Guidelines for Organizing Visual Exhibits in Invenio 2.x-Based Institutional Repositories

Aleksandra Żukowska and Mateusz Kowalski

Aleksandra Żukowska, Department of Information Studies, University of Warsaw; Mateusz Kowalski, School of Computing, University of Gdańsk

Abstract

First version of Invenio considered albums as a first-class citizen of a system while treating photos as a subrecord of an album limiting user's ability to add extra metadata to single pictures or index/search them. Additionally, management issues came into play, when user wanted to place one photo in multiple albums the only way to achieve that was to upload it twice, therefore introducing high redundancy.

The goal of photo albums changes in Invenio 2 is to solve aforementioned issues, by extracting pictures from albums, creating new records out of them and translating them to JSON from previous marc21 format to gain even more flexibility.

Table of Contents

1 Introduction

The main terms connected to the project are: Python, JSON, jsonref, Marc21. Here is the short description of all of them:

Python – General-purpose high-level programming language. Created by Guido van Rossum in 1991. Python supports many programming paradigms. It's a dynamic typing language with automatic memory management and with huge community and support.

JSON – JavaScript Object Notation, human-readable open standard format that transmits attribute-value pairs. JSON I a language-independent data format, despite its name.

Jsonref – standard for dereferencing JSON reference objects. It allows to split JSONs to multiple documents and join them by including appropriate pointers. It solves the problem of cyclic references and adds opportunity to use lazy evaluation.

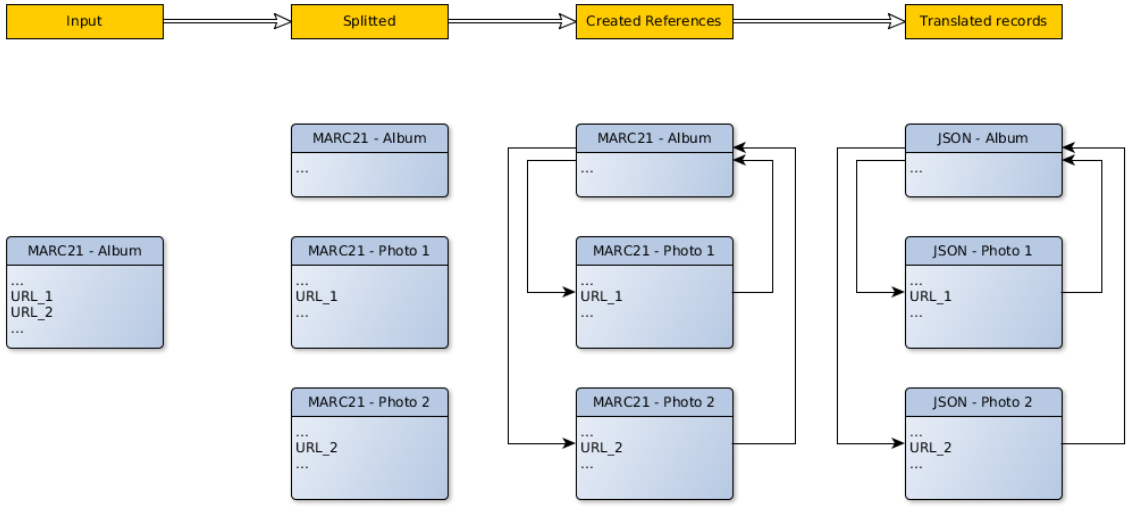
Marc21 – standard for bibliographic data. It consists of key-value pairs, where value can contain subfields. Key has a three-digit tag and two indicators, which might be empty – shown as '_' – or a number describing particular meaning of a field.

2 Previous state

Previous versions of Invenio treated Albums as a first-class citizen – internally it was a "record". Record is an object that can be indexed, searched and referenced by its ID. Photos on the other hand were just a document objects included in an album, without the possibility of being referenced outside the system. This solution prevented single pictures from appearing isolated from albums, therefore limiting features of Invenio.

Additionally, usage of Marc21 format introduced many compatibility issues due to the lack of native support of media in this standard. Therefore the decision was made to translate all media photos and albums to JSON format with appropriate json-schema.

3 Implementation



Picture above shows the main stages of the project. It includes splitting albums to extract photos, create references and translating from marc21 to json.

1. Splitting record included filtering URL which might be in an album and create new "record seed" from the ones that satisfy defined predicate. For example current module accepts all photos from cds.cern.ch apart from icons. After extracting new records, predefined key-value pairs from album are copied to newly-created photos – particularly those which make sense from the perspective of a single, isolated picture.
2. Creating references is a straight forward process of inserting subdocuments like {'$ref': <id>} to all of the records. After that system is able to resolve those links and replace them seamlessly with the whole – external – JSON
3. Translation is carried out as a last step in the process. Module collects all the translation rules from different places – like a common set for global Invenio, as well as album/photo specific from local modules. After that the translation is launched and as a result user gets JSON.

Example of translation

Input marc21:

001 __ 1800700
269 __ $$aGeneva$$bCERN$$cApr 1962

Output JSON:

```
{
        "control_number": "1800700",
        "imprint": {
                "date": "Apr 1 962" ,
```

```
                ”name_of_publisher ”: ”CERN” ,
                ”place_of_publication”: ”Geneva”
        },
}
```

4 Conclusion

Through this project we achieved:

- Way of adding metadata to single photos
- Indexing and searching single photos
- Possibility of referencing particular photo without an album
- Freedom of changes in the structure with JSON format

5 Future work

This project can be expanded in many ways. For example :

- Add new JSON Resolvers to cover various data sources like MongoDB, FTP, Redis, etc.
- Use fixtures in unit tests
- Integrate scripts with the whole system
- Integrate scripts with celery to run it periodically