Interoperability for digital repositories
Towards a Quality and Policy framework

OR2010 - Madrid, 6th July 2010

Dr Giuseppina Vullo and Perla Innocenti, HATII, University of Glasgow
Prof. Seamus Ross, Faculty of Information, University of Toronto
Today

- DL.org & the interoperability challenge
  - Addressing the interoperability challenge
    - DL.org Policy Working Group
    - DL.org Quality Working Group
  - First results
Project mission:
“mobilising Digital Library* designers, developers, end-users and researchers towards interoperability, best practices and modelling foundations for the enhanced development of next-generation Digital Libraries”

EU co-funded project, FP7 Coordination Action

DL.org Consortium:
- CNR-ISTI
- NKUA
- HATII
- TRUST-IT

http://www.dlorg.eu/
DL.org Strategic Alliances

- 14 EC-funded projects
- 8 National Initiatives
- 5 Coalitions & Think-Tanks
Interoperability definitions

• “The ability of two or more systems or components to exchange information and to use the information that has been exchanged” (IEEE, 1991)

• “the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires minimal knowledge of the unique characteristics of those units” (ISO/IEC 2382-2001)
Interoperability levels

- **Organisational interoperability**: refers to cooperation between and within organisations, business goals and process modelling. This is the most challenging level of interoperability, especially at a machine-readable and automation level.

- **Semantic interoperability**: refers to understanding the meaning of information.

- **Technical interoperability**: refers to interconnection, presentation and exchange of digital objects, accessibility and security issues.

*European Interoperability Framework for eGovernment services (IDABC, 2004)*
DL.org Working Groups

- Content Working Group
- User Working Group
- Functionality Working Group
- Policy Working Group
- Quality Working Group
- Architecture Working Group

https://workinggroups.wiki.dlorg.eu/index.php/Main_Page
A DL may operate within an organisation which defines over-arching policies (not necessarily specific to Digital Libraries) which affect interoperability *eg*:

- Subject community
- University
- A repositories’ network
Policy WG Participants

Kevin Ashley, ULCC

Perla Innocenti, HATII at UG

Seamus Ross, UoT

Hans Pfeifferberger, AWI

John Faundeen, USGS

Antonella De Robbio, UniPd

Mackenzie Smith, MIT Libraries

*Steve Knight, NLNZ

Testimonial

### Identified Policy Interoperability Issues

<table>
<thead>
<tr>
<th>Concept definition</th>
<th>Underpinning every digital library, there is an organisation governed by an organisational policy framework, that makes the digital library viable. The policy domain is a meta-domain, situated both outside the DL and any technologies used to deliver it, and within the DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interoperability level</td>
<td>Policy permeates the digital library from conceptualisation through to operation and needs to be so represented at these various levels <a href="https://workinggroups.wiki.dlorg.eu/index.php/Definition_of_Policy_and_Policy_Interoperability">https://workinggroups.wiki.dlorg.eu/index.php/Definition_of_Policy_and_Policy_Interoperability</a></td>
</tr>
</tbody>
</table>
| State of the art | Unexplored territory at global organisational (rather than only technical) level & interdisciplinary research  
Passing the baton from DL.org! |
<p>| Time dimension | Handling policy drift over time |</p>
<table>
<thead>
<tr>
<th>Concept definition</th>
<th>Policy Interoperability defined as <strong>Business Level Interoperability</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interoperability level</td>
<td><strong>At high (organisational) level, then instantiated at process level</strong> - whether those processes are being handled by human or machine</td>
</tr>
</tbody>
</table>
| Policy representation      | - **PLEDGE classification** (automated assessment of trustworthiness, iRODS rules, where policies are coded as functions, and TRAC)  
                              - **SHAMAN Assessment Framework** (TRAC criteria, DRAMBORA risk registry and mitigation strategies, iRODS rules)  |
| State of the Art and time dimension | - **Policy user scenarios**  
                                 - Evaluation of **current targeted DLs policies** |

**Towards a Policy Interoperability Framework**

[https://workinggroups.wiki.dlorg.eu/index.php/Policy_Interoperability_Approaches_Summary](https://workinggroups.wiki.dlorg.eu/index.php/Policy_Interoperability_Approaches_Summary)
Policy Scenario

• Digital Libraries and Archives in a consortium need to replicate (or backup) their content both for access continuity and as part of a preservation strategy, when that is a requirement of the library. Technically, there are many options for how to do it. These choices should be specified by the library's and archives policy and exchanged across consortium members.

