Development of usage statistics for RepositóriUM
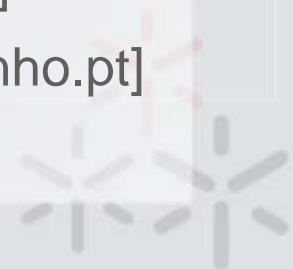
Eloy Rodrigues [eloy@sdum.uminho.pt]

Angelo Miranda [amiranda@sdum.uminho.pt]

http://repositorium.sdum.uminho.pt

Universidade do Minho Serviços de Documentação
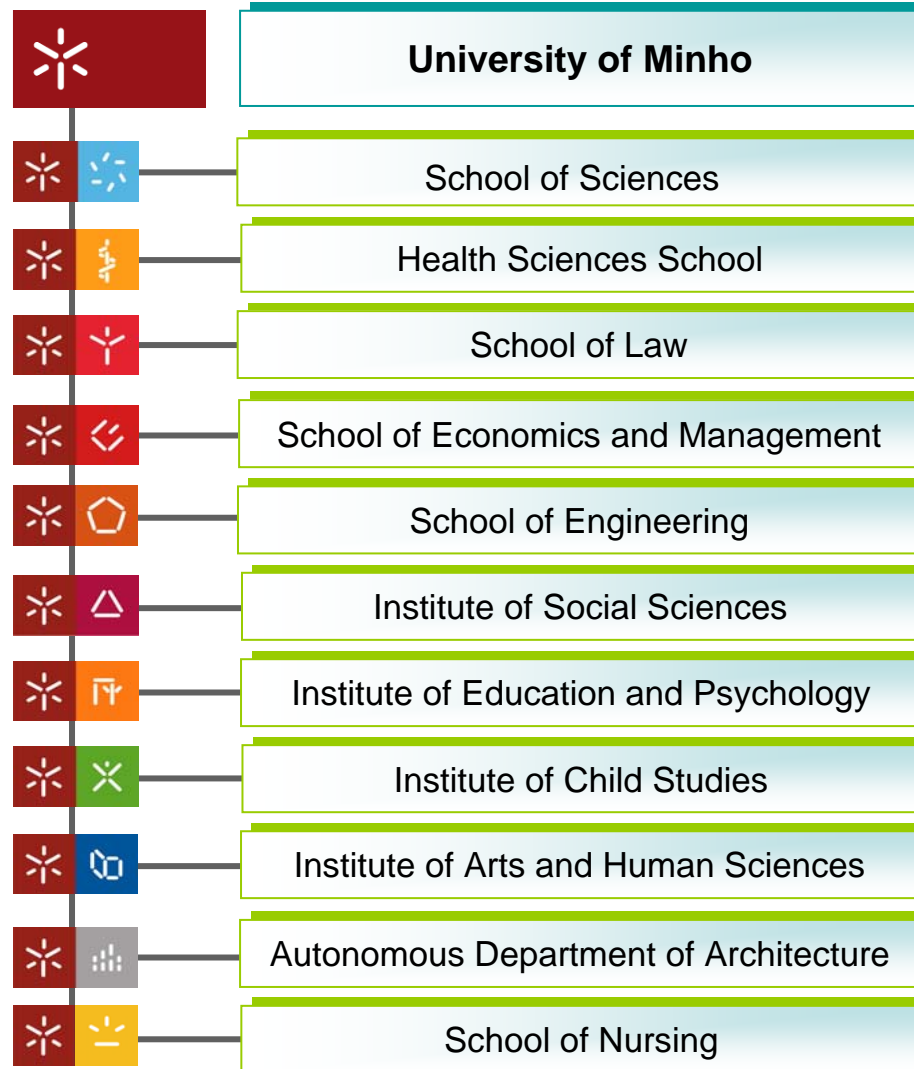
Summary

- Introduction

- Objectives and general principles

- Architecture

  - Log Processor

  - Data Model

  - Stats Processor

- Future work

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 (teachers and researchers)

- 11 Schools/Institutes

- 30 Research Centers

University of Minho

School of Sciences

Health Sciences School

School of Law

School of Economics and Management

School of Engineering

Institute of Social Sciences

Institute of Education and Psychology

Institute of Child Studies

Institute of Arts and Human Sciences

Autonomous Department of Architecture

School of Nursing

RepositóriUM

- 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)

- After a review of available systems, DSpace was chosen to implement the I.R.

