

ISTI Open Access - Feature #20382

Reinforce the BibTeX export

Dec 29, 2020 07:47 PM - Leonardo Candela

Status:	Closed	Start date:	Dec 29, 2020
Priority:	High	Due date:	
Assignee:	Michele Artini	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	ISTI Open Portal Tasks		

Description

The bibtext file resulting from the export is not missing some mandatory information, e.g.

- **article** records are missing the **Journal**
- **inproceedings** records are missing the **Booktitle**

We should try to extract suitable information from the "source" field.

I kindly ask @michele.artini@isti.cnr.it to clarify what are the typologies of bibtext items resulting from our product typologies and the rest of mappings and transformations governing the production of the bibtext file.

History

#1 - Jan 05, 2021 10:04 AM - Michele Artini

The bib records are generated using the JBibTeX library (v.1.0.17).

The mapping produces the bib entities according to the following rule:

```
"Conference object"      -> BibTeXEntry.TYPE_INPROCEEDINGS
"Report"                 -> BibTeXEntry.TYPE_TECHREPORT
"Article"                 -> BibTeXEntry.TYPE_ARTICLE
"Part of book or chapter of book" -> BibTeXEntry.TYPE_INBOOK
"Book"                   -> BibTeXEntry.TYPE_BOOK
"Master thesis"          -> BibTeXEntry.TYPE_MASTERSTHESIS
"Bachelor thesis"        -> BibTeXEntry.TYPE_MASTERSTHESIS
"Doctoral thesis"        -> BibTeXEntry.TYPE_PHDTHESIS
"Patent"                 -> BibTeXEntry.TYPE_MISC
"Other"                  -> BibTeXEntry.TYPE_MISC
```

For each typology the following information are added:

```
record.getTitle()        -> BibTeXEntry.KEY_TITLE
record.getCreators()      -> BibTeXEntry.KEY_AUTHOR
record.getPublisher()     -> BibTeXEntry.KEY_PUBLISHER
record.getDate()          -> BibTeXEntry.KEY_YEAR
record.getDois()          -> BibTeXEntry.KEY_DOI
```

The "source" field was excluded from the mapping because the values were not cleaned.

#2 - Jan 05, 2021 06:03 PM - Leonardo Candela

- Status changed from New to In Progress

We will never have a dc:source field suitable for a net transformation because of the mismatch between the dc:source semantic and the bibtext fields.

We might follow a **"quick & dirty option"**, i.e. take the dc:source content and compile the BibTeXEntry as follows

- KEY_BOOKTITLE for Conference object;
- KEY_INSTITUTION for Report;
- KEY_JOURNAL for Article;
 - in this case we could also take only the dc:source string before the "(" and compile the KEY_JOURNAL field;
- KEY_BOOKTITLE for Part of book or chapter of book
- nothing for Book;
- nothing for Master thesis, Bachelor thesis, Doctoral thesis ... it seems no suitable source is exposed by People;

- KEY_HOWPUBLISHED for Patent;
- nothing for Other ... it seems no suitable dc:source is exposed by People;

Another strategy could be to use services that given a DOI return a bibtex ... this approach might be smart (at least for objects with a DOI).

Dunno whether OpenAIRE can offer something about that, I know that Crossref might offering something

<https://www.crossref.org/labs/citation-formatting-service/>

#3 - Jan 07, 2021 03:03 PM - Michele Artini

We might follow a "**quick & dirty option**", i.e. take the dc:source content and compile the BibTeXEntry as follows

- KEY_BOOKTITLE for Conference object;
- KEY_INSTITUTION for Report;
- KEY_JOURNAL for Article;
 - in this case we could also take only the dc:source string before the "(" and compile the KEY_JOURNAL field;
- KEY_BOOKTITLE for Part of book or chapter of book
- nothing for Book;
- nothing for Master thesis, Bachelor thesis, Doctoral thesis ... it seems no suitable source is exposed by People;

- KEY_HOWPUBLISHED for Patent;
- nothing for Other ... it seems no suitable dc:source is exposed by People;

I extended the mapping with your proposal.

Regarding the articles I try (using a regex) to extract the following info from the source field:

```
KEY_JOURNAL
KEY_VOLUME
KEY_PAGES
KEY_YEAR
```

NB. The KEY_YEAR is extracted from the **source** field only for articles to ensure the right references to the journal , otherwise it is obtained from the **date** field (that contains the publication date in People).

#4 - Jan 07, 2021 08:34 PM - Leonardo Candela

- *Status changed from In Progress to Closed*

- *% Done changed from 0 to 100*

I would consider the task complete for the time being, we improved the bibtex export by counting on the metadata we have in People.

The idea to use third party services (for items having a DOI) might be interesting for the future yet it might introduce mismatches between the information we have in People and what we expose.