Driver Guidelines
Working on repository interoperability

Martin Slabbertje
Utrecht University (NL)
About me...

- Repository Manager @ Utrecht University Library since 2000
- Project Manager
- Expertise: Metadata, Digital Preservation, Repositories, Digitization
- Current Projects: Enhanced Publications, new website Digitized Special Collections
- Co-author of DRIVER Guidelines
DRIVER Guidelines?
Create New Ticket

Welcome to the DRIVER helpdesk ticketing system. To "open" a "ticket" please fill in this form specifying: your enquiry, the type of enquiry, your e-mail address and the email addresses of the people you want to include in the ticket conversation. Your request will be "assigned" to the members of the DRIVER support team who will contact you as soon as possible. "Opening" a ticket will initiate a conversation, i.e. an history of messages, between you and the support team members. Please remember that:

- Any reply from the team members will be saved in the ticket and delivered to you by email
- Any reply from you to the team members will have to be specified through this interface, NOT by email, and will be saved in the ticket; to ease this process, all emails from team members contain a link you should click on to reply to the message
- As a general rule, if it is more than one person interacting with the team members, please sign your replies; this will notably improve the quality of the conversation.

Your email:
DRIVER-project ended in December 2009, but:

- DRIVER-Search Portal will be maintained
- DRIVER-Guidelines are still valid (but need to be reviewed)
- DRIVER-Validator is still working
- DRIVER-Helpdesk is still active
- COAR: continue to work in WGs
- Speeding up services: OpenAIRE Plus?
Driver Safety Awareness
Outline DRIVER Guidelines

- Why comply with the DRIVER Guidelines?
- Overview DRIVER Guidelines
- Implementation in Utrecht (DSpace)
Why comply with the DRIVER-Guidelines?

- As a local Content Provider: connect to Service Providers (DRIVER Search Portal, BASE etc.)
- Base for a common exchange format (on a national and international scale)
- Easier to comply with new initiatives (Open-AIRE, COAR-Guidelines?)
But why do we need guidelines anyway?
Since 2009: 3 times as much contents and a small increase of participants
38 participating countries from all over the world → a lot of different policies
All 289 repositories are somehow “unique”
→ a lot of different practices
A look inside the repository metadata

- Variety of vocabularies
- No standards (internal metadata-formats)
- Variety of interpretations (semantics)
Example dc:type – theses

- text.thesis.masters - 28.671
- Text.Thesis.Doctoral - 46.594
- Doctoral thesis - 32.636
- Dissertation - 29.287
- Electronic Thesis or Dissertation - 366.783
- Tese ou Dissertacao Eletronica - 357.567
- Thesis - 95.626
- Study - 30.203
Example dc:type – article

- Article in monograph or in proceedings - 43.230
- JournalArticle - 53.606
- Article/Letter to editor - 58.978
- Article / Letter to editor - 73.058
- Article / Letter to the editor - 136.160
- article - 345.093
- Article - 337.314
- Artikel - 310.858
- 雑誌掲載論文 Journal Article - 12.774
- peer-reviewed article - 56.534
- journal article - 84.717
- Journal Article - 77.745
- a - 109.455
- Correspondence - 108.033
- Peer-reviewed Article - 49.885
- PeerReviewed - 146.348
- preprint - 19.1784
- NonPeerReviewed - 173.643
Examples

**Generic dc:type**
- text - 54.068
- texte - 15.051

**dc:date**
- 2006-11-07
- 2004-05-31T00:00:00Z
- 21st century

**dc:language**
- Proven x00E7 al Old to 1500 - 1
- pt, es. - 1
- spa;cat - 1
- eng; slv - 1
- noreng - 1
- [BLANK] - 1
- [language = Uvean, West] - 1
- [language = Madngele] – 1
Fine when your repository is the center of the universe...

It becomes a mess when repositories get aggregated!
What to do about all that ambiguity?

○ **Guidelines**
  ◦ **Talking about the issues and provide theoretical solutions**

○ **Validator**
  ◦ **Going from theory to implementation**

○ **Helpdesk**
  ◦ **Opening channels of communication with actual repository community**
DRIVER Guidelines – the basics

- Interoperable metadata (field content)
- Interoperable OAI-PHM (behaviour)
- Clear distinguishable Open Access set
Interoperable metadata

- Minimal DC, Recommended QDC
- Use of publication type vocabulary (dc:type)
- Use of version type vocabulary (dc:type)
- Use of persistent URL (dc:identifier)
- Optional: Use of MPEG-21 DIDL
  - Full-text harvesting
  - Representation of complex documents
  - Separation of Handle and Bitstreams
  - Multiple Bitstreams: level of Access
Local implementation in Utrecht (DSpace)

- **Interoperable Metadata**
  - Mapping internal metadata-fields into DRIVER metadata-fields
  - Mapping local vocabularies into DRIVER vocabularies
  - Configuration of DSpace-crosswalks

- **Interoperable OAI-PMH**
  - Implementation of OAI-PMH DRIVER-refinements
  - Define (Open Access) collections to be harvested and to be mapped to the DRIVER Set
  - Define Metadata Formats to be supported (nl-didl and oai-dc)
## Mapping Utrecht types – DRIVER types

