

Thursday, October 12, 2006 ---- M.Barone

If we want to have something limited in scope, but well done, before Valencia, what is still missing is:

- 1) how to define queries in order to fill -in the right and complete way- the virtual collections containers
- 2) define what we would like to see in the output format
- 3) what is missing in the harvested metadata
- 4) define at least one document type and corresponding submission page

1) how to define queries in order to fill -in the right and complete way- the virtual collections containers

- a. **MB:** I think that the easiest way to do it is by keywords. Actually, we could have a list (no free-text) with just ***one*** keyword for each virtual collection, that is about 35-40 different keywords. That should be enough, because we just need a criterion to link a talk with one or more virtualcollections. The speaker should choose keywords from a list (multiple choice) and the manager/convener could be responsible for checking it, otherwise the manager could select the keywords when creating the entry for the talk. Also, there could be default keywords associated to categories, ex: "controls" will be the default keyword for the categId=50 (" Machine Design > Area, Global and Technical Systems > Controls"). Having default keywords could be a turn around to avoid the hard wiring (that make me worried).
JY: To do that, we would need to hack Indico as this is not an existing option. It is not immediate (for Valencia).
JF: keywords are one solution potentially, but they cannot be implemented in Indico before
- b. **JY:** keep perfect matching between Indico top level categories and Invenio collections (simplest solution)
MB: we have to consider that a talk about "Electron Source" can belong to more than one ILCagenda category: Institutions, Area...Systems, Conference & Workshops, ... Keeping perfect category matching bw the 2 systems will reduce their potentiality
MB: a "special" query can be defined for some virtual collection (as Beam Delivery System) that has correspondent categories in Indico. This could be showed in November as an example, waiting for a more stable solution.
- c. **JY:** when harvesting Indico, you can decide the matching with pre-defined values: Indico records can be converted with BibConvert module to be enriched with the appropriate keywords used to define the collections. It requires that you define for all Indico categories and subcategories the corresponding virtual collection(s).
MB: Yes, but BibConvert has to determine which keyword is appropriate and which is not... Therefore, the information has to be in the Indico metadata. Is that correct?
JY: This can be done at harvesting time by keeping within the BibConvert configuration the default keywords.
MB: How?

2) define what we would like to see in the output format

- a. **TWM**: we would like to see in the detailed output format of a talk/meeting something like:

----- **TALK** -----

Title (ok)
Speaker (ok), Institution (OK)
Date (ok)
Abstract (ok)

Title of the Meeting (not available)

List of files (ok) , file size (not yet available)

----- **MEETING** -----

Title (ok)
Date (ok)
Abstract (ok)
convener/chairperson (?) (not available)

speaker: list of files (ok) , file size (not yet available)

MB: In order to implement it, we need to have the meeting title in the talk metadata, and the speaker/talk info in the meeting metadata harvested from Indico

JY: Well, as suggested, drop the detailed view. I think it is useless... isn't it ? [...]

Before going in UDF BibFormat details, please let me know if providing only the link to Indico and not to detailed view is okay ? This is what we are doing on

<http://indicosearch.cern.ch> .

3) add what is missing in the harvested metadata

- a. See 2a
- b. See 1a , 1c
- c. File size (in progress; useful, but not necessary)

4) define at least one document type and corresponding submission page

- a. **TWM:** We could implement the submission of general files ('random files'), like spreadsheets.

MB: we could define an appropriate document type ('miscellanea', 'random files',...?), including several sub-categories like "Spreadsheets", "Letters", "Memo", ... [...] We should also define a list of keywords to be used for the submission of the miscellanea documents.

JY: My impression is that you are now calling "keyword" are what we call "subjects" at CDS (field **65017_a**). They are rather general large concepts that are defined in advance and that of course, could be used to create the virtual collections.

JY: Isn't it "Internal Notes" type of document ?

Summary:

MB: It would be nice to show that clicking on "Beam Delivery System" or any other virtual collection, you get the list of all the talks/meetings/'random files' related to it. Then, we could show how to get a sub-list by selecting just one collection between talks/events/'random files'.

JY: Do you have a backlog of internal documents (or 'random files'; hum, I do not like too much this term as it can include everything, also peer reviewed published article!) ? If yes, it would be useful to upload them so that the cross-search will be more visible.

Legenda

TWM: Tom Markiewicz;

JY: Jean-Yves Le Meur;

JF: Jim Fromm;

MB: Maura Barone