

## What can be done, right now!

Andrea Bollini, 4Science (CTO)

<http://www.4science.it/en/>



# Who I am?



...a proud member of the COAR Next Generation Repository Working Group

...the deputy leader of the euroCRIS CERIF & Architecture Task Group

...an active developer in the open scholarly field since 2004

...a strong advocate of the DSpace platform, committer and Lead of the REST API sub-team for DSpace 7



# Agenda: how to implement that?



The mission of a repository is to **manage and provide access to** the valuable and **diverse intellectual output** of the community it serves.

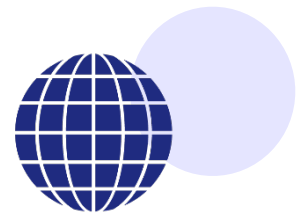
Repositories are **nodes in a larger network, contributing their collective contents** to a global knowledge commons on top of which value added services can be built.

Repositories provide access to published articles as well as to a **broad range of artifacts beyond traditional publications** such as datasets, pre-prints, working papers, images, software, and so on.

We also aim to create a global brand for repositories that establishes repositories as a **central place for the daily research and dissemination activities** of researchers.



Provide access to digital content  
doesn't mean to allow the  
download of the bits sequence...



Datasets need to be usable:  
preview, sampling, visualization,  
remote computation & more



# Dataset: good examples

CKAN: Preview, sampling, visualization and open webservices for tabular data

<https://ckan.org/>

Dataverse: allow exploration of tabular dataset and Geospatial files

<http://dataverse.org/>

DSpace-CRIS & DSpace-CKAN addon:

<https://github.com/4Science/dspace-ckan>



# DSpace-CRIS: Dataset context



Please use this identifier to cite or link to this item: <http://hdl.handle.net/123456789/13>

Title: Example of XLS to setup CRIS data model

Authors: [Bollini, Andrea](#)

*person(s)*

Issue Date: 2016

Project: [Gamma ray burst](#)

*project(s)*

Related Publication: [Research Data sample item](#)

*publication(s)*

Conference: [Conference on High Energy](#)

*event(s)*

Abstract: This is the sample data model configuration shipped with the DSpace-CRIS application. See [org.dspace.app.cris.batch.StartupMetadataConf](http://org.dspace.app.cris.batch.StartupMetadataConf) for more details.

URI: <http://hdl.handle.net/123456789/13>

Appears in Collections: [Datasets](#)

Show full item record



DOWNLOAD... 100MB CSV? 1GB+ ...

SCOPUS™ Citations  
**10**  
checked on Nov 8, 2016

WEB OF SCIENCE™ Citations

Title: DSpace-CRIS 5 Technical documentation

Authors: [Bollini, Andrea](#)   
[Pascarelli, Luigi Andrea](#)

Issue Date: 2016

Project: DSpace-CRIS Open Source: CRIS made easy with DSpace

- Related Dataset:
- [Example of CSV to import items into repository](#)
  - [Example of XLS to import organizations into repository](#)
  - [Example of XLS to import projects into repository](#)
  - [Example of XLS to import researchers into repository](#)
  - [Example of XLS to import organizations into repository](#)

*back link*

# Dspace-CKAN: preview tabular & geospatial data



*Paginated and filterable*

Grid Graph Map 639 records << 1 - 100 >> Search data ... Go » Filters

lon	lat	clase	municipio	nombre	dirección	actividad
-16.637585	28.1580...	turismo	Vilaflor	Hotel Vill...	Cm San ...	alojamie...
-16.637835	28.1549...	turismo	Vilaflor	El Somb...	Cl Santa...	alojamie...
-16.662599	28.1216...	turismo	Vilaflor	El Nogal	Cm Real...	alojamie...
-16.636974	28.1593...	turismo	Vilaflor	Tio Paul...	Pz La Igl...	alojamie...
-16.636297	28.1585...	turismo	Vilaflor	Los Gira...	Cl Dolor...	alojamie...
-16.66551	28.1187...	turismo	Vilaflor	Los F...	Cm G...	turismo
-16.666047	28.116...	turismo	Vilaflor	Los F...	Cm G...	turismo
-16.636497	28.157...	turismo	Vilaflor	Los F...	Cm G...	turismo
-16.366122	28.512...	turismo	Vilaflor	Los F...	Cm G...	turismo
-16.35427	28.496...	turismo	Vilaflor	Los F...	Cm G...	turismo
-16.3502...	28.525...	turismo	Vilaflor	Los F...	Cm G...	turismo
-16.425668	28.502...	turismo	Vilaflor	Los F...	Cm G...	turismo
-16.300875	28.408...	turismo	Vilaflor	Los F...	Cm G...	turismo

Grid Graph Map 3 records << 1 - 3 >> albatros Go » Filters

lon	lat	clase	municipio	nombre	dirección	actividad
-16.783282	28.1292...	turismo	Adeje	Albatros	Av Jable ...	alojamie..
-16.6439...	28.0056...	turismo	Arona	Parque Albatros	Lg Costa...	alojamie..
-16.610889	28.0260...	turismo	San Miguel	Parque Albatros	Ur Golf d...	alojamie..

