

## Ghent University Library

---

### New Developments in Digital Repository Infrastructure

October 29, 2010

Inge Van Nieuwerburgh



- Scholarly communication
- Dissemination of research results
- Prestige

Open Access

# DRIVER motivation

---

- Scholarly communication changes towards distributed provision of text, data and services
- Repositories are thought as a saviour in this development building such a distributed system
- An infrastructure supporting distributed repositories and services is needed

# Distributed provision

- People and institutions, including cultural heritage institutions, like to keep (control over) their own data
- Given this, building the global infrastructure of publications and data in a bottom-up fashion is the only approach that makes sense.
- Let things "live where they are born" and only add whatever is needed to organize the distributed aggregation of those results.

# Repositories

- Dissemination, not archiving
- Institutional / subject oriented
- Text/audiovisual/research data/...
- Open / closed

# Building repository networks

- Not primarily a technical challenge
    - free software or commercial hosting
    - (relatively) simple set-up
  - Not primarily a copyright problem
  - Not primarily author-persuasion about Open Access
- Challenges are in effective support for cultural, policy and procedural change management within institutions and research communities

## Top-down support - Institutional

- Shared vision with stakeholders
- Encouragement
  - Statements, policies, funding
- Embedding in research process and work-flows
  - Prestige measures to match
- Representation to powerful lobbies
- Ensure legal framework is supportive
- Support for centralised support services
- Recognition of value of bottom-up work

## Top-down support - Funding Bodies

- Recognition of value of Open Access to mission of funders
- Policies/ mandates to ensure Open Access and/or deposition
- Recognition/reward of compliance from authors
  - and sanctions for non-compliance from authors
  - work with repository managers
- Promotion of open access work to their stakeholders (government, general public, researchers, institutions, learned societies)

## Side-to-side support

- Networking amongst peers
  - email, events, wikis, blogs
- Professional training
  - advocacy, technical issues, legal issues
- Share best practice, standards
- Self-help - create:
  - mentoring arrangements
  - peer-networks
  - professional support groups

# Bottom-up support - Repository Managers

- Establish repositories
- Create effective policies for/about repository use
- Advocacy to researchers and authors - and library staff
- Tackle practical problems
- Identifying work-flows and structures within institutions to support Open Access deposit
- Act as institutional focus to drive repository agenda

# Repositories can solve access problem

- „Everybody can be a publisher“
  - Common description standards
    - e.g. Dublin Core Metadata Initiative
    - Many subject-specific standards
  - Common transfer protocols
    - e.g. OAI-PMH, but also FTP, XML-RPC, WS, etc.
  - Searchability is possible!

- Still: many data are lost to re-use/remix
  - Closed: too sensible, weakly described, unimportant, ....
  - Missing service frameworks / infrastructures

Problems: Data and service interoperability  
Solution: „Infrastructure“

# Infrastructure

# Needs

- Standards
- Protocols
- Guidelines
- Unique identifiers (PURL, handle, URI, ...)
- Cooperation
  - Improve full-text access
  - Standardize metadata quality
  - Create a reliable infrastructure for permanent identification, resolution, traceability and storage
  - Resolve semantic and classification issues

# Open!

- Open Access to scientific publications
- Open Access to research data
- Open Access to public domain cultural heritage

# Interoperability

- Create a network of repositories:
  - a common vision, guidelines, etc...: a grid of interoperable repo's

