Contribution ID: **1**   Type: **not specified**

Things you can do dumping your Invenio database into a flat file

Wednesday, 22 March 2017 09:00 (15 minutes)

Invenio database design and interfaces are optimized for fast end user search and retrieval. As administrators, we can add indexes at will and use them via web or API. However, many maintenance tasks are not well covered with those indexes.

For most of those cases, reading the records sequentialy is the optimal solution. However, if the database is large enough, reading them via Invenio API may take hours, while the system slows down and it may become unresponsive.

In this presentation I'll show a small Python tool that uses Invenio API and a SQLite database as cache to keep an up to date flat file with your bibliographic records.

We'll see how whith this flat file it is much faster and easier to do tasks like generate specialised statistics, quality control, automatic record enrichment or cleaning, or even creating exotic indexes or counters.

Primary author: JORBA, Ferran (Universitat Autònoma de Barcelona)

Presenter: JORBA, Ferran (Universitat Autònoma de Barcelona)

Session Classification: Legacy

Track Classification: Main track