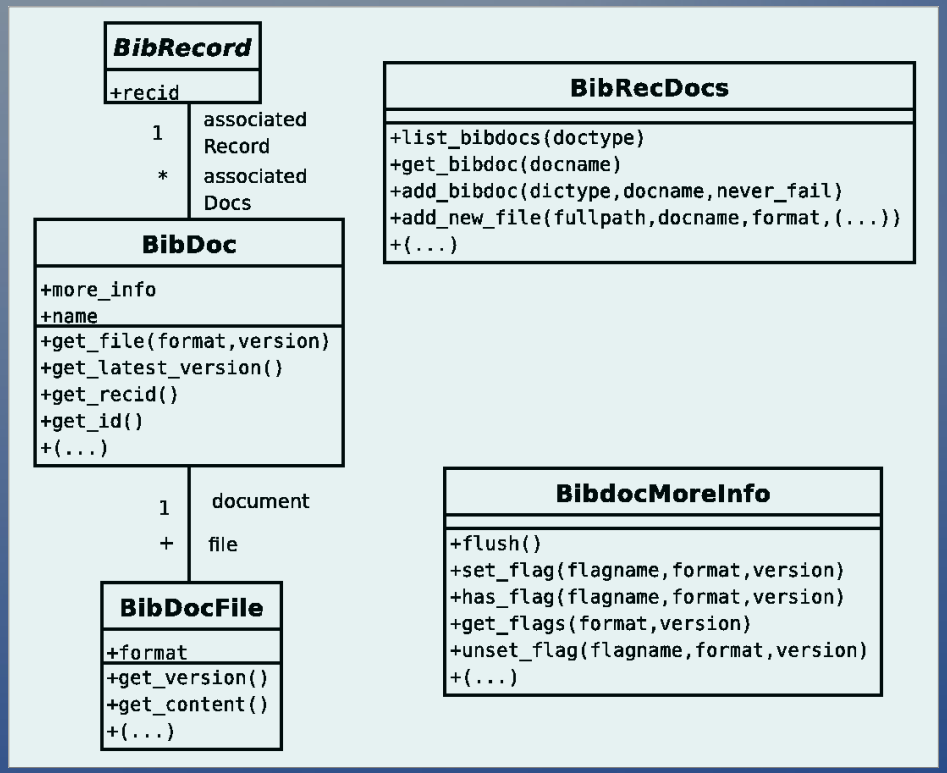
BibDocfile extensions

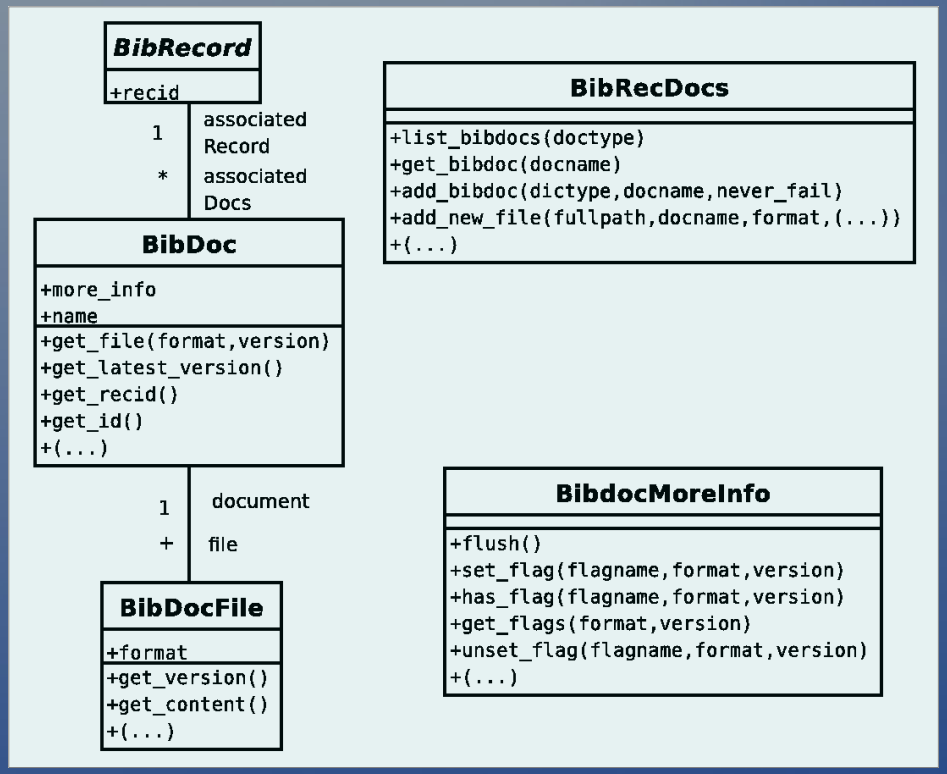
Piotr Praczyk piotr.praczyk@cern.ch
Somewhere between Meyrin and Saint Genis-Pouilly, 14.11.2012

