The „many ways“ to interoperability

Outcomes and Challenges within DRIVER & OpenAIRE

Wolfram Horstmann
Universität Bielefeld
This Talk

• Discusses „Interoperability“ of DLs in the OA-domain (repositories)
• Exemplified by DRIVER- & OpenAIRE-initiatives
• „Already history“, story often told: but what can we actually learn WRT interoperability?
• Proposes a multi-faceted perspective on interoperability
• Suggests accommodating interoperability for autonomy in DLs by strict semantic focus
TERMS USED
“Interoperability“: Definitions

• “the ability of two or more systems or components to exchange information and to use the information that has been exchanged”

• “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”

cf. Vullo et al., Quality interoperability within digital libraries: the DL.org perspective
„Interoperability“: Levels

- **Technical** agreements cover formats, protocols, security systems, so that messages can be exchanged.

- **Semantic**: Content agreements cover the data and metadata, and include semantic agreements on the interpretation of the information.

- **Organisational** agreements cover the ground rules for access, preservation of collections and services, payments, authentication, etc.

Also adopted by *IDABC: European Interoperability Framework for pan-European eGovernment Services European Commission, Luxembourg (2004)*, now introducing *political* and *legal* level (here subsumed in organisational).

*cf. Vullo et al., Quality interoperability within digital libraries: the DL.org perspective*
Digital Library

- **DL.org Reference Model**
  - Digital Library
  - Digital Library System
  - Digital Library Management System
  - Users
    - DL End users
    - DL Designer
    - DL System Administrator
    - DL Application Developer
  - Content
  - Functionality
  - Policy
  - Architecture
  - Quality
Digital Libraries & Interoperability

- Digital libraries are complex systems, intrinsically interdisciplinary [and heterogeneous]. They involve collaboration support, digital preservation, digital rights management, distributed data management, hypertext, information retrieval, human-computer interaction, library automation, publishing.

- **Autonomy challenge** -- degrees of freedom in DLs needed for representing heterogeneous requirements; standards for interoperation
„Open Access“

• “By open access, we mean its immediate, free availability on the public internet, permitting any users to read, download, copy, distribute, [export], search or link to the [materials], crawl them for indexing, pass them as data to software or use them for any other lawful purpose.”

Squared brackets indicate changes of the original wording of the Budapest Declaration
„Repositories“

• „Vessels with tentacles and glands“
  – Here: for resources of scholarly communication
OA-INTEROPERABILITY IN DRIVER & OpenAIRE
DRIVER & OpenAIRE

• DRIVER
  – „Digital Repository Vision for European Research“
  – Construction plans for a distributed DL backbone
  – Generic infrastructure development & production

• OpenAIRE
  – „Open Access Infrastructure for Research in Europe“
  – Support for European Open Access Policy (pilot phase)
  – Multi-faceted application of DRIVER-results

• Ultimate goal: Multi-sited OA-DLS system

> Perfect example for OA interoperability challenges with complete walkthrough coverage of all levels
e.g. semantic interoperability

- DCMES as content agreements
- OAI-PMH „standard“ as implementation
  - Addressing technical (transfer) as well as semantic (content) interoperability – with known problems

E.g. Field Contents heterogeneity: `<dc:type>`

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>2597764</td>
</tr>
<tr>
<td>Text</td>
<td>2540080</td>
</tr>
<tr>
<td>TEXT</td>
<td>1128327</td>
</tr>
<tr>
<td>Dataset</td>
<td>580448</td>
</tr>
<tr>
<td>Image</td>
<td>522643</td>
</tr>
<tr>
<td>image</td>
<td>514585</td>
</tr>
<tr>
<td>Electronic Thesis or Dissertation</td>
<td></td>
</tr>
<tr>
<td>Tese ou Dissertacao Eletronica</td>
<td></td>
</tr>
<tr>
<td>article</td>
<td>345093</td>
</tr>
<tr>
<td>Article</td>
<td>337314</td>
</tr>
<tr>
<td>Artikel</td>
<td>310858</td>
</tr>
<tr>
<td>Other</td>
<td>219331</td>
</tr>
<tr>
<td>NULL</td>
<td>8082</td>
</tr>
<tr>
<td>Clippings&lt;br&gt;Text&lt;br&gt;</td>
<td>5004</td>
</tr>
<tr>
<td>This record contains content</td>
<td>1473</td>
</tr>
<tr>
<td>Dark Item</td>
<td>1145</td>
</tr>
<tr>
<td>Photographs; Mosaics;</td>
<td>1</td>
</tr>
<tr>
<td>1 photographic print mounted on cardboard: b&amp;w; 22 x 16 in.</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>80534</td>
</tr>
<tr>
<td>Section V Flow Control</td>
<td>1</td>
</tr>
<tr>
<td>Acrylic on board: 63 x 114 cm.</td>
<td>1</td>
</tr>
<tr>
<td>79508</td>
<td>1</td>
</tr>
<tr>
<td>Colored graphite on paper: 47 x 63 cm.</td>
<td>1</td>
</tr>
<tr>
<td>texts, vocabulary ('stories, monologues, word list')</td>
<td>1</td>
</tr>
</tbody>
</table>
e.g. semantic interoperability

