

# **GeoNetwork Remote Harvesting Capabilities**

Andrea Carboni

FAO

# Summary

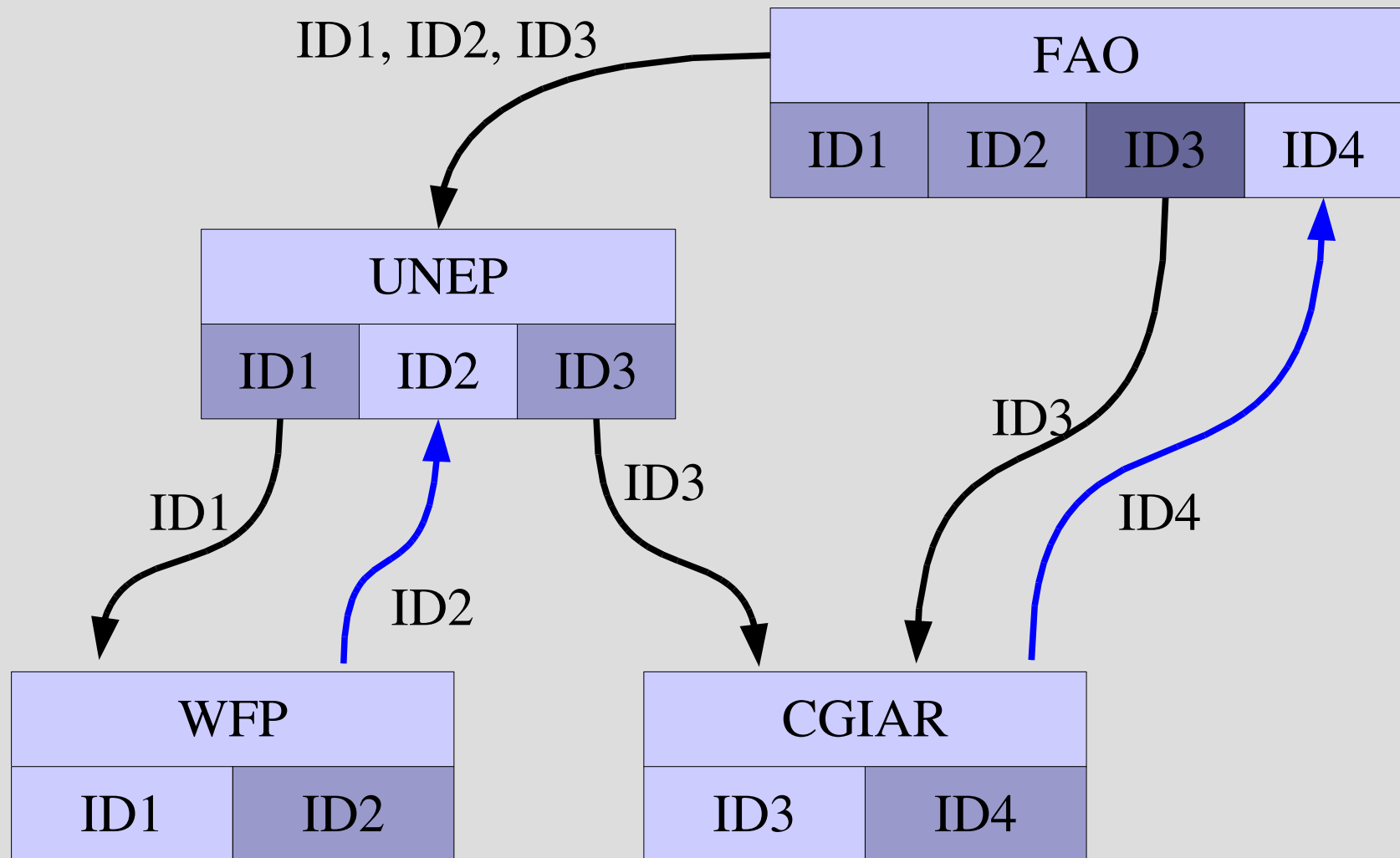
- Harvesting overview
- Protocol
- Configuration
- Example
- Configuring logos and thumbnails

# Harvesting Overview

# Harvesting Overview

- It is a way to share metadata between GeoNetwork nodes
- It uses the UUID for metadata synchronization
- It works with ISO 19115 metadata (will work with 19139)
- Harvested metadata are not editable
- Harvesting hierarchies are allowed

