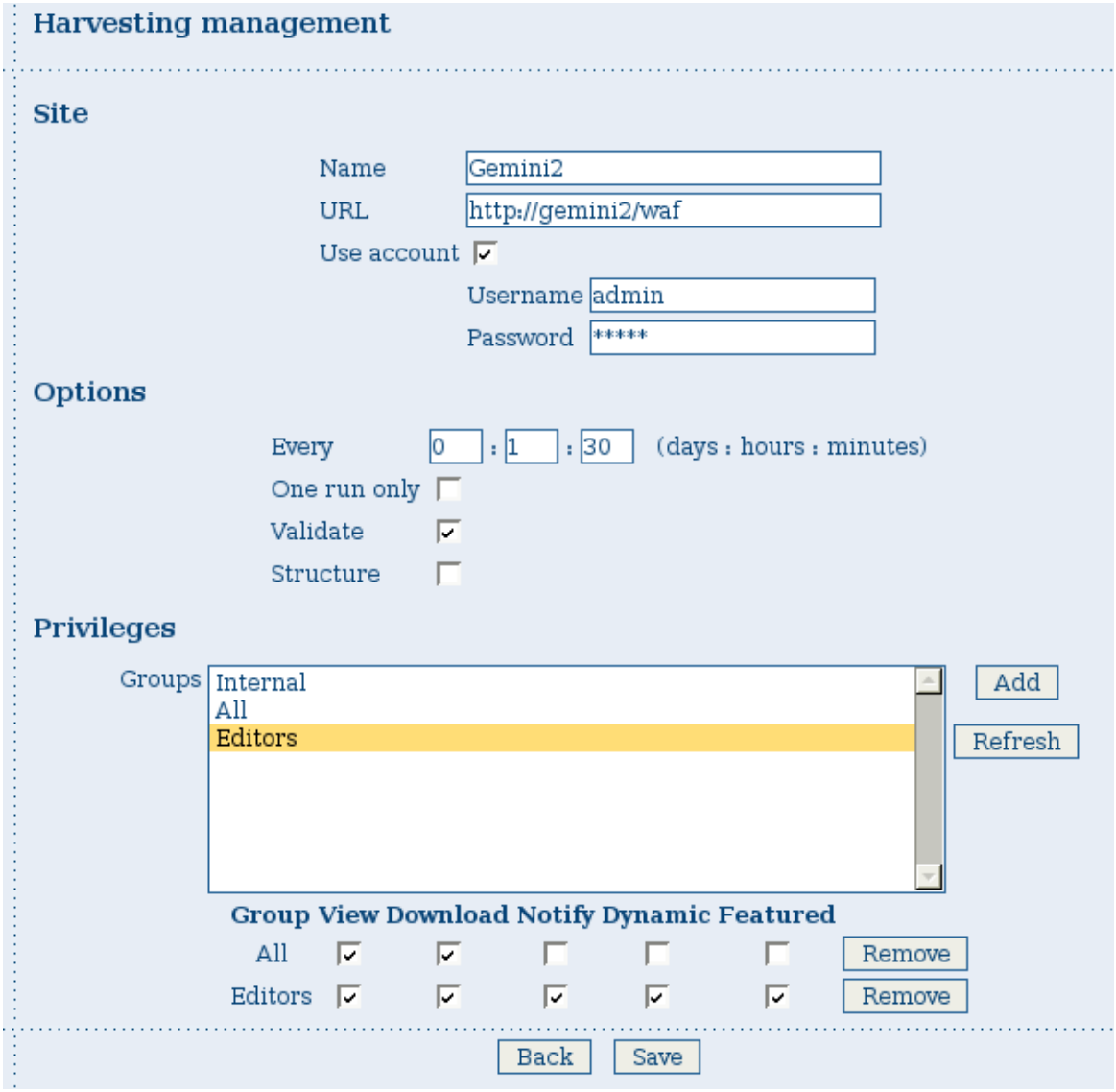

Figure 3.8: Adding a web accessible folder node

Adding a Web Assessible Folder node

In this type of harvesting, metadata are retrieved from a remote web page. The available options are shown in figure 3.8 and have the following meaning:

Site Here are the connection information. The available options are:

- **Name** This is a short description of the node. It will be shown in the harvesting main page.
- **URL** The remote URL from which metadata will be harvested. This URL must return a valid HTML page containing the metadata links. Each link that ends with **.xml** will indicate a metadata and the link content will be retrieved, converted into xml and imported.

Use account Account credentials for a basic HTTP authentication toward the remote URL.

Options General harvesting options

Every This is the harvesting period. The smallest value is 1 minute while the greatest value si 100 days.

One run only If this option is checked, the harvesting will do only one run after which it will become inactive.