- RepositóriUM was publicly released on the 20th November 2003.

Evolution of RepositóriUM

- The evolution of RepositóriUM in the 1st semester of 2004 was slower than expected.

- Strategy defined:

  - Communication plan and promotion of RepositóriuM inside and outside Minho University

  - Active participation in the international community related with Open Access, IR's and DSpace

  - Definition of institutional policy requiring self-archiving in the IR (Defined in December 2004, applied in 2005)

  - Development of value added services for authors and their communities

    - Statistics

    - Bibliographical lists, reports, etc.

Evolution of RepositóriUM

	December 2004	December 2005	April 2006
N.º communities	7	33	34*
N.º documents	626	3105	3852
N.º registered users	717	2410	2844

* 4 New communities in process of adhesion

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

- Based on ANU software

- Real time statistics

- Data stored in database

- Customizable queries

- Customizable web interface

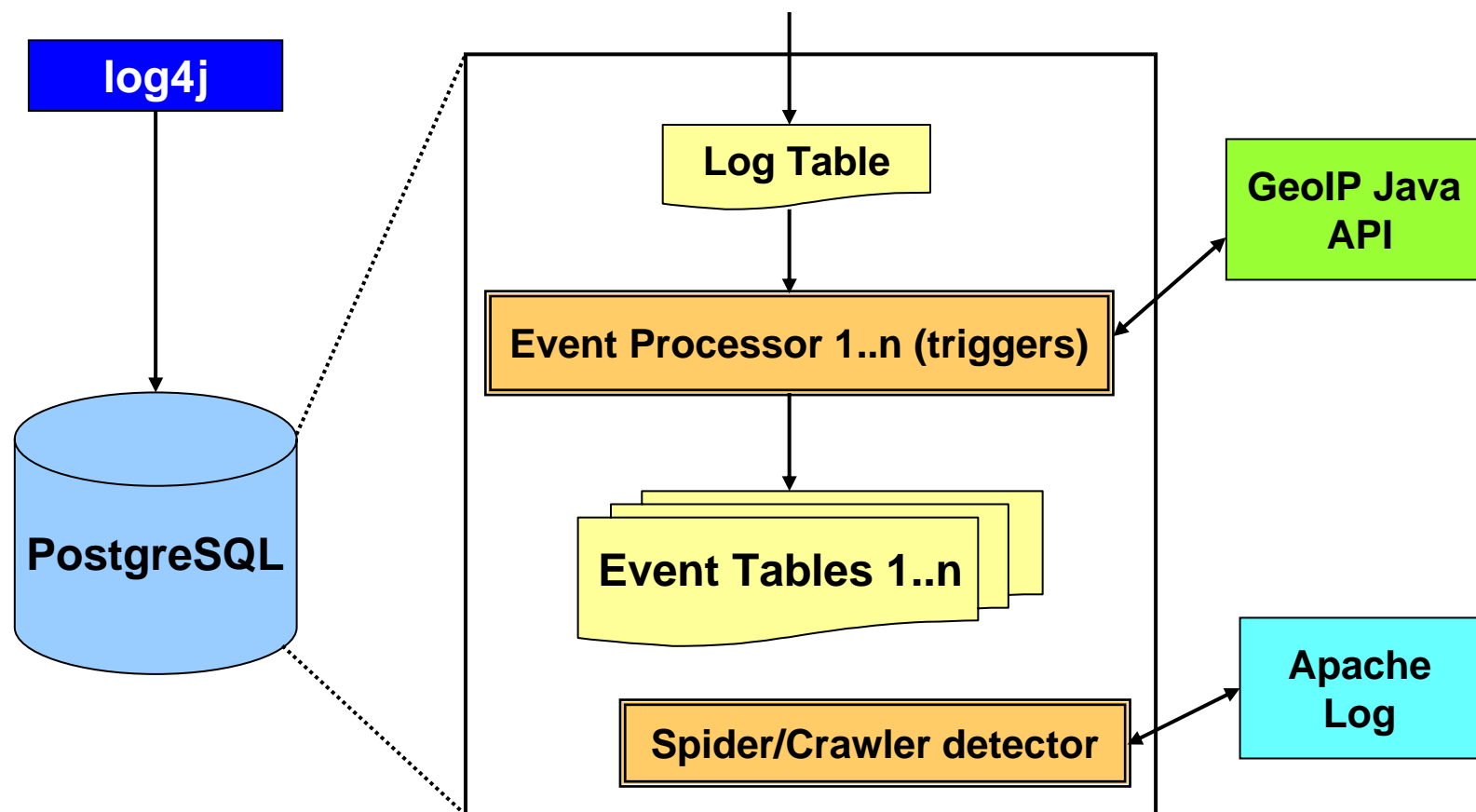- Customizable access policies

Requirements

- DSpace 1.3.x

- MaxMind GeoLite Country (free)

- MaxMind GeoIP Java API (free)
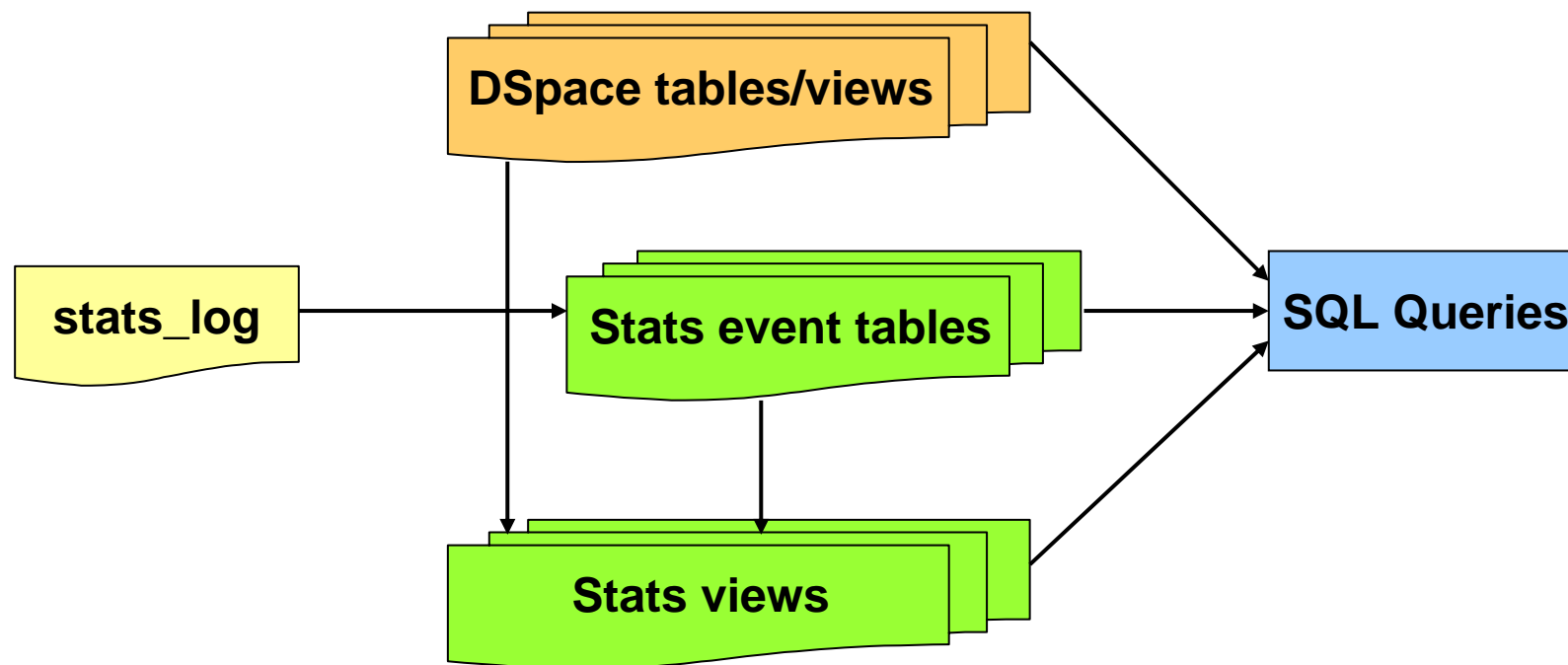
- Apache combined log format