<table>
<thead>
<tr>
<th>DSpace types</th>
<th>DRIVER types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>annotation</td>
</tr>
<tr>
<td>Article</td>
<td>article</td>
</tr>
<tr>
<td>Article in proceedings</td>
<td></td>
</tr>
<tr>
<td>Bachelor thesis</td>
<td>bachelorThesis</td>
</tr>
<tr>
<td>Book</td>
<td>book</td>
</tr>
<tr>
<td>Digitized book</td>
<td></td>
</tr>
<tr>
<td>Part of book or chapter of book</td>
<td>bookPart</td>
</tr>
<tr>
<td>Book review</td>
<td>bookReview</td>
</tr>
<tr>
<td>Conference report</td>
<td>conferencelItem</td>
</tr>
<tr>
<td>Pamphlet</td>
<td>conferencePaper</td>
</tr>
<tr>
<td>Conference object</td>
<td>conferenceObject</td>
</tr>
<tr>
<td>Contribution for newspaper or weekly magazine</td>
<td>contributionToPeriodical</td>
</tr>
<tr>
<td>Newspaper article</td>
<td></td>
</tr>
<tr>
<td>Doctoral thesis</td>
<td>doctoralThesis</td>
</tr>
<tr>
<td>Dissertation</td>
<td></td>
</tr>
<tr>
<td>Conference lecture</td>
<td>lecture</td>
</tr>
<tr>
<td>Educational material</td>
<td></td>
</tr>
<tr>
<td>Lecture</td>
<td></td>
</tr>
</tbody>
</table>
Mapping to DRIVER Set

<table>
<thead>
<tr>
<th>Collection</th>
<th>collid</th>
<th>Naar</th>
<th>webdir</th>
<th>xml naar...</th>
<th>darest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Select Collection (30)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Select collection</td>
<td>66</td>
<td>NEE</td>
<td></td>
<td></td>
<td>NEE</td>
</tr>
<tr>
<td><strong>Audio- en videocollectie UU (33)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diergeneeskunde AV</td>
<td>71</td>
<td>JA</td>
<td>av</td>
<td>av</td>
<td>NEE</td>
</tr>
<tr>
<td><strong>Bijzondere Collecties (1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gedigitaliseerde boeken</td>
<td>1</td>
<td>JA</td>
<td>bijzcoll</td>
<td>omega</td>
<td>NEE</td>
</tr>
<tr>
<td>Gedigitaliseerde kaarten</td>
<td>67</td>
<td>JA</td>
<td>kaart</td>
<td>omega</td>
<td>NEE</td>
</tr>
<tr>
<td>Gedigitaliseerde objecten (afgeschermd)</td>
<td>136</td>
<td>NEE</td>
<td></td>
<td></td>
<td>NEE</td>
</tr>
<tr>
<td><strong>Biologie (5)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biologie</td>
<td>5</td>
<td>JA</td>
<td>bio</td>
<td>omega</td>
<td>JA</td>
</tr>
<tr>
<td><strong>CERES (3)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERES</td>
<td>3</td>
<td>JA</td>
<td>CERES</td>
<td>omega</td>
<td>JA</td>
</tr>
<tr>
<td><strong>Cursusjaar 2010 - 2011 (48)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2011-05-13T12:07:05Z</responseDate>
  <request verb=ListMetadataFormats>
    http://dspace.library.uu.nl:8080/dspace-oai/request
  </request>
  <ListMetadataFormats>
    <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://purl.org/dc/elements/1.1/</schema>
      <metadataNamespace>http://purl.org/dc/elements/1.1/</metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>ubu_mods</metadataPrefix>
      <metadataNamespace>urn:mpeg:mpeg21:2002:02-DIDL-NS</metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>rd_ore</metadataPrefix>
      <metadataNamespace>urn:mpeg:mpeg21:2002:02-DIDL-NS</metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>srli_didl</metadataPrefix>
      <metadataNamespace>urn:mpeg:mpeg21:2002:02-DIDL-NS</metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
Lessons Learned

- Use of internal granular metadata
- DSpace has shortcomings when using MPEG-21 DIDL to express metadata on Bitstream-level
DRIVER guidelines 2.0
Guidelines for content providers - Exposing textual resources with OAI-PMH

[November 2008]

[Guidelines for Repository Managers and Administrators on how to expose digital scientific resources using OAI-PMH and Dublin Core Metadata, creating interoperability by homogenising the repository output.]
Validator

http://validator.driver.research-infrastructures.eu

DRIVER Score 84 out of 100

Rule: dc:creator (mandatory)

Rule: dc:date (mandatory)

Rule: dc:date (occurrence: 1)

Rule: dc:date (pattern: YYYY[-MM[-DD]])

Rule: dc:format (pattern: type/subtype)
Helpdesk

http://helpdesk.driver.research-infrastructures.eu/

Create New Ticket

Your email:
please enter your e-mail

Short summary:

Full description (you may use WikiFormatting here):

Type of enquiry: Other

e-mails of people to be included in the ticket conversation (Please separate with ;):

I have files to attach to this ticket

Are you human? Please complete this test before submitting:

http://helpdesk.driver.research-infrastructures.eu/ticket/109

Ticket #109 (new help) ...
... add operator specific instructions here ...

I can't see my repository in Driver, why?

Reported by: anna-kaisa.aalund@seamk.fi
Assigned to: somebody

Priority: minor
Assigned to: somebody

Component: Repository validation/Guidelines compliance
Version:

Keywords: CC:

Attachments

Attach File

Change History

29/04/09 21:42:36 changed by friedrich.summenn

Dear Anna-Kaisa,

please check ticket #70 and my last comments on that:

Dear Anna-Kaisa,

with the basicurl from ticket #63 it works now and the interface delivers 1103 metadata records.

Some first comments:

Using https for the basicurl communication is not OAI-PMH compatible but some more repositories do it that way.

Looking for the dc tags gives this list (the figures in brackets show the occurrence)
Guidelines, Validator and Helpdesk have helped increase repository interoperability

To be continued through COAR
Thank you

Martin Slabbertje
m.slabbertje@uu.nl

Many thanks to:
Maurice Vanderfeesten - SURF (NL)