Validate If checked, the metadata will be validate during import. If the validation does not pass, the metadata will be skipped.

Structure Reserved for future use

Privileges Here it is possible to assign privileges to imported metadata. The **Groups** area lists all available groups in GeoNetwork. The **Refresh** button can be used to refresh this list in case it has changed. One a group has been selected, it can be added through the **Add** button (each group can be added only once). For each added group, a row of privileges is created at the bottom of the list to allow privilege selection. To remove a row simply press the associated **Remove** button on its right.

At the bottom of the page there are the following buttons:

Back Go back to the main harvesting page. The harvesting is not added.

Save Saves node's data creating a new harvesting node. Then it will go back to the main harvesting page.

3.5 Mixing harvesting between different versions

Part II

GAST

Chapter 4

Introduction

4.1 What is GAST?

GAST stands for "GeoNetwork's Administrator Survival Tool" and is a standalone application whose purpose is to simplify some low level tasks like change of the servlet, configuration of the JDBC account, setup the database and so on. Most of the GAST's facilities work only for the GeoNetwork's installation where GAST is in. This implies that if you are using a servlet container other than Jetty (like Tomcat) you will not be able to change some options (like the servlet's name). Other facilities work for any servlet container but you have to specify the GeoNetwork's URL into the GAST's configuration dialog.

4.2 Starting GAST

GAST belongs to the core components so it is installed by default. It is located into the **gast** folder and to start it you can either use the **java** command or just click its jar's icon. To use the java command you have to:

- from the GeoNetwork's root folder, issue the command:

  ```
  java -jar gast/gast.jar
  ```

- from the gast folder, issue the command:

  ```
  java -jar gast.jar
  ```

To use your window manager, simply open the GeoNetwork's folder, go inside the **gast** folder and click the **gast.jar** icon. If you have Java installed, GAST should start in a few seconds.

To run, GAST needs at least Java 1.5. It will not work on Java 1.4 and it should run on Java 1.6 (anyway, it has not been tested).