**Filters**

Field: lon

Add



# DSpace-CKAN: preview tabular & geospatial data



*different visualization*

Grid Graph **Map** 639 records << 1 - 100 >> Search data ... Go » Filters

lon	lat	clase	municipio	nombre	dirección	actividad
-16.637585	28.1580...	turismo	Vilaflor	Hotel Vill...	Cm San ...	alojamie...
-16.637835	28.1549...	turismo				
-16.662599	28.1216...	turismo				
-16.636974	28.1593...	turismo				
-16.636297	28.1585...	turismo				
-16.66551	28.1187...	turismo				
-16.666047	28.1168...	turismo				
-16.636497	28.1572...	turismo				
-16.366122	28.5127...	turismo				
-16.35427	28.4963...	turismo				
-16.3502...	28.5253...	turismo				
-16.425668	28.5024...	turismo				
-16.300875	28.4085...	turismo				

Grid Graph **Map** 639 records << 1 - 100 >> Search data ... Go » Filters

Latitude / Longitude fields  
 GeoJSON field  
Latitude field: lat  
Longitude field: lon  
Update  
 Auto zoom to features  
 Cluster markers

Leaflet | Map data © 2011 OpenStreetMap contributors, Tiles Courtesy of MapQuest m

# All that is powered by REST webservices (CKAN datastore)

- Machine to machine interaction  
(researchers can use the datasets from  
R or other workbench tools)
- New applications can immediately re-use  
such data (for example a mobile app)



# What about images?

- An astrophysics image can go over 5GB!
- High-quality scanned book have images typically over 100MB for each page
- Medical images have similar dimensions
- The structure of images sequences are complex and relevant (page sequences, evolution of phenomena in medical images, etc.)



# Beyond traditional publications



Image files (medical, astrophysics, cultural heritage: digitalized manuscript, rare books, etc.) need to be consulted online, discussed and commented / annotated

IIIF protocols and formats allow you to meet these requirements in a standard and understandable way (for both humans and machine)



Image 7 of 8 Go

CONTENTS << Dante con l'esposizioni di Christoforo Landino et d'Alessandro Vellutello sopra...

Index Thumbnails

Indice

- Canto I
- Canto II
- Canto III
- Canto IV
- Canto V
- Canto VI
- Canto VII
- Canto VIII

com-  
Inui-  
refcit  
endo  
mento  
o dif-

A R G O M

GIUNTO Dante sopra il se  
peccato della invidia, trou  
licio, lequali haueano cuciti g  
tra queste troua Sapia donna S'

MORE INFORMATION

About the item

Title

Dante con l'esposizioni di Christoforo Landino et d'Alessandro Vellutello sopra la sua Comedia

Other Titles

Divina Commedia

Authors

- Dante Alighieri
- Christoforo Landino
- Alessandro Vellutello

Issue Date

1578

Publisher

Appresso Giovambatista (In Venetia)

Type

Book

Related Event

page Dante007.tif of -

Previous Result 3 results found for 'Dante' Clear Next Result

Download Bookmark Share



Discussion around the IIF Support in DSpace: <https://wiki.duraspace.org/display/DSPACE/IIF+and+DSpace>



IIF Image API  
allows a smooth  
interaction with the  
image files

The screenshot displays a digital library interface. At the top, there is a search bar with the text "Go" and navigation icons. Below the search bar, a search result is shown for "Landino et d'Alessandro Vellutello sopra...". The main content area features a large image of a manuscript page with the text "com- Inui- rescit endo mento o dif- A R G O M G I V N T O Dante sopra il se peccato della inuidia, trou licio, lequali haueano cuciti g tra queste troua Sapia donna S". The word "Dante" is highlighted in blue. To the right of the image is a sidebar with the heading "MORE INFORMATION" and sections for "About the item", "Other Titles", "Authors", "Issue Date", "Publisher", "Type", and "Related Event". The "About the item" section includes the title "Dante con l'esposizioni di Christoforo Landino et d'Alessandro Vellutello sopra la sua Comedia". The "Authors" section lists "Dante Alighieri", "Christoforo Landino", and "Alessandro Vellutello". The "Issue Date" is "1578". The "Publisher" is "Appresso Giovambatista (In Venetia)". The "Type" is "Book". At the bottom of the interface, there is a search bar with the text "page Dante007.tif of -" and navigation buttons for "Previous Result", "3 results found for 'Dante' Clear", and "Next Result".

