

Frontiers & Invenio

Angel Luis Castro & Jorge Jardines

Who we are

Frontiers is an online platform for the scientific community to publish open-access articles and network with colleagues. Our grand vision is to build an Open Science platform where everybody has an equal opportunity to seek, share and generate knowledge.

- Fast, open-access publication
- Rigorous peer-review in real-time thanks to our interactive, online interface
- Detailed metrics to follow the impact of articles
- A networking platform made especially for the scientific community

Repositories

- PubMed comprises more than 23 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central (PMC) and publisher web sites.
- PMC is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM). 2.9 MILLION Articles are archived in PMC
- arXiv is an e-print service in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance and statistics. 890 thousand e-prints.
- CrossRef. Its API allows us to perform queries for DOIs using citation metadata, and to retrieve deposited metadata for a DOI. Over 60 million journal articles and other content items

Why Invenio

- Is a free & open-source software
- Enables us to run our own document repository.
- Strong implementation and community.
- Good solution for disambiguation.
- Support OAI-PMH out of the box.
- Can act as provider of content.
- OAI server is included.
- Allows post-processing

Our Goals

- Have our own document repository in Invenio.
- Harvesting/manually importing articles from different repository sources to Invenio.
- Make most of data available by the repository.

Example of ArXiv:

- a. The metadata is extracted from PDF
 - b. The PDF full text is attached
 - c. The plots are extracted and attached
 - d. The references are extracted from PDF and metadata is enriched
 - e. The keywords are extracted from PDF and the metadata is enriched
 - f. The rich paper is finally submitted to the system.
- Provide API for our Journal platform
 - API built out of the box
 - Support multiple matching criteria.

Expectations of the workshop

- Learn more about Invenio.
- Meet the community.
- Get a few hint / paths to test performance
 - Try & understand Shard Manager
 - Alternative would be not keeping unmatched record in the live system but rather in the holding pen only.

Thank you