# An example of harvesting hierarchy



# Protocol

# Protocol

- Uses XML over HTTP to call remote services
  - An HTTP POST request is used to send XML data to the remote node and get the results back
- Protocol steps (each step is a service call):
  - Login on remote node (optional)
  - Search execution
  - Metadata retrieval based on change date and UUID
  - Logout (optional)

# Protocol

- Without authentication
  - All public metadata are returned (those that belong to the *internet* group)
- With authentication
  - Same as without authentication
  - Plus all metadata *visible* to the specified user

# Protocol

- The harvesting is incremental: after the first run, only changed metadata are harvested
- Categories are automatically assigned if they exist locally
- Group privileges are not imported and must be set manually

# Configuration

# Configuration

- Jeeves schedules
- Harvesting schedule
- Remote host and account
- Search parameters
- Privileges of harvested metadata
- Harvesting services

# Jeeves schedules

- A schedule is a periodically executed activity (30 secs, 1 hour, etc...)
- It is managed by Jeeves
- User specifies the java class to execute and the time period
- This information is provided by the GeoNetwork's configuration file (web/WEB-INF/config.xml)

# Jeeves schedules

Example of a configuration file with a schedule:

```
<geonet>
  ...
  <schedules package="org.fao.geonet">
    <schedule name="geonetwork sync"
      class=".schedules.MetadataSync" when="12 hours">
      ...
    </schedule>
  </schedules>
  ...
</geonet>
```

# Harvesting schedule

```
<schedule name="..." class="..." when="30 min">
  <param name="sites">
    <site name="geonetwork">
      ... <!-- configuration elements go here -->
    </site>
    ...
    <site name="geonetwork">
      ... <!-- configuration elements go here -->
    </site>
  </param>
</schedule>
```

# Remote host and account

- You have to specify host, port and servlet name
- If user and password are omitted the search is performed with the **Guest** account

```
<general>  
  <host>geonetwork.fao.org</host>  
  <port>8080</port>  
  <language>en</language>  
  <username>test</username>  
  <password>test</password>  
</general>
```

# Search parameters

- Search parameters can be taken from the browser
- All parameters are optional except for the **siteld**

Example:

```
<search>
  <siteld>mysite-id</siteld>
  <any>...</any>
  <paper>off</paper>
  <relation>equal</relation>
  ...
</search>
```

# Privileges of harvested metadata

- For each local group, a set of privileges can be specified

Example:

```
<group id="1">  
  <privilege>view</privilege>  
  <privilege>download</privilege>  
  <privilege>edit</privilege>  
  <privilege>dynamic</privilege>  
  ...  
</group>
```

# Harvesting services

- Usually you don't have to change them
- You can call these services for other purposes

```
<services>  
  <login>xml.user.login</login>  
  <logout>xml.user.logout</logout>  
  <search>xml.search</search>  
  <get>xml.metadata.get</get>  
</services>
```

# **A complete example**

# A complete example

```
<schedule name="sync" class=".schedules.MetadataSync" when="30 sec">  
  <param name="sites">  
    <site name="geonetwork">  
      <general>  
        <host>vam.wfp.org</host>  
        <port>80</port>  
        <language>en</language>  
        <username></username>  
        <password></password>  
      </general>  
      <services>  
        <login>xml.user.login</login>  
        <logout>xml.user.logout</logout>
```

# A complete example

```
<search>xml.search</search>
  <get>xml.metadata.get</get>
</services>
<search>
  <siteId>wfp</siteId>
</search>
<group id="1">
  <privilege>view</privilege>
  <privilege>download</privilege>
  <privilege>dynamic</privilege>
</group>
</site>
</param>
</schedule>
```

# **Configuring logos and thumbnails**

# Configuring logos and thumbnails

- Organization logos are stored in 'web/images/logos'



- Logos can be associated to harvested metadata and shown in the search results page
- Edit the 'web/xml/sources.xml' file to setup your logos

# Configuring logos and thumbnails

- Format of a **source** element:
  - **code** : this is the siteld of the local or remote node
  - **logo** : absolute web path of the logo image
  - **name** : name of the node. Used in the advanced search on the main page to fill the siteld drop down control.
  - **baseURL** : for harvested metadata, point to the basic GeoNetwork url

# Configuring logos and thumbnails

Example:

```
<source>  
  <code>fao</code>  
  <logo>/images/logos/fao.gif</logo>  
  <name>Food and Agriculture Organization</name>  
  <baseURL>http://www.fao.org/geonetwork</baseURL>  
</source>
```

That's all