DRIVER Infrastructure

The Demo

“European Information Space: Infrastructures, Services, and Applications Workshop”

Paolo Manghi

ISTI-CNR

paolo.manghi@isti.cnr.it
DRIVER Infrastructure

• Targets VREs addressing “data integration”
  – Examples in Europe: DARE (NL), BASE (DE), PUMA (IT), DART Europe, etc.

• In order to
  – Overcome high maintenance costs, low degrees of scalability, and data sharing
**DRIVER Resource KITs**

- DRIVER middleware currently outsources two application paradigms
  - Data aggregation KIT
  - End-user functionality KIT
- Open to other application paradigms
DRIVER Data Aggregation KIT

• Harvesting tools
  – Easy “harvesting” of heterogeneous external data resources

• Aggregation tools
  – Data manipulation: e.g. enrichment, modification, transformations
  – Definition of Information Spaces: new data, of higher quality

• Data storage and access outsourcing
  – Automatic management of data storage and data access (sharing, replication, indexing, etc)
DRIVER End-User functionality KIT

• User Interfaces
  – Automatic adaptation to different Information Spaces features

• User profiling tools
  – Acquiring and exploiting user interest info

• User Community management
  – Enabling and easing the process of data selection activities

• User Recommendation mechanisms
**DRIVER European VRE**

- **European VRE for Open Access publications**
  - Harvests content from European Repositories
    - Now: 51 Repositories for a total of 250,000 Open Access documents
    - Soon: up to 73 Repositories by the end of the year
  - Maintains one European Information Space
    - **DRIVER Metadata Format (DMF)**

- Other VREs are to join DRIVER in the near future
  - Adding new content
  - Sharing existing content
**DRIVER infrastructure: the DEMO**

**Aggregation**: populating the DRIVER Information Space (DMF)

**Running queries** over the DRIVER Information Space (DMF)

**Harvesting**: enriching DRIVER with content from OAI-PMH Repositories

**User Functionality KIT**

**Repositories**
DEMO links

- http://search1.driver-repository.eu/webInterface/
- http://driver10.isti.cnr.it/
- http://129.70.12.42:8007/cgi-bin/ASmanager.pl
- Open Alpha release on DRIVER Information Space, final release by the end of the year
- Operates over the DRIVER Information Space and DMF documents
- It includes many functionalities, such as user registration and profiling, collections and communities, others will be introduced soon, such as browsing or user private collections
- From here we can search the whole information Space or focus on a subset by using collections
  - For example:
    - we run the query: “american election”
    - Select the second document
    - Show the DMF: data cleaning, repository info, country
    - Enables definition of collections not only content-based but location-based
    - Show the document splash page
    - Get back to the search page and leave it there (it will be used for the same query)
- Move to the administrative interface, from which we can manage all infrastructure resources
  - For example: monitoring of Nodes/Services and Repositories, thus external resources
    - New instances of the services can be added or removed and the infrastructure will keep the VRE’s operative by exploiting the resources currently available
  - We can also register a new Repository
    - New repositories are registered and put in “pending” mode, waiting for validation from an administrator
    - The validation of a repository to be harvested requires the allocation of storage space, of indexes, and the notification to the administrators that will have to configure the harvesting operation
  - For example we can validate the Repository British Library as we know it contains documents related with “american election” to see the immediate effect of the harvesting
- Open Aggregator in other Browser TAB
- These effects can be tracked down through the list of data structures created by the infrastructure
- Run again in the first tab the same query
Welcome to the DRIVER website, portal to the European research information.

It may therefore run relatively slow. DRIVER endeavours to rectify this in due course.

Use this site to access various DRIVER products and services, such as the DRIVER Support site. Here you will find information about the project itself and its relevance to researchers, institutional and repository managers, funding agencies and others, as well as information about building repositories and activities in European countries. The DRIVER website also gives you access to the DRIVER Guidelines and the DRIVER wiki. The DRIVER portal offers a number of features and characteristics that are ready to use and others that will become available soon. More specifically:

Features & characteristics

- Access to public documents in a consistent and harmonized way from over 70 repositories (click here to get a full list)
- Advanced searching capabilities
- Use of Collections and Communities allowing virtual views on the data
- Personalized services that allow registered users to systematically filter their searches based on their individual research interests
- Dynamic growth of the DRIVER content by real-time repository data harvesting
- Distributed and scalable infrastructure supporting
  - Dynamic deployment of DRIVER nodes and services
  - Integration of external services requiring minimal effort
  - Provision of existing OPAC-like services to other service providers

Coming soon

- Browsing and navigating capabilities
- User and community related recommendations and alerting services
- Repository managers tools for easy registration and data quality feedback

