

Setting the scene...

Just to clear up the scenario in Spain regarding the National CRIS...

Little by little Dialnet CRIS is consolidating as a service to provide research information management and public eye window to research institutions as well as a tool to build multi-organization research portals.

The latest milestone has been the confluence with Hercules Project to build a National portal. This will provide Dialnet CRIS with the whole picture of the Research Information in Spain.

Data Aggregation to build a National CRIS

Favorable background

Already mentioned in previous sessions, a leading role in this kind of ambitious project is particularly necessary. The Dialnet non-profit Foundation, originally created at the last corner of the library of the smallest University in Spain, is eventually playing that role nationally.

The IT Department has been working side by side with the Library professionals on a regular basis and personally sharing the same workspace. It was a question of time that Dialnet ended up working along with the Spanish Libraries Consortium REBIUN. That relationship came with the implicit benefits of the direct connection of expertise between both worlds in designing those requirements and implementing the solutions.

Thanks to the IT team support, Library professionals forget about manual work that can be done by automated processes and allows them to play the leading role in validating its institutional information.

Integration

Challenges

1. Individual Portals, 1º stage on the roadmap (example UR, UCM, not only Universities, agricultural research center, SACYL, etc.)
2. Regional Portals, 2º stage on the roadmap (example BUCLE) pivoting on our local identifier (dialnetId)

3. National Portal, 3^o stage on the roadmap (INV.ES)

Solution: Centralized DB and multi-tenancy (Mode of operation where multiple independent instances (tenants) operate in a shared environment)

Pitfalls

1. Administrative issues (create the necessity, bottom-up approach, not easy at the beginning, not supported by the Spanish administrations)
2. Interoperating with multiple sources UXXI, CERIF, HERCULES, SIGMA, MANUAL
3. Interoperating with multiple formats CSV, XMLs, JSON, OAI-PMH
4. Dealing with PIDs in the source. 100% PIDs and deduplication in data providers is unachievable. Dialnet perspective has always been aligned with conforming standardization but with a multi-identifier interoperability.

Solution: Inclusive perspective rather than exclusive.

Workflow

- Aggregation

Instead of making institutions (with more or less sophisticated/digitized Research Systems) adapt to a specific format, the more operative workaround is their aggregation into a centralized structure. At this stage of collecting multiple data sources it is of paramount importance updating their content independently, preventing overlapping.

- Mapping

Mapping the content into a common structure by pivoting on the Dialnet local identifier. Best way of pinpointing deduplication.

- Data curation

After aggregating, mapping and deduplicating it is simpler to deal with PIDs resolving (supported by multiple identifiers matching).

Automated processes are in charge of populating PIDs:

DOIs (manual inputs from Crossref and automated inputs from retrospective data in Crossref)

DATAcite DOIs provider

OPEN_FUNDER_IDs (Crossref from DOIs)

ORCIDs (OJS Journals, VIAF, manual inputs from Crossref)

ROR

On the agenda: imports from Wos / Scopus and Wikidata

Pipeline (Chart)

1°. Multiple Data Providers:

UXXI

CERIF compliant sources like sources following 'OpenAIRE guidelines for CRIS managers'

Proprietary CSV files

HERCULES RIS

SIGMA

MANUAL data management for tiny/small organizations

2°. Data Package

3°. Validation

4°. Generate Deltas (between two data packages and between data package and DB)

5°. Load

Conclusion

Lessons learned

...

On schedule

Provided interoperation for Literature Production with OpenAIRE, it is time to implement the interaction providing CRISs information.

To keep adding the contents of all the CRIS systems in Spain.

Start working with Latinamerican countries in order to follow a similar strategy to build organization portals, regional and in the long run, may be national CRISs (by gathering all the informacion, attaching impact metrics and framing all of them in National Portals).

Generating multipurpose reports for the different actors that interact with the portal, researchers, evaluators, funders, general public, etc.