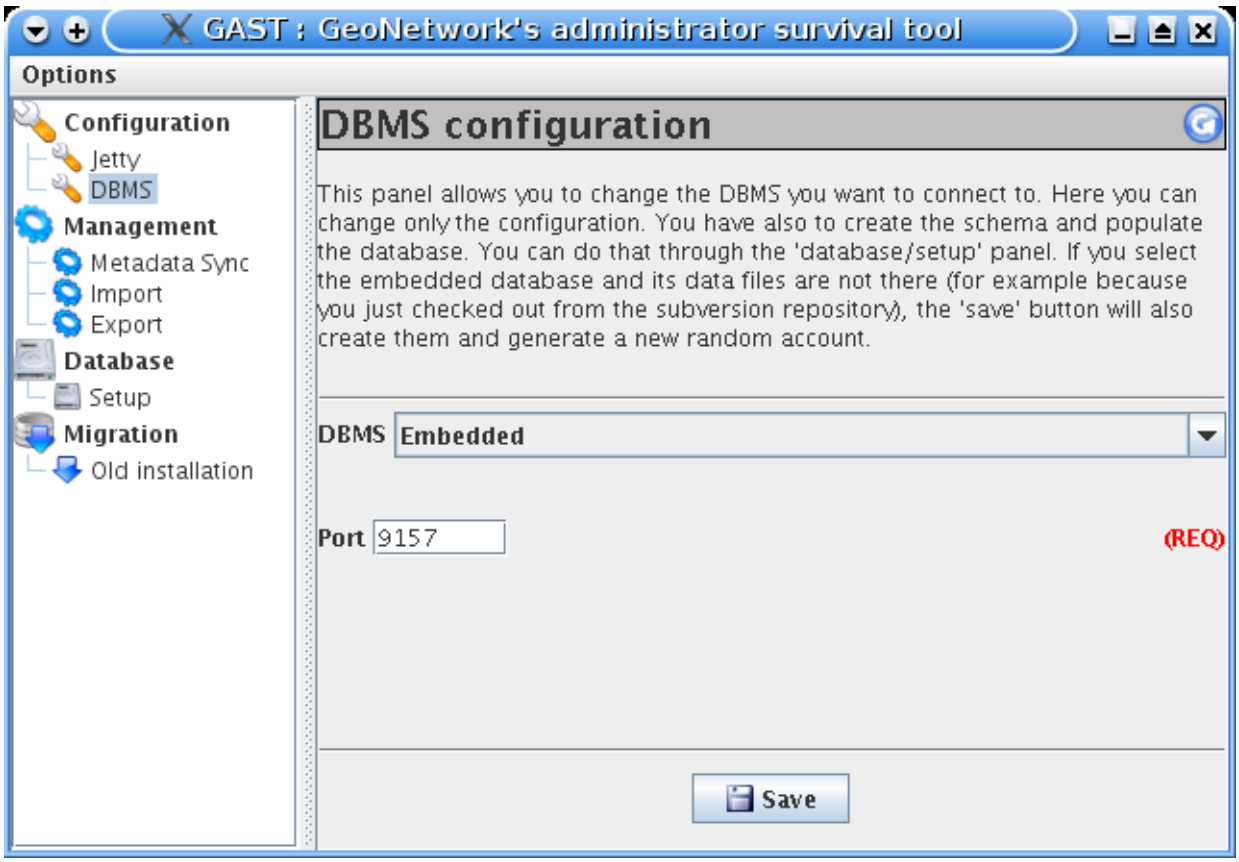

Figure 4.1: GAST's main window with a tool selected

4.3 Operating modes

When you start GAST, you get an application window like the one in figure 4.1. On the left side you have a panel with the tools you can use. After selecting a tool, on the right side you get the tool's options panel.

Every GAST tool has an **operating mode**, which defines the condition under which the tool can be used. The tool's mode is shown with an icon on the right side of the tool's name. The operating modes, with their icons are summarized in the following table:

Mode	Icon	Description
Restarted		The tool can be always used, but GeoNetwork must be restarted in order to make the change effective.
Running		The tool can be used only if GeoNetwork is running.
Stopped		The tool can be used only if GeoNetwork is stopped. This is important because some tools change the database's account or create the database from scratch. These are sensitive operations that cannot be performed while GeoNetwork is running.

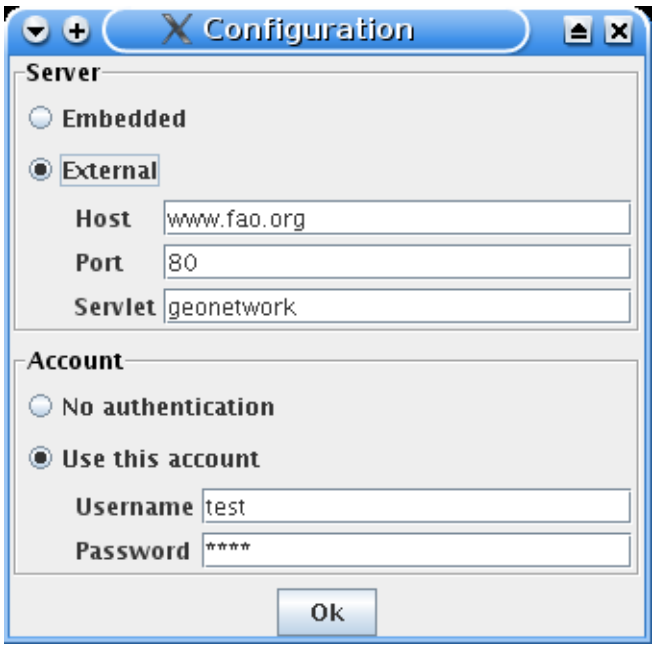

Figure 4.2: The configuration dialog

4.4 Tools subdivision

All GAST tools present into the left panel are logically subdivided into groups. Each group represents a GeoNetwork's aspect for which GAST allows you a graphic interface. The groups are:

Configuration You can change some configuration parameters, like the servlet's name, JDBC account etc...

Management General purpose tools related to the site's administration.

Database Operations that regard the database. Here you can find tools to create a database from scratch, creating the schema and filling it with proper data.

Migration Tools that allow you to migrate metadata from old installation.

4.5 Configuration dialog

Some of the GAST's tools access a running GeoNetwork. Usually, GAST connects to GeoNetwork using the connection parameters it finds on the installation folder but you can specify other parameters in order to connect to other instances. This is mandatory when the GeoNetwork instance is not running on the embedded Jetty server. In addition to that, some tools require authentication so account parameters must be provided.

To provide these parameters, you have to use the GAST 's configuration dialog. To open the dialog, simply select **Options** ▷ **Config** from the menubar. You will get the dialog shown in figure 4.2.

The dialog is subdivided into 2 areas:

Server Tells GAST how to connect to a running GeoNetwork. If you select the **embedded** option, GAST will get the connection parameters from the installation directory. Alternatively, if you use Tomcat or an external servlet container you have to choose the **external** option and provide the connection parameters yourself. Remember that this will work only for tools which operating mode is *Running*. For all the others, GAST will access the parameters from the installation directory.

Account Some tools require authentication. To authenticate, simply select the **Use this account** option and provide the username and password of a valid account. These parameters will work for both the embedded instance and for any external instance.

