Improving usage statistics for RepositóriUM

Eloy Rodrigues [eloy@sdum.uminho.pt]
Ângelo Miranda [amiranda@reitoria.uminho.pt]
Ricardo Saraiva [rsaraiva@sdum.uminho.pt]
https://repositorium.sdum.uminho.pt
Summary

- Introduction
- Objectives, principles and architecture of the Statistics Add-on
- Objectives and Improvements of the Statistics v2.0
- Brief Demo
- Future Work and Concluding Remark
University of Minho

- Created in 1974
- Two main campus in two different towns (Braga and Guimarães)
- +13 500 undergraduate students
- +1 500 graduate students

- 1 116 FTE academic staff
- 11 Schools/Institutes
- 30 Research Centers
Building an I.R. was defined as a strategic objective in 2003.

The I.R. was included in the University proposal to the national Program E-U Campus Virtual, was approved and integrated E-UM (the University of Minho project).

RepositóriUM was publicly released on the 20th November 2003.
# Evolution of RepositóriUM

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N.º communities</strong></td>
<td>7</td>
<td>33</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td><strong>N.º documents</strong></td>
<td>626</td>
<td>3105</td>
<td>4990</td>
<td>5942</td>
</tr>
<tr>
<td><strong>N.º registered users</strong></td>
<td>717</td>
<td>2410</td>
<td>3390</td>
<td>3958</td>
</tr>
</tbody>
</table>
Evolution of visitors, visits and downloads
Aims of RepositóriUM statistics

- Promote RepositóriUM by showing its significant usage
- Promote author self-archiving/deposit in the IR by
  - Demonstrating the usage (access/downloads) of archived documents
  - Demonstrating the worldwide accessibility of archived documents
- Provide usage, content and administrative statistics to IR and community/collection administrators or coordinators
General principles

- Customizable statistics
- Customizable web interface
- Customizable access policies
- Data stored in database
- Access origin detection and processing
- Crawler exclusion
- Real time/Batch statistics
Impact of the Statistics Add-on

- Wide usage of RepositóriUM statistics by Minho University members (e.g. community managers and individual authors.)

- Several contacts from other interested DSpace users and some repositories using it (e.g. Universiteit Hasselt (Belgium), Brunel University Research Archive (UK) :
Problems and limitations of v1

- Increasing degradation of performance/response time as the usage of RepositóriUM and the period of the log information have been growing;

- Rigidity on the statistics definition and structure of the Web interface;

- Poor layout and data presentation (e.g. Inexistence of charts; reduced navigation and limited data presentation).
Objectives and improvements of v2

- Improvement of the performance and response time
  - Temporal data aggregations;
  - Dimensional aggregations (e.g. Country, Document type, etc.)

- Improvement of the presentation and navigation
  - Customizable menu navigation Structure (100% XML)
  - Customizable statistics layout (100% XSL style sheets)
  - Charts

- New statistics

- Better system documentation
  - Renewed Installation Guide
  - New User Guide
Overall architecture

1. DSpace logs
2. Log Processor
3. Data Model
4. Web Interface
Detailed architecture

- PostgreSQL
- Log Table
- Log processor
- Data Tables/Views
- Access origin detection and processing
- Spider/Crawler detector
- log4j
- XML stats definition
- XSL Stylesheets
Downloads e consultas - Períodos Fixos - Total
As estatísticas disponibilizadas referem-se a dados registados desde 1 de Janeiro de 2006.

Período: Total; Ano; Mês

Downloads e consultas

<table>
<thead>
<tr>
<th>Ano</th>
<th>Downloads</th>
<th>Consultas</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>840.16</td>
<td>775</td>
</tr>
<tr>
<td>2007</td>
<td>1615.08</td>
<td>1172</td>
</tr>
</tbody>
</table>

Downloads e consultas por ano

Downloads por país (top 10)

<table>
<thead>
<tr>
<th>País</th>
<th>Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1200</td>
</tr>
</tbody>
</table>

Consultas por país (top 10)

<table>
<thead>
<tr>
<th>País</th>
<th>Consultas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>900</td>
</tr>
</tbody>
</table>
Brief Demo
Future Work

- Multilanguage support
- Drill-down and drill-up capabilities
- Exporting (CVS, PDF)
- ...
Concluding remark

- DSpace needs good statistical tools/module;

- We want to contribute, with others (GSOC a etc), to a global, single, comprehensive and highly customizable solution for this need.
Grazie/Thank You!

https://repositorium.sdum.uminho.pt