Search in Repositories
Query results
**DRIVER Metadata Format**

---

**Document Information**

<table>
<thead>
<tr>
<th>Creator(s):</th>
<th>Heng, Michael S.H.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This paper adopts an information systems perspective to study the recent American presidential election. The election may be seen as a decision making process to choose the leader to represent the wishes and interests of the people. The system supporting the process is an information system, with information technology inputs, outputs, users, human operators, procedures, assumptions, and goals. The information system is supposed to serve as an objective instrument to support the decision making process of political election. But its functioning is deeply embedded in the political life, with its history, institutions, procedures, norms and strengths and flaws. The output of the system is thus a good enough answer, rather than the most accurate answer. All these aspects feature prominently in the recent election. In addition, the impasse in Florida reveals several problems; the more important ones are (1) the controversy surrounding voting cards that not clearly punched (2) the use of outdated technology, (3) the tension between democratic ideal and legal and political expediency.</td>
</tr>
<tr>
<td>Publisher(s):</td>
<td>Vrije Universiteit Amsterdam</td>
</tr>
<tr>
<td>Contributors(s):</td>
<td>-</td>
</tr>
<tr>
<td>Repository:</td>
<td>DSpace at Vrije Universiteit Amsterdam</td>
</tr>
<tr>
<td>Name:</td>
<td>DSpace at Vrije Universiteit Amsterdam</td>
</tr>
<tr>
<td>Links:</td>
<td>Repository Info</td>
</tr>
<tr>
<td>Country:</td>
<td>NL</td>
</tr>
<tr>
<td>Institution:</td>
<td>-</td>
</tr>
<tr>
<td>Subject(s):</td>
<td>history; L98; political expediency; truth; meaning; American presidential election; interpretation; socially embedded; procedures;</td>
</tr>
<tr>
<td>Type(s):</td>
<td>Research paper</td>
</tr>
<tr>
<td>Language(s):</td>
<td>English</td>
</tr>
<tr>
<td>Doc Link:</td>
<td>View this document...</td>
</tr>
<tr>
<td>Model:</td>
<td>OA1</td>
</tr>
<tr>
<td>Metadata Format(s):</td>
<td>local_dc:</td>
</tr>
<tr>
<td>Publication Date(s):</td>
<td>2004-01-06</td>
</tr>
<tr>
<td>Collection Date(s):</td>
<td>2007-10-07T08:09:19Z</td>
</tr>
</tbody>
</table>
Virtual collections
Administrative interface
Repository management
Node and Service management
Repository registration and validation

DRIVER Repository Management

<table>
<thead>
<tr>
<th>Repository Name</th>
<th>Size</th>
<th>Country</th>
<th>Status</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archimer, Archive Institutionelle de l'Emmer</td>
<td>2159</td>
<td>FR</td>
<td>valid invalidate</td>
<td>delete</td>
</tr>
<tr>
<td>BieSOn - Bielefelder Server fuer Online-Publikationen (University of Bielefeld, GERMANY)</td>
<td>883</td>
<td>DE</td>
<td>valid invalidate</td>
<td>delete</td>
</tr>
<tr>
<td>Bristol Repository of Scholarly Eprints (ROSE)</td>
<td>852</td>
<td>UK</td>
<td>valid invalidate</td>
<td>delete</td>
</tr>
<tr>
<td>British Library Research Archive</td>
<td>41</td>
<td>UK</td>
<td>valid invalidate</td>
<td>delete</td>
</tr>
<tr>
<td>DSpace at Open Universiteit Nederland</td>
<td>112</td>
<td>NL</td>
<td>valid invalidate</td>
<td>delete</td>
</tr>
<tr>
<td>DSpace at Radboud Univ. Nijmegen</td>
<td>8913</td>
<td>NL</td>
<td>valid invalidate</td>
<td>delete</td>
</tr>
<tr>
<td>DSpace at UGent</td>
<td>2938</td>
<td>BE</td>
<td>valid invalidate</td>
<td>delete</td>
</tr>
</tbody>
</table>
Repository harvesting and aggregation

DRIVER Aggregator Manager - OAI Admin Panel

List of Repositories

Repository Information

Identify - ListSets - ListMetadataFormats

Harvesting

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Format unit</th>
<th>Status</th>
<th>Last Harvesting Date</th>
<th>Harvestingtype</th>
<th>Harvesting schedule</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>oai_dc</td>
<td>inactive</td>
<td>2007-10-20T17:25:46Z</td>
<td>REFRESH</td>
<td>WEEKLY</td>
</tr>
</tbody>
</table>

Aggregating

<table>
<thead>
<tr>
<th>Start aggregating</th>
<th>Test Mapping</th>
<th>Re-Feed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repository:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>British Library Research Archive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repository ID:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-8a2cc8b-b42c-c442f-a67c-a5a3fde13c34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>