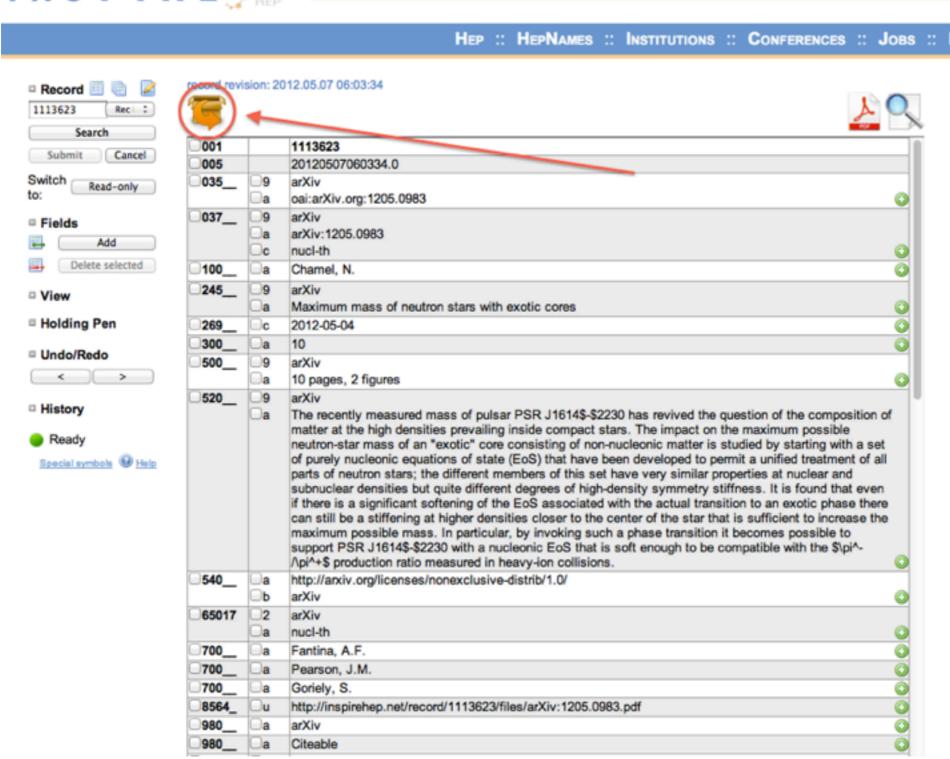
Invenio - refextract

Alessio Deiana INSPIRE

From bibedit



Welcome to INSPIRE! INSPIRE is now in full operation and supersedes SPIRES. SPIRES front end at all mirrors



Command line

\$ /opt/invenio/bin/refextract

- bibtask
- deals with records

\$ /opt/invenio/bin/docextract

- standalone
- deals with pdfs directly

Docextract

• \$ docextract example.pdf --out example.xml

```
<record>
<controlfield tag="001">1091122</controlfield>
<datafield tag="999" ind1="C" ind2="5">
        <subfield code="o">1</subfield>
        <subfield code="s">PHRVA,D42,3613</subfield>
</datafield>
<datafield tag="999" ind1="C" ind2="5">
        <subfield code="o">2</subfield>
        <subfield code="s">APHYE,7,263</subfield>
</datafield>
</datafield>
```

Refextract

- refextract --recids I-100
- refextract --recids 2 --create-ticket
- refextract --collections 'HEP'
- refextract -a (runs on added records)

KBs

- Is /opt/invenio/etc/docextract/*.kb
 authors.kb collaborations.kb journal-titles inspire.kb journal-titles-re.kb report numbers.kb books.kb journal-titles.kb
 publishers.kb special-journals.kb
- Journals KB is being moved to the database

Great

Done