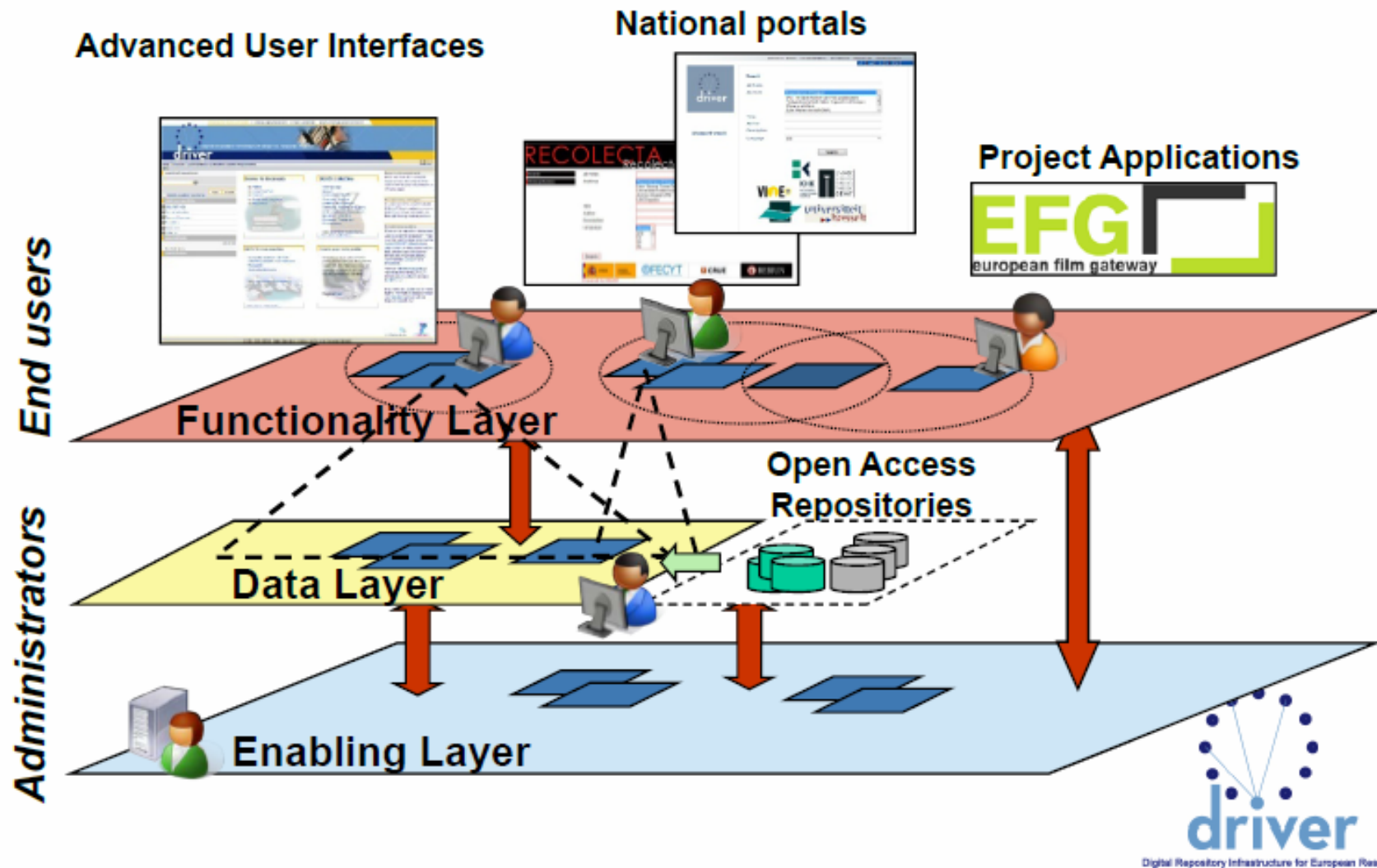
# DRIVER Objectives

- Organisational structure for repositories
  - ⊙ e.g. the „Confederation“
- Improving quality and standards in local rep.
  - ⊙ e.g. validation procedures
- Building a distributed runtime system
  - ⊙ e.g. service and data sharing
- Target Groups
  - ⊙ Repository Managers
  - ⊙ Service Providers
  - ⊙ Information System Executives

# The DRIVER approach is incremental

- Start with publications
  - ⊙ Existing distributed system, somehow connected
  - ⊙ Considerable homogeneity and formats: OAI-PMH
- Extend geographical coverage
  - ⊙ From 5 countries, to 10, to 27, to ???
- Extend towards other contents
  - ⊙ From publications to enhanced publications, i.e. representations of „texts + data“
- Learn about subject specificity
  - ⊙ Data bring in disciplinary requirements

