

ALBUMS IN INVENIO 2.X

Łukasz Chrząszcz

August 25, 2015

CERN Openlab Summer Student

Albums and photos

- Previously...
- Goal – Major tasks
- Results

PREVIOUSLY

```
001__ 1800700
005__ 20150505091919.0
037__ $$aCERN-PHOTO-62045364_a
110__ $$aCERN PhotoLab
245__ $$aNo caption
260__ $$c1962
269__ $$aGeneva$$bCERN$$cApr 1962
300__ $$aPhotographic negative$$c10cmx12.5cm
500__ $$3Box 06_4651-5800$$aImage scanned from original photo negative on 09 Dec 2014
541__ $$e5364
595__ $$912092014-06_4651-5800-10cmx12.5cm$$aArchive Collection$$bScanned by Contentra Technologies
690C_ $$aPHOTO
774__ $$aALBUM$$r762638
8564_ $$s2277475$$uhttps://cds.cern.ch/record/1800700/files/5364_a.jpg$$yImage 5364_a
8564_ $$s1231475$$uhttps://cds.cern.ch/record/1800700/files/5364_a.jpg?subformat=icon-640$$xicon-640
8564_ $$s3370955$$uhttps://cds.cern.ch/record/1800700/files/5364_a.jpg?subformat=icon-1440$$xicon-1440
8564_ $$s7070355$$uhttps://cds.cern.ch/record/1800700/files/5364_a.jpg?subformat=icon-180$$xicon-180
916__ $$r$$w201503
960__ $$a86
963__ $$aPublic$$bnotvisible
980__ $$aPHOTOARCIMAGES
999__ $$aIMAGE
```

- All albums and photos in marc21
- Albums contains direct urls to photos
- Problems?
 - Management issues if one photo should be in many albums
 - Searching/indexing one particular photo
 - Lack of metadata connected to the single photo
 - JSON would be better than marc21

- JSON Reference – allow records to reference other records
- Translation – marc21 to JSON
- Separate records – Album and photos

```
{  
  'photo':  
  {  
    '$ref': '123456'  
  }  
}
```

- When accessed, the \$ref part is replaced with proper structure
- Lazy evaluation
- Easily expandable solution – Flask Registry

Record translation

- It has to be reversible
- A lot of edge cases, exceptions or malformed records
- Batch of 26k records and... try, catch, try, catch, try, catch...

TRANSLATION – RESULT

```
[
  {
    "album": "762638",
    "control_number": "1800700",
    "date_and_time_of_latest_transaction": "20150505091919.0",
    "image": [
      {
        "link_text": "Image 5364_a",
        "photo_url": "http://cds.cern.ch/record/1800700/files/5364_a.jpg"
      },
      {
        "photo_url": "http://cds.cern.ch/record/1800700/files/5364_a.jpg?subformat=icon-640"
      },
    ],
    "immediate_source_of_acquisition_note": [
      {
        "accession_number": [
          "5364"
        ]
      }
    ],
    "imprint": {
      "date": "Apr 1962",
      "name_of_publisher": "CERN",
      "place_of_publication": "Geneva"
    },
    "internal_note": [
      {
        "internal_note": "Archive Collection",
        "subject_note": "Scanned by Contentra Technologies"
      }
    ],
  },
  ...
]
```


Album splitting

- Extract photos from album creating new records
- Some metadata will be copied, some will be only in an album
- Again – a lot of exceptions

ALBUM SPLITTING

```
001__ 39135
003__ SzGeCERN
005__ 20150818103838.0
035__ $$PHOPHOC$a0000258
037__ $$aCERN-EX-8904340
100__ $$aPatrice Loiez
...
6531_ $$9CERN$$aGolden Jubilee Photos $$uhttp://cdsweb.cern.ch/record/801663
8564_ $$qhttp://preprints.cern.ch/.../8904340.gif$$xicon
8564_ $$qhttp://preprints.cern.ch/...$$x1$$yAccess to the pictures
8567_ $$2MediaArchive$$d\\cern.ch\...\8904340.jpg$$xAbsolute master path
8567_ $$2MediaArchive$$uhttp://mediaarchive.cern.ch/.../8904340-A4-at-144-dpi.jpg
8567_ $$2MediaArchive$$uhttp://mediaarchive.cern.ch/.../8904340-A5-at-72-dpi.jpg
8567_ $$2MediaArchive$$uhttp://mediaarchive.cern.ch/.../8904340-Icon.jpg
916__ $$r$$w199800
923__ $$rMichelini
923__ $$pPoint 6-LEP
960__ $$a81
961__ $$c20050511$$h1437$$mMMD0$$x19980114
963__ $$aPUBLIC
970__ $$a000000258MMD
980__ $$aPHOTOLAB
```

Thank you!