Chapter 5

Import / export tools

5.1 Introduction

Using GAST, you can import and export metadata as will. These facilities allow you a broad variety of tasks:

- Backup of the entire metadata set. Each metadata has its own file with maps and other data files. Once you have the backup, you can decide to import all or only some of them.
- Move your metadata from one GeoNetwork's installation to another. This can be done to mirror your metadata or to upgrade an old installation. In the last case, you export your metadata from your old installation and then reimport them into the new one.
- Fill the system with test data. Using the 'skip uuid' option, you can reimport the same metadata over and over again. This is usefull, for example, if you want to perform stress tests.

Metadata are exported using the MEF format.

5.2 Export

This facility is located inside the **Management** node on the left panel and allows you to export a set of metadata using the MEF format. Clicking the **Export** node you get its option panel on the right, as you can see from figure 5.1.

Here follows a detailed explanation of the options:

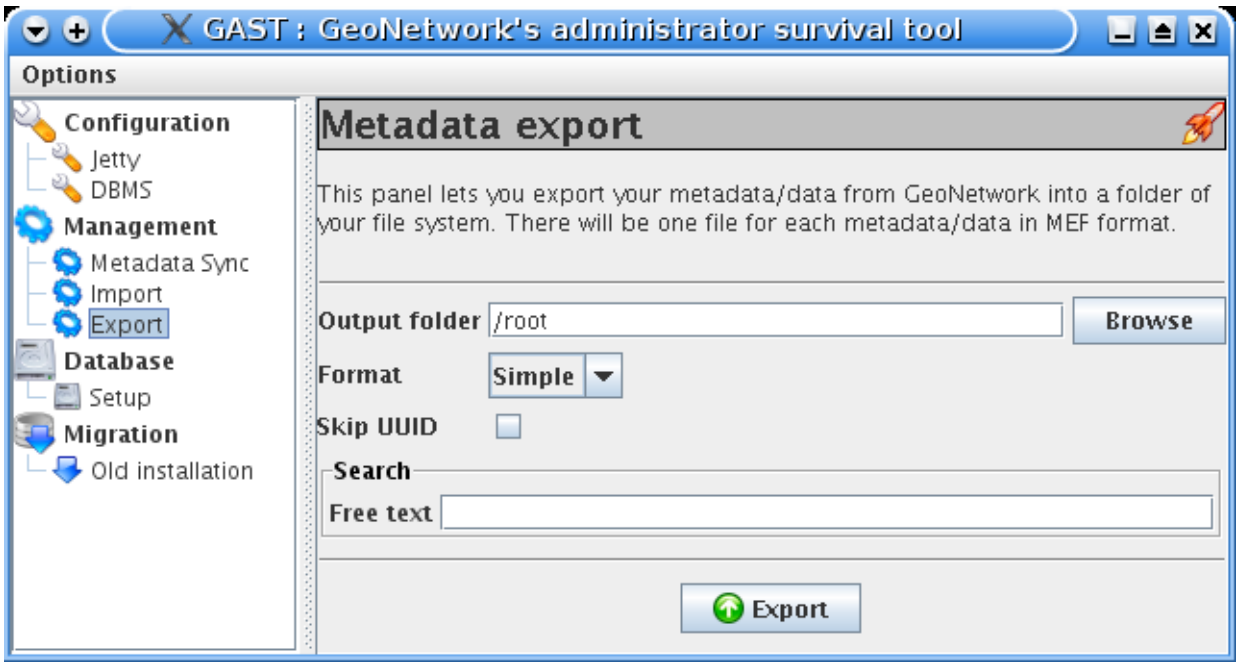

Figure 5.1: The metadata export panel

Output folder This is the target folder in your system where GAST will put the exported metadata. You can either press the Browse button to navigate through your file system to choose a better location or enter it manually into the textfield.

Format Here you can specify the metadata's output format. See the MEF specification for more information.

Skip UUID Normally is not selected. If you select it, you will loose the metadata's unique identifier (uuid) but you will be able to reimport that metadata over and over again. This is usefull to fill the system with test data.

Search This panel contains some search criteria that you can enter when exporting metadata.

Please notice that the export's result will depend on the metadata visible to the searching user. If you do not authenticate, you will get only public metadata.

Pressing the Export button will start the process. A progress dialog will be opened to show the export status.

5.3 Import

This facility is located inside the **Management** node on the left panel and allows you to import a set of metadata that have been previously exported using the export facility described above. Clicking the **Import** node you get its option panel on the right, as you can see from figure 5.2.

Figure 5.2: The metadata import panel

Here follows a detailed explanation of the options:

Input folder This is the source folder in your system that GAST will scan to collect metadata to import. GAST will try to import all files with the **mef** extension. Please notice that subfolders are not scanned. You can either press the Browse button to navigate through your file system to choose a better location or enter it manually into the textfield.

Pressing the Export button will start the process. A progress dialog will be opened to show the import status.