Overall architecture

```
        DSpace logs
             |
             v
      +---------------+
      | Log Processor |
      +---------------+
             |
             v
      +---------------+
      |  Data Model   |
      +---------------+
             |
             v
      +---------------+
      | Stats Processor |
      +---------------+
             |
             v
```
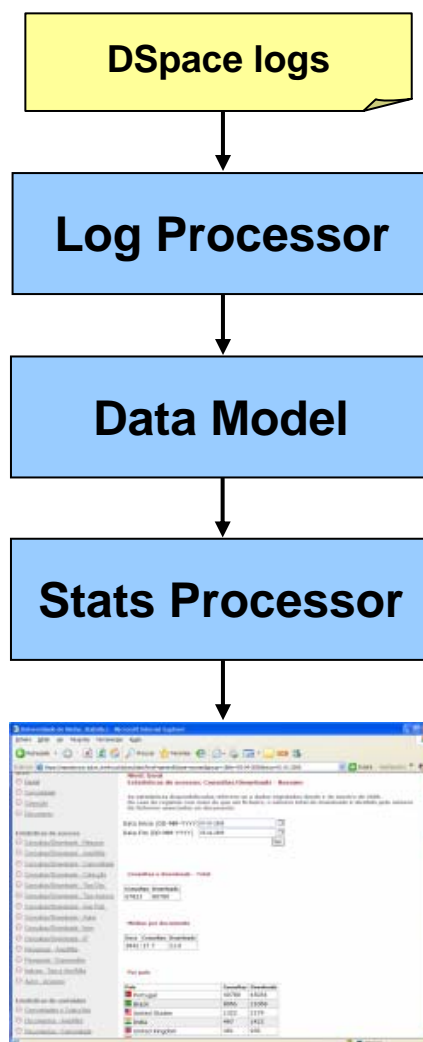
Log Processor

Log Processor

- Log events we are currently processing
  - view_item, view_bitstream
  - search, browse
  - start_workflow, advance_workflow, claim_task
  - login

- Other events can be processed

Data Model

Query model

- Organized by Type of Statistic
  - Access
  - Content
  - Administrative

- and aggregation level
  - Global
  - Community
  - Collection
  - Item

Query model

- Queries are configured in XML

- Query Groups

  - Definition of groups of queries based on type of statistic and aggregation level

- SQL Query

  - Individual sql query definition. Each query can be used in more than one group

- Model is used to build the navigational component of the web interface

Query model

- ## Groups

```
...
<group type="access" level="community" accessGroups="1" name="Views/Downloads">
    <querys name="view-down-community" display-type="html"/>
    <querys name="view-down-average-community" display-type="html"/>
</group>
....
```

- ## SQL Query

```
...
<query name="view-down-community" title="Consultas e downloads - Total">
  <option name="use-xsl-transform" type="html" stylesheet="resultset2table.xsl" render-to="table"/>
  <option name="use-xsl-transform" type="xml" stylesheet="identity.xsl" render-to="xml
document"/>
  <param src="communitylist" name="Community" id="object-id"/>
  <param name="Start Date (DD-MM-YYYY)" id="inicio"/>
  <param name="End Date (DD-MM-YYYY)" id="fim"/>
  <sql> select views as "Views" from .... </sql>
</query>
....
```

Stats Processor

Web Interface

Future Work

- **Improve navigational component**
  - Navigational side bar
  - Navigation between queries

- **Improve web look and feel**
  - Improve XSL stylesheets
  - Develop Chart XSL stylesheets
  - Develop XSL stylesheets (Excel, Pdf)

- **Extend log processor**
  - Develop additional event processors
  - Develop additional SQL queries

