

Invenio at UAB

11 years after

Invenio User Group Workshop 2017
Heinz Maier-Leibnitz Zentrum (MLZ)
21-24 March 2017

Ferran Jorba
Universitat Autònoma de Barcelona
Ferran.Jorba@uab.cat

Quick summary

- 2006: started with UAB institutional repository, <http://ddd.uab.cat> (153.000 records).
- 2010: second instance, Catalan Language and Literature Database, <http://traces.uab.cat> (107.000 records).
- 2014: third instance, Catalan Musical Sources Catalogue, <http://ifmuc.uab.cat> (10.000 records).
- DDD is the main installation, others follow.
- Single server, flat system, many shared resources.
- Invenios at `/home/{ddd,traces,ifmuc}/invenio/`, Apache-itk.



traces



What do we use from standard Invenio?

- Submission forms and conversions.
 - *With extra Python scripts to complete Marc21 conversions.*
- Bibliographic Marc21 records.
 - *Changing \$\$ to \$ and extra space after subfield marks.*
- Collections, real and virtual; portalboxes.
 - *Adding a statistics link for each portalbox.*
- Indexes and searches.
 - *With pre-generated .txt files to speedup fulltext indexing.*
- OAI server and client
 - *With extra helper scripts.*
- Users, roles, permissions.

On our own: bibformat elements

Why?

- Marc21 usage in Atlantis (the basis for any new installation) site is so substandard, that we have practically rewritten all Python logic so it accepts real world records.

How?

- Lots of work.

Interested in going mainstream?

- As much as possible, with 3.x.

On our own: local files and URLs

Why?

- When we started, it was not clear how to upload documents attached to records for imported records.
- We wanted to share documents between old and new system.

How?

- They are plain files and directories, in shared NFS disk, URLs just point to them. Independent from any particular software.

Interested in going mainstream?

- Not really.

On our own: thumbnails

Why?

- When we started, there were no thumbnails for text documents and no standardised procedure.

How?

- Two thumbnails per document: static (.ico) and dynamic or zoom (.gif). Python, mupdf and Imagemagick.

Interested in going mainstream?

- Maybe.

On our own: Marc editor

Why?

- We really don't understand it, it seems too cumbersome, rigid, or both.

How?

- Basically a simple textarea, with some extras for file uploading and web page capture.

Interested in going mainstream?

- Maybe. We miss accessing older record versions.

On our own: journals

Why?

- When we started with DDD, we had to load several thousands articles for some dozen of UAB journals, and there was no established procedure.

How?

- Using Marc tags 866 (in journal records), 773 and 973 (in article records) and creating searches and links in bibformat elements.

Interested in going mainstream?

- Maybe.

On our own: usage and content statistics

Why?

- They don't address our needs.

How?

- We follow our local consortium definitions and requirements.
- Content statistics: from database dumps.
- Usage statistics: from Apache logs, cleaning robots and librarian's IPs
- Python, jdresolve, geoip, SQLite, etc.

Interested in going mainstream?

- Probably not.

On our own: submission review workflow

Why?

- We don't know how it works. Is it documented anywhere?

How?

- Alexander Wagner explained once how they distributed incoming records to collections to be reviewed.
- We tried to implement this workflow, but it is fragile.

Interested in going mainstream?

- We'd love it!
- Similar workflow for web forms and OAI harvesting.

On our own: gateways and imports

- Several scripts for importing records from different sources and databases, local (UAB) and remote.
- Some are batch (unattended, cron jobs), some with web fronted, manually activated.
- Gateway between local research system and DDD to upload documents.
- Procedures to double check OAI imports.
- etc.

On our own: maintenance scripts

- Lots of scripts to complete, correct or check records.
- Librarian-driven global changes.
- Subject of my other talk.

What we are pondering at UAB

- RDA implementation.
- Authority control.
- Review workflow.
- More import and export gateways.
- Bibliographic export to ORCID site, ResearchGate, etc.
- RDA.
- Research data, reusing DDD?
- Digital preservation, data fixicity.

What we expect from new releases

- Authority control.
- Review workflow.
- Bibliographic export to ORCID site, ResearchGate, etc.
- Better Marc21 implementation, RDA compatibility.
- Research data handling.
- Data fixicity and digital preservation.
- The wonderful Zenodo look and feel.

What we fear from 3.x

- Too much complexity: I cannot follow developments any more.
- The spirit of free software: be able to freely fix and modify.
- Too much departure from administration of a standard Debian box: pip, virtualenv, etc.
- Too much isolation among our Invenio instances: docker.
- Too many hot-and-new (but-maybe-unproven) buzzwords and technologies.
- Non standard Marc and/or RDA implementation.

CERN, your turn!

Thanks for listening!