Discussion around the IIF Support in DSpace: <https://wiki.duraspace.org/display/DSPACE/IIF+and+DSpace>

IIIF Presentation API  
allows to keep  
metadata close to the  
resource and build a  
native web UI

Discussion around the IIIF Support in DSpace: <https://wiki.duraspace.org/display/DSPACE/IIIF+and+DSpace>

A native web UI allows easy integration with other services/ technologies as web annotation

The screenshot shows a digital manuscript viewer. On the left, a table of contents lists 'Canto I' through 'Canto VIII', with 'Canto VII' highlighted in blue. The main area displays a scan of a manuscript page with the word 'Dante' highlighted in blue. A red box is drawn around the word 'Dante'. On the right, a sidebar titled 'MORE INFORMATION' contains metadata for the item, including 'About the item', 'Other Titles', 'Authors', 'Issue Date', and 'Publisher'. The authors listed are Dante Alighieri, Christoforo Landino, and Alessandro Vellutello. The issue date is 1578 and the publisher is Appresso Giovambatista (In Venetia). At the bottom of the viewer, there are icons for download, bookmark, and share.

here, actually a transcription driven by IIF Search API

Discussion around the IIF Support in DSpace: <https://wiki.duraspace.org/display/DSPACE/IIF+and+DSpace>



# Beyond traditional publications



... the same requirements apply to audio and video content

- Streaming
- Internal structure
- Annotation / commenting / transcript

Adopt an open standard: the MPEG-DASH format allows adaptive streaming over simple html client with full support for multiple tracks, ToC, subtitles



Discussion around the IIF Support in DSpace: <https://wiki.duraspace.org/display/DSPACE/IIF+and+DSpace>



# Open peer-review

**Central place for the daily research and dissemination activities** of researchers

An open peer-review module has been released for the DSpace platform



<https://github.com/arvoConsultores/Open-Peer-Review-Module/wiki>



# Open peer-review

Regardless to the specific implemented functionalities, I want to highlight the workflow:

1. Start from the repository («live» content: working ~~paper~~ research)
2. Capture new content: review, comments
3. Expose this content to allow further analysis
4. Capture or produce the analysis
5. Expose the analysis results (reviewer ranking, quality indicators, etc.)



...repositories are **nodes in a larger network, contributing their collective contents** to a global knowledge commons on top of which value added services can be built.



# Signposting - <http://signposting.org/>



Signposting is an approach to make the scholarly web more friendly to machines exposing relations as Typed Links in HTTP Link headers

The following discovering patterns are currently defined:

- Author
- Bibliographic Metadata
- Identifier
- Publication Boundary
- Resource Type

The Signposting approach is fully aligned with hypermedia (REST, HATEOAS) lines of thinking regarding web interoperability.



Signposting - <http://signposting.org/>



As an example, Herbert Van de Sompel and Michael L. Nelson are the authors of the paper with DOI <https://doi.org/10.1045/november2015-vandesompel>; their respective ORCIDs are <http://orcid.org/0000-0002-0715-6126> and <http://orcid.org/0000-0003-3749-8116>



# Signposting - <http://signposting.org/>



curl -I "<https://doi.org/10.1045/november2015-vandesompel>"

**HTTP/1.1 303 See Other**

**Location:** <http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html>

**Link:** <<http://orcid.org/0000-0002-0715-6126>> ;

**rel="author",**

<<http://orcid.org/0000-0003-3749-8116>> ;

**rel="author"**



- Successor of the OAI-PMH protocol and much more...
- Faster, reliable and scalable
- Allows real-time notification (and recovering of missed messages)
- Drives resource synchronization: content and metadata are both managed





# DSpace 7

- A new single UI built on top of a freshly REST API based on the HATEOS principles is under development: <https://wiki.duraspace.org/display/DSPACE/DSpace+7+UI+Working+Group>
- A ticket to stimulate discussion about the implementation of the signposting pattern has been created: <https://jira.duraspace.org/browse/DS-3589>
- A first implementation of resourcesync for DSpace was produced: <https://github.com/CottageLabs/DSpaceResourceSync> a ticket now exists to resume such implementation: <https://jira.duraspace.org/browse/DS-3590>



- A revision of the current data model is needed
- Precise identification of persons, organizations, projects, concepts and linked resources (dataset, different versions etc.)
- Avoid loss of details to allow a fine grain and effective interoperability



Thank for you attention!

Andrea Bollini

[andrea.bollini@4science.it](mailto:andrea.bollini@4science.it)

skype: a.bollini

linkedin: andreabollini

orcid: 0000-0002-9029-1854