Credits

> Programming
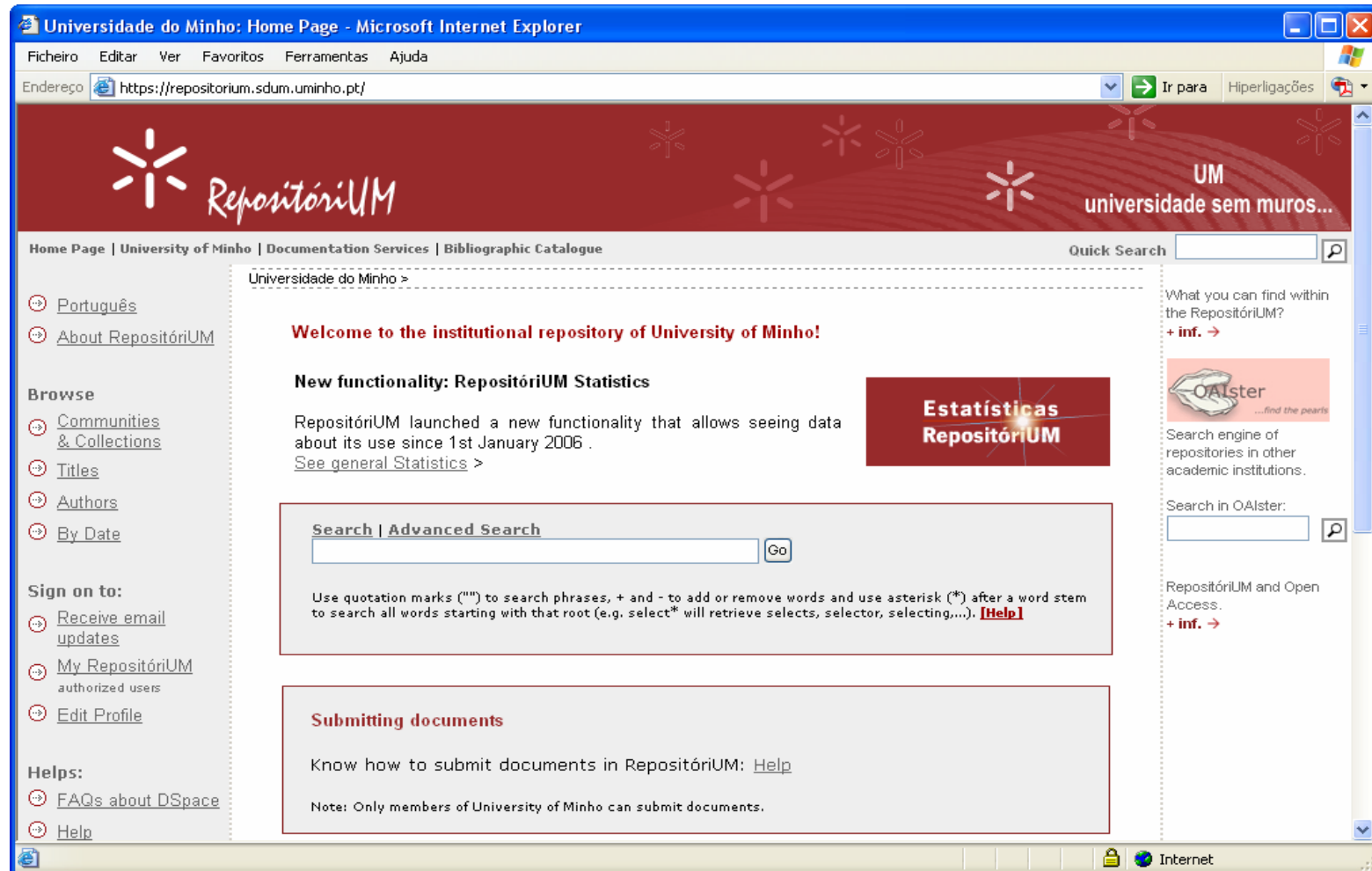
Arnaldo Dantas [arnaldo@sdum.uminho.pt]

> Interface Design

Ricardo Saraiva [rsaraiva@sdum.uminho.pt]
Daniela Castro   [dcastro@sdum.uminho.pt]

Thank you!

https://repositorium.sdum.uminho.pt