Aggregating metadata from educational repositories the ODS experience

Edrene Conference, 6 May 2014, Athens
Kostas Vogias (GRNET), Enayat Rajabi (UAH)

vogias@admin.grnet.gr
Enayat.rajabi@uah.es
Harvesting 

Repositories

OAI PMH

ODS
Processing and importing

Process

Updater
## Statistics for 2\textsuperscript{nd} Harvesting Cycle

<table>
<thead>
<tr>
<th>Repository</th>
<th>Harvesting</th>
<th>Usable</th>
<th>Repository</th>
<th>Harvesting</th>
<th>Usable</th>
<th>Repository</th>
<th>Harvesting</th>
<th>Usable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dryades</td>
<td>251,047</td>
<td>250,591</td>
<td>SET for the resources Natural Europe</td>
<td>3,089</td>
<td>2,926</td>
<td>Photodentro/Educational Video repository</td>
<td>676</td>
<td>676</td>
</tr>
<tr>
<td>MACE</td>
<td>186,152</td>
<td>94,725</td>
<td>OpenScout</td>
<td>56,247</td>
<td>40,851</td>
<td>CARNet School Portal</td>
<td>513</td>
<td>513</td>
</tr>
<tr>
<td>Moodle for CARNet's Users</td>
<td>89,615</td>
<td>89,423</td>
<td>I2G Intergeo</td>
<td>2,847</td>
<td>2,846</td>
<td>Bulgarian National Educational Repository /Resursi</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Discover the COSMOS</td>
<td>81,520</td>
<td>81,380</td>
<td>SET for the pathways Natural Europe</td>
<td>125</td>
<td>123</td>
<td>Rural Observatory</td>
<td>409</td>
<td>399</td>
</tr>
<tr>
<td>LA FLOR</td>
<td>56,858</td>
<td>54,036</td>
<td>OER Commons (Green OER)</td>
<td>2,770</td>
<td>2,389</td>
<td>Digis</td>
<td>389</td>
<td>386</td>
</tr>
<tr>
<td>Znam.bg</td>
<td>60,986</td>
<td>60,986</td>
<td>Photodentro/School LOR</td>
<td>2,702</td>
<td>2,702</td>
<td>OSR</td>
<td>1977</td>
<td>343</td>
</tr>
<tr>
<td>Organic.Edunet</td>
<td>10,937</td>
<td>9,169</td>
<td>LaProf</td>
<td>752</td>
<td>752</td>
<td>SIVECO LRE Repository</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>Miksike's LeFo</td>
<td>3,798</td>
<td>3,797</td>
<td>Educanext (LMS.at, Bildungspool)</td>
<td>322</td>
<td>322</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total harvesting</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>814,323</strong></td>
<td></td>
<td><strong>Usable</strong></td>
<td><strong>699,927</strong></td>
<td></td>
</tr>
</tbody>
</table>
Workflow Representation

- Repository Registry
- XSL Registry
- Harvesting
- Metadata Analysis
- Transformation
- Identification
- Validation
- Filtering
- Lang Detection
- Link Checking
- Metadata Analysis
- Logging
- Log Visualization
Logging System

- Live automatic updates.
- Good for repository administrators and Data providers.
Metadata Aggregation Workflow

Aggregation Results Visualization
Brainstorming

* Distributed metadata indexing to Elastic Search cluster.
  * Scalability.
  * Replication.
  * Distributed backup.
  * Availability.
  * Search API over Repository.
* Data Providers registry content exposure using Elastic Search.