• DCMES Interoperability solutions
  – Simplicity and degrees of freedom make it applicable and widespread, at least a common denominator
  – OAI-PMH readily packs in into a technical solution

• DCMES problems
  – Low specificity causes semantic variation
  – Integration with OAI-PMH prevents healing directed only to one aspect

• DRIVER / OpenAIRE always in the dilemma to address a critical mass and to sort out the problems
  – Alternative not building on OAI-PMH endangers to loose standards at all and leave materials unused
e.g. technical interoperability

- D-NET software for **multi-sited operation** at the level of **service interoperability**
e.g. technical interoperability

• D-NET: Interoperability Solutions
  – Division of labour in operation and dev. between partners (Athens, Bielefeld, Pisa, Warsaw)
  – Functional responsibilities: e.g. aggregation, UI etc.
  – Versatile construction of various DLs w. service reuse

• D-NET: Interoperability Problems
  – Transitions between technical and organisational interoperability introduce overheads
    • MoUs, SLAs, Specialization, local interests
  – Threshold for uptake to high due to the multiple problem domains to be solved for service interoperability

• DRIVER/OpenAIRE always in the dilemma between controlling sites and expanding services to partners
e.g. organisational interoperability

- Interoperability policy compliance requires organisational implementation and includes technical and semantic interoperability
  - DRIVER guidelines as OAI-PMH „interpretation“
    - „Become part of the network“ through validation
  - OpenAIRE guidelines as European OA-policy interpretation („SC39“)
  - OpenAIRE National Open Access Desks and DRIVER helpdesk as support structures est.
e.g. organisational interoperability

• Guidelines: interoperability solutions
  – Enhances adherence to standards
    • Also implementation in platforms (DLMS)
  – Provides glue between all interoperability levels

• Guidelines: interoperability problems
  – Introduce compliance enforcement
  – Cause extra work at all levels (repositories, support etc.)
PUTTING IT TOGETHER
Multi-Level Interoperability

• Interoperability levels not really separable
• Technical: Operation of distributed repositories and services
  – e.g. http, OAI-PMH (transfer), Web-Services, Platforms
• Semantic: Harmonization of distributed content
  – e.g. OAI-PMH (DCMES), Guidelines, Profiles/Rules
• Organisational: Guidelines compliance, policy implementation
  – Guidelines, Helpdesk/NOADs, Networking
  – Policy (EC-OA-policy), Legal (Copyright, SC39)
Observations WRT Levels

• Ambiguity of interoperability levels
  – OAI-PMH integrates technical and semantic level
  – Guidelines span all interoperability levels
  – Policy and legal level indeed can be added but also introduce more complexity

• Interpretation: „Direction“ of interoperability unclear
  – Reducibility: Organisational wouldn‘t exist without technical and technical not without semantic but not vice versa > semantic is the core interoperability level > „Library-Content here to stay“
  – Interoperability in heterogeneous systems benefits from organisation/policy

• Suggestion
  – Interoperability measures should address semantic level
Summary & Conclusion

• Interoperability is **multi-leveled** but rather a **network than** a layer-model is required
• **Simplicity** is the clue for uptake of **standards**
• **Semantics** are **core** DL interoperability **challenge**
• **Content** (semantic) and **functionality** (technical) to be **decoupled** as far as possible Look deeper into the DL.org-RM?
• Focus on semantic interoperability allow **DLs/repositories** the **autonomy** they need for meeting heterogeneous requirements
Tentative Outlook

• Semantic interoperability can be enhanced...
  – (not immediately by semantic web; probably still too powerful and versatile for DLs to encode content)
  – through rigid terminology standards applied on the local site (DL/repository)
    • E.g. authority files for Journal Titles, Names, Institutions, ...
  – through adaptive algorithms (unsupervised, dynamic, fault-tolerant...) that tame semantic variability on the central site (specialized service providers)

• Separation of „semanticists“ & technologists?
Thanks.