# The DRIVER Infrastructure



# Follow up

- COAR (confederation of open access repositories)  
<http://coar-repositories.org/>
- Different portals based on D-NET:  
e.g. Recolecta: Spanish digital repository network
- OpenAIRE: making use of the infrastructure and network



# OpenAIRE

Open Access Infrastructure for Research in  
Europe

<http://www.openaire.eu>



# Main Goals

- deliver “an electronic infrastructure supporting mechanisms for the identification, access, and monitoring of FP7 and ERC funded articles”  
**helpdesk**
- additionally, offer “a specialised repository that can be stored neither in institutional nor in subject-based/thematic repositories”.  
**Orphan repository**
- All deposited articles will be accessible worldwide through a new portal to the products of EU-funded research, built as part of this project.  
**OpenAIRE portal**

# Heritage institutions

# Content

- Managed locally
- Expertise!
- Policies:
  - Openness (with open licences)
  - Roles different stakeholders
  - Support
  - Aggregation

## Open data

⇒ linked data: make links, link data with related data

⇒ Enrich content

⇒ OPEN the data

Download & API

search.ugent.be/meercat/x/api\_nl.html?language=en

Web Slice-galerie

Andere bladwijzers

inlichtingen | aleph | mijn bibliotheekgegevens | mobile | download&api | contact | chat | en | nl

Home Boeken & meer Tijdschriften E-Artikels Databanken Afbeeldingen

## Downloads | API's

### Plaats een Meercat zoekscherm op je eigen webpagina.

#### Beschrijving

Kopieer deze HTML code in je website of blog om het onderstaande Meercat zoekscherm te activeren:

Boeken & meer  Tijdschriften  Databanken  Afbeeldingen

[OPEN SERVICE](#)

#### Service Details

Kopieer en plak de onderstaande code in je website:

```
<form action="http://search.ugent.be/meercat/x/all">
  <div id="msearch">
    <input type="text" name="q" value="" size="40"/> <input type="submit"
value="Zoek">
  </div>
  <div id="mcheck" style="font-size: 70%;">
    <input type="radio" name="type" value="books" checked
onChange="this.form.action='http://search.ugent.be/meercat/x/all';"> Boeken &
meer
    <input type="radio" name="type" value="journals"
onChange="this.form.action='http://search.ugent.be/meercat/x/journals';">
Tijdschriften
    <input type="radio" name="type" value="databases"
onChange="this.form.action='http://search.ugent.be/meercat/x/databases';">
Databanken
  </div>
</form>
```

#### Gerelateerde open data collecties

- Ghent University Academic Bibliography & Institutional Repository

Download & API

search.ugent.be/meercat/x/api\_nl.html?language=en

Web Slice-galerie

Andere bladwijzers

---

## Dagelijkse/Wekelijkse exports

### Beschrijving

Hieronder staan enkele links naar datasets die door de Meercat zoekmachine worden geïndexeerd. De meeste files zijn in het Aleph sequentiële MARC formaat.

[OPEN DATA](#)

### Export details

Omschrijving	Type	Link	Frequentie
UGent catalogus	GZIP-TXT; MARC; Aleph sequential	<a href="#">Link ~ 50MB</a>	Wekelijks
UGent tijdschriftencatalogus (print)	GZIP-TXT; MARC; Aleph sequential	<a href="#">Link ~ 20MB</a>	Wekelijks
UGent steekkaartencatalogus	GZIP-TXT; MARC; Aleph sequential	<a href="#">Link ~ 350MB</a>	Wekelijks
UGent brievencatalogus	GZIP-TXT; MARC; Aleph sequential	<a href="#">Link ~ 4 MB</a>	Wekelijks
UGent Google Books catalogus	GZIP-TXT; MARC; Aleph sequential	<a href="#">Link ~ 50 MB</a>	Wekelijks

---

## Record exports

### Beschrijving

Verschillende Meercat RSS-feeds geven informatie over de nieuwe aanwinsten van faculteiten, vakgroepen en bibliotheken.

