



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?





Digital Repository Infrastructure Vision for European Research

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 in context

- 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



Harvesting

Metadata

DRIVER Guidelines

OAI-PMH

DRIVER Search Portal

Vocabulary Mapping

DRIVER Helpdesk

DRIVER Validator

DC



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?



© 2006 www.ArtStudioSeven.com

combscartoons@yahoo.com

DRIVER Search Portal



Digital Repository Infrastructure Vision for European Research

MyDriver

email password

[Sign in](#)

[SEARCH THE REPOSITORIES](#)

[LEARN ABOUT DRIVER](#)

[FIND SUPPORT](#)

[REGISTER YOUR REPOSITORY](#)

[Home](#) | [Search](#) | [Browse](#) | [Communities](#) | [Collections](#)

[Repositories](#)

Search for more than
3.340.000 documents
in 294 repositories
from 38 countries

[Search](#)

[Advanced Search...](#)

News and Announcements

- ❖ **25 feb 2011: New repository** - EconStor has joined DRIVER with 28,172 publications
- ❖ **25 feb 2011: New repository** - peDOCS (FachPortal.Paedagogik.de) has just joined DRIVER with 2,182 publications
- ❖ **25 feb 2011: New repository** - UnissResearch, the Institutional Repository of the University of Sassari, has joined DRIVER with 5203 publications

Communities

[Open Access Community](#)
[UNAM - EAE](#)

[View all](#) ⇨

[Create your communities...](#)

Collections

[European Countries](#)
[Specialized Communities](#)
[Subject classification](#)

[View all](#) ⇨

[Create your collections...](#)

Are you a repository manager?

Is your repository registered with DRIVER? If not, visit our dedicated site "[Register your Repository](#)" where you may register your repository with a few clicks.

Can we help your organization?

The DRIVER software ([D-NET v 1.0](#)) may be used to set up similar portals to harvest and aggregate content from various sources, provide search, browsing and personalized services using customized user interfaces.

Check out two example applications: the [Belgium DRIVER portal](#) and the [Spanish publications portal Reselecta](#)



DRIVER Search Portal

Contents: 20 October 2009
254 repositories – 31 countries
1,116,755 documents

Contents: 9 May 2011
289 repositories – 38 countries
3,360,000 documents

Since 2009: 3 times as much contents
and a small increase of participants



DRIVER Search Portal

38 participating countries from all over the world → a lot of **different policies**



DRIVER Search Portal

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
- 雜誌掲載論文 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

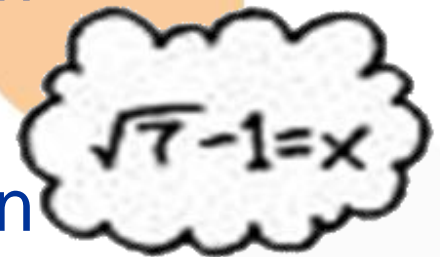
- ⊙ Talking about the issues and provide theoretical solutions

○ 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

DSpace types	DRIVER types
Annotation	annotation
Article Article in proceedings	article
Bachelor thesis	bachelorThesis
Book Digitized book	book
Part of book or chapter of book	bookPart
Book review	bookReview
Conference report	conferenceItem
Pamphlet	conferencePaper
Conference object	conferenceObject
Contribution for newspaper or weekly magazine Newspaper article	contributionToPeriodical
Doctoral thesis Dissertation	doctoralThesis
Conference lecture Educational material Lecture	lecture

Mapping to DRIVER Set

DSpace commur

	collid	Naar Ω	webdir	xml naar...	daset
1 Select Collection (30)					
1 Select collection	66	NEE			NEE
Audio- en videocollectie UU (33)					
Diergeneeskunde AV	71	JA	av	av	NEE
Bijzondere Collecties (1)					
Gedigitaliseerde boeken	1	JA	bijzcoll	omega	NEE
Gedigitaliseerde kaarten	67	JA	kaart	omega	NEE
Gedigitaliseerde objecten (afgeschermd)	136	NEE			NEE
Biologie (5)					
Biologie	5	JA	bio	omega	JA
CERES (3)					
CERES	3	JA	CERES	omega	JA
Cursusjaar 2010 - 2011 (4R)					

Supported Metadata Formats

```
- <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>
      - <schema>
        http://purl.lanl.gov/STB-RL/schemas/2003-12/DIDL.xsd
      </schema>
      <metadataNamespace>urn:mpeg:mpeg21:2002:02-DIDL-NS</metadataNamespace>
    </metadataFormat>
    - <metadataFormat>
      <metadataPrefix>nl_ore</metadataPrefix>
      - <schema>
        http://purl.lanl.gov/STB-RL/schemas/2003-12/DIDL.xsd
      </schema>
      <metadataNamespace>urn:mpeg:mpeg21:2002:02-DIDL-NS</metadataNamespace>
    </metadataFormat>
    - <metadataFormat>
      <metadataPrefix>nl_didl</metadataPrefix>
      - <schema>
        http://purl.lanl.gov/STB-RL/schemas/2003-12/DIDL.xsd
      </schema>
      <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

Validator

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

driver

Home | Registration | Validation | Studies | Statistics

Validation > Content report

Content validation report summary

Repository Information

<http://arno.unimaas.nl/oai/dare.cgi>

Validated set(s)

withfulltext:yes

DRIVER Score 84 out of 100

[Find out more about the DRIVER score...](#)

More info

Validation started on Thu Oct 15 20:54:44 CEST
2009

Elapsed Time 0 minutes and 8 seconds.

Checked 20 records (out of 21)

Current status - **Completed**

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/



Digital Repository Infrastructure Vision for European Research






[Login](#) | [Settings](#) | [About](#)

[New Ticket](#)

Create New Ticket

Your email:

Short summary:



Full description (you may use [WikiFormatting](#) here):
B **I** **A**     

Type of enquiry:

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:



stop spam.
read books.

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



Digital Repository Infrastructure Vision for European Research

logged in as [maurice.vanderfeesten](#) | [Logout](#) | [Settings](#) | [Help/Guide](#) | [About Trac](#)

[Wiki](#) | [View Tickets](#) | [New Ticket](#) | [Timeline](#) | [Roadmap](#) | [Search](#) | [Admin](#) | [Browse Source](#)

Ticket #109 (new help)

... add operator specific instructions here ...

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

Opened 5 months ago
Last modified 5 months ago

Reported by:	anna-kaisa.sjolund@seamk.fi	Assigned to:	somebody
Priority:	minor	Milestone:	
Component:	Repository validation/Guidelines compliance	Version:	
Keywords:		Cc:	

Attachments

Change History

29/04/09 21:42:36 changed by [friedrich.summann](#)

Dear Anna-Kaisa,
please check ticket [#70](#) and my last comments on that:

Dear Anna-Kaisa,
with the basicurl from ticket [#83](#) 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)

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



Powered by Trac 0.10.4
By Edgewall Software.



Visit the Trac open source project at
<http://trac.edgewall.org/>



**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)