• Additional challenges in real-life DLs include policy representation and classification, machine-encoding, policy drift over time.
Policy Interoperability Survey
first set of organisations

- ACM Digital Library
- California Digital Library
- (CDL) - Calisphere
- DANS
- DRIVER
- ELis
- Europeana
- ITHAKA: JSTOR, PORTICO
- Libor Liber
- NARA
- Nemertes
- National Science Digital Library
- Library (NSDL)
- Padua@Research
- UK Data Archive
- Univ. Chicago Digital Repository
- USGS Digital Library
The Quality WG Members

1st DL.org Workshop WG Testimonial

Scientific leader

Giuseppina Vullo
HATII, University of Glasgow

Seamus Ross
University of Toronto

Sarah Higgins
Digital Curation Centre (UK)

Dirk Roorda
Data Archiving and Networked Services (NL)

Sarantos Kapidakis
Ionian University

Nicola Ferro
University of Padua
WG Scientific Chair

Genevieve Clavel
Swiss National Library

Wolfram Horstmann
University of Bielefeld

Tefko Saracevic
Rutgers University

Coordinator

Start date: March 2009 (M4) - End date: July 2010 (M20)

Vullo, Innocenti, Ross – Interoperability for digital repositories – OR2010
Identified Quality Interoperability Issues

- Quality Interoperability, i.e. how different DLs can share a common Quality framework
- Data quality
- Quality Parameters
- DL Evaluation
- Towards a **Quality Core Model**

• **Our motivating scenario**: consider that representatives of two (or more) DLs have a round table to negotiate a service level agreement (SLA) defining their interoperability requirements and for this establish a quality threshold that each individual DL has to meet or exceed; “Quality” would provide transparent qualitative or quantitative parameters for defining the threshold

• **Our approach is practical**: Quality Interoperability Survey, Quality scenarios, best practices and Checklist
Identified Quality Interoperability Issues

The Quality Core Model

Quality Parameter
- Integrity
- Provenance
- Metadata Evaluation

Content Parameter
- Policy Consistency
- Policy Precision

Policy Parameter

Generic Parameter
- Interoperability Support
- Impact of Service
- Compliance to Standards

The Quality Core Model wikipage: https://workinggroups.wiki.dlorg.eu/index.php/The_Quality_Core_Model
Quality WG: some results

- Quality: dynamic, subjective, systems vs users
- Implement the **Quality Core Model** with the Quality Interoperability survey
- Quality Certifications and Guidelines: DINI, DRIVER, TRAC, DRAMBORA, Data Seal of Approval
- **Provenance** = the resource story = how to establish quality
- Identify and disseminate quality interoperability **best practices**, **Quality Interoperability Checklist**
## Quality Case studies

### Template

<table>
<thead>
<tr>
<th>Aspect</th>
<th>DINI Certificate</th>
<th>DRIVER Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit quality policy for protocol and metadata implementation</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Explicit policy for operations (personell, support etc.)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Personal quality check (questionaire, on-site review)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Intellectual quality check (remote)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Automatic self validation</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Organized through sustainable Organisation</td>
<td>DINI</td>
<td>COAR</td>
</tr>
<tr>
<td>Explicit branding when checked</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Translation in English, Spanish, Portuguese, Japanese</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Green and Gold</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Strictly full-text oriented</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Quality scenario
Policy Consistency

The DRIVER repository network has guidelines for content providers that define **how to expose fulltexts** with OAI-PMH. This is to make clear that DRIVER expects repositories to expose fulltexts rather than catalogue entries. At the same time DRIVER has registration policies for including repositories in the network. Consistency can be checked by whether or not the content policy is reflected in the registration policy. During registration DRIVER offers repositories a validator tool to check their compliance with the DRIVER-Guidelines.

However, for logical and technical reasons a **binary decision for or against compliance cannot be made and repositories (and therefore also DRIVER) may still offer records to users that do not lead to a fulltext**.

As a consequence, an **inconsistency** between **content policy** and **registration policy** could be stated. However, DRIVER applies a **quantitative compliance rate**.
Get involved ☺

New paths to interoperability
Best practices and modelling foundations for digital repositories

Content, Functionality, User, Policy, Quality and Architecture

Tomorrow morning
DL.org Birds of Feather 11.00-12.30
Room Reino Unido A
Thank you!

Policy WG:
p.innocenti@hatii.arts.gla.ac.uk

Quality WG:
g.vullo@hatii.arts.gla.ac.uk