[OPEN DATA](#)

Gegeven een faculteit-id: **lw**

Formaat	Link
RSS	<a href="http://search.ugent.be/rss/lw.rss">http://search.ugent.be/rss/lw.rss</a>

Gegeven een vakgroep-id: **lw17**


Formaat	Link

Ghent University Academ... x

biblio.ugent.be/input/home?func=developer

Web Slice-galerie

Andere bladwijzers

 Ghent University Academic Bibliography & Institutional Repository

Home Contact FAQ Open Access Demo's Download|API

login

As of July 1st 2010, only records submitted **with** full text will be accepted in the academic bibliography. [more info](#)

## Downloads | API's

### Daily/Weekly Exports

**Description**  
For easy reporting and statistics we provide daily and weekly dumps of the complete datasets of the Ghent University Academic Bibliography which can easily be processed which office tools such as Excel, Access and various RSS readers.

[OPEN DATA](#)

Description	Type	Link	Frequency
Complete export I	Tab-delimited; CSV; TXT	<a href="#">Link</a> ~50MB	Daily
Complete export I	Tab-delimited; CSV; ZIP-ed	<a href="#">Link</a> ~40MB	Daily
Complete export II (all fields)	Tab-delimited; CSV; TXT	<a href="#">Link</a> ~90MB	Daily
Complete export II (all fields)	Tab-delimited; CSV; ZIP-ed	<a href="#">Link</a> ~26MB	Daily
Complete export III (MARC-ish)	TXT; MARC; Aleph sequential	<a href="#">Link</a> ~23MB	Weekly
Daily additions	RSS	<a href="#">Link</a>	Daily
Sitemap of all publications	Sitemap	<a href="#">Link</a>	Weekly
Twitter channel of all open access additions	Twitter	<a href="#">Link</a>	Hourly

### Related open data collections


- [Ghent University Library](#)


Groups of Data Packages ... x

ckan.net/group/

Web Slice-galerie

Andere bladwijzers

Login with  OpenID



comprehensive knowledge archive network

Home Data Packages Tags **Groups** Authorization Groups Revision History

Home Create a new group

CKAN.net

## Groups

There are 20 groups.

Title	Number of packages	Description
<a href="#">Open Linguistic Resources</a>	2	
<a href="#">Economics</a>	10	Packages related to economics.
<a href="#">Dictionaries</a>	21	A group for dictionaries, thesauri and suchlike.
<a href="#">Art</a>	5	Art and the artworld. For data regarding artworks, collections, museums, art...
<a href="#">Public Domain</a>	20	A group for collections of public domain works - including images, music,...
<a href="#">MKEdata.net: Milwaukee government data</a>	1	The mkedata.net group of packages.
<a href="#">Chemistry</a>	2	Open data packages in chemistry.
<a href="#">Open Geospatial Data</a>	13	A group for open geospatial data with an emphasis on use in teaching and

**Groups section**

Whilst tags are great at collecting packages together, there are occasions when you want to restrict users from editing a collection. A **group** can be set-up to specify which users have permission to add or remove packages from it.

[Create a new group](#)

## Summary e-infrastructure

- Distributed approach: repositories
- Repository challenge = support by different stakeholders
- Interoperability is needed
- Open data

## Links

- DRIVER: <http://www.driver-community.eu>
- OpenAIRE: <http://www.openaire.eu>
- Linked data (Tim Berners-Lee) <http://www.w3.org/DesignIssues/LinkedData.html>
- CKAN: <http://ckan.net>

# Contact

ngé Van Nieuwerburgh

gent University Library

[ngé.VanNieuwerburgh@UGent.be](mailto:ngé.VanNieuwerburgh@UGent.be)

<http://lib.ugent.be>