SIMBAD Pipeline entry

A&A Meeting 06/03/2020



OBERTO A. MANTELET G.

Current process A&A → SIMBAD

- Table Of Content:
 - → (FTP XML, python script) bibliography in SIMBAD
- Text:
 - → (EDP PDF, DJIN) object name & small tables data in SIMBAD
- Tables (big):
 - → (VizieR, COSIM) table data in SIMBAD

New process

- Table Of Content + Article text
 - → (FTP XML, bash script) **XCDS**
 - → (XCDS) bibliography in SIMBAD
 - → (XCDS, DJIN) object name & small tables in SIMBAD
- Tables (big):
 - → (XCDS/VizieR, COSIM) table data in SIMBAD

□ 1 - Download

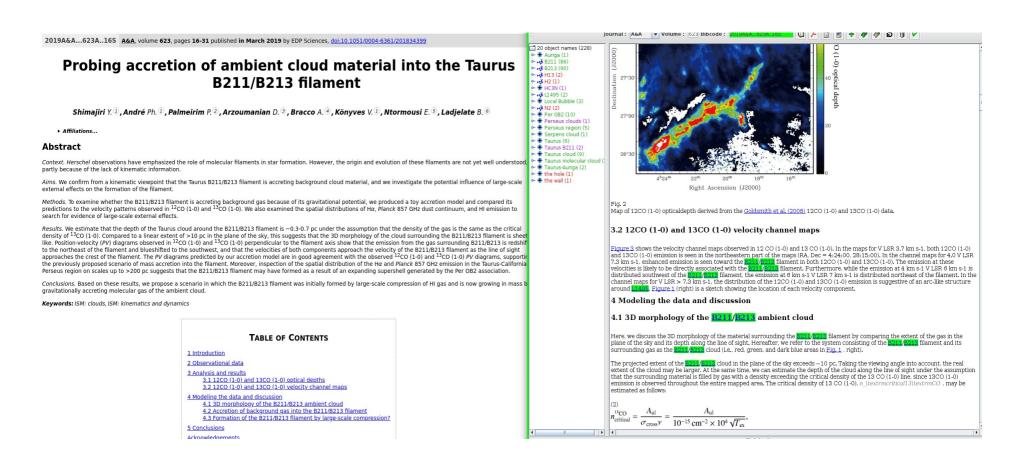
- Once a volume id finished : download from FTP cdsarc all gz files
- Extract all files from one volume (some articles are replaced)
 - XML, JPEG, PDF ...

2 - Convert

- XSLT Filters to transform XML
 - → XCDS for DJIN
 - → TOC for SIMBAD
- + extra characters / formula

3 - Name recognition

DJIN parsing to extract object names



□ 4 - Table data

- COSIM extract columns from tables (VizieR)
 - Xmatch with SIMBAD to add some data