Non-bibliographical data in Invenio



- Documents represented as BibDoc instances
- Document supports versions and different formats
- Internal data stored in a BibdocMoreInfo instance

Non-bibliographical data in Invenio

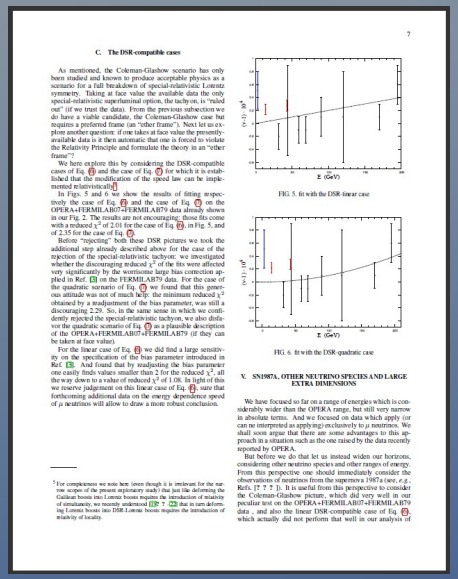


- Internal meta-data stored in a MoreInfo instance
- Link between a MARC record and the document
  - Every document must belong to exactly one record

Initial assumptions

- We need to make documents independent of the records
- We need a more elastic storage for document-dependent meta-data

Example: Figures



C. The DSR-compatible cases

As mentioned, the Coleman-Glashow scenario has only been studied and known to produce acceptable physics as a scenario for a full breakdown of special-relativistic Lorentz symmetry. Taking at face value the available data the only special-relativistic superluminal option, the tachyon, is "ruled out" (if we trust the data). From the previous subsection we do have a viable candidate, the Coleman-Glashow case but requires a preferred frame (an "ether frame"). Next let us explore another question: if one takes at face value the presently-available data is it then automatic that one is forced to violate the Relativity Principle and formulate the theory in an "ether frame"?

We here explore this by considering the DSR-compatible cases of Eq. (6) and the case of Eq. (7) for which it is established that the modification of the speed law can be implemented relativistically[3].

In Figs. 5 and 6 we show the results of fitting respectively the case of Eq. (6) and the case of Eq. (7) on the OPERA+FERMILAB07+FERMILAB79 data already shown in our Fig. 2. The results are not encouraging: those fits come with a reduced χ^2 of 2.01 for the case of Eq. (6), in Fig. 5, and of 2.35 for the case of Eq. (7).

Before "rejecting" both these DSR pictures we took the additional step already described above for the case of the rejection of the special-relativistic tachyon: we investigated whether the discouraging reduced χ^2 of the fits were affected very significantly by the worrisome large bias correction applied in Ref. [8] on the FERMILAB79 data. For the case of the quadratic scenario of Eq. (7) we found that this generous attitude was not of much help: the minimum reduced χ^2 obtained by a readjustment of the bias parameter, was still a discouraging 2.29. So, in the same sense in which we confidently rejected the special-relativistic tachyon, we also disfavor the quadratic scenario of Eq. (7) as a plausible description of the OPERA+FERMILAB07+FERMILAB79 (if they can be taken at face value).

For the linear case of Eq. (6) we did find a large sensitivity on the specification of the bias parameter introduced in Ref. [8]. And found that by readjusting the bias parameter one easily finds values smaller than 2 for the reduced χ^2, all the way down to a value of reduced χ^2 of 1.08. In light of this we reserve judgement on this linear case of Eq. (6), sure that forthcoming additional data on the energy dependence speed of μ neutrinos will allow to draw a more robust conclusion.

FIG. 5. fit with the DSR-linear case

FIG. 6. fit with the DSR-quadratic case

V. SN1987A, OTHER NEUTRINO SPECIES AND LARGE EXTRA DIMENSIONS

We have focused so far on a range of energies which is considerably wider than the OPERA range, but still very narrow in absolute terms. And we focused on data which apply (or can be interpreted as applying) exclusively to μ neutrinos. We shall soon argue that there are some advantages to this approach in a situation such as the one raised by the data recently reported by OPERA.

But before we do that let us instead widen our horizons, considering other neutrino species and other ranges of energy. From this perspective one should immediately consider the observations of neutrinos from the supernova 1987a (see, *e.g.*, Refs. [7 ? ?]). It is useful from this perspective to consider the Coleman-Glashow picture, which did very well in our peculiar test on the OPERA+FERMILAB07+FERMILAB79 data , and also the linear DSR-compatible case of Eq. (6), which actually did not perform that well in our analysis of

[3] For completeness we note here (even though it is irrelevant for the narrow scopes of the present exploratory study) that just like deforming the Galilean boosts into Lorentz boosts requires the introduction of relativity of simultaneity, we recently understood [? ? ?] that in turn deforming the Lorentz boosts into DSR-Lorentz boosts requires the introduction of relativity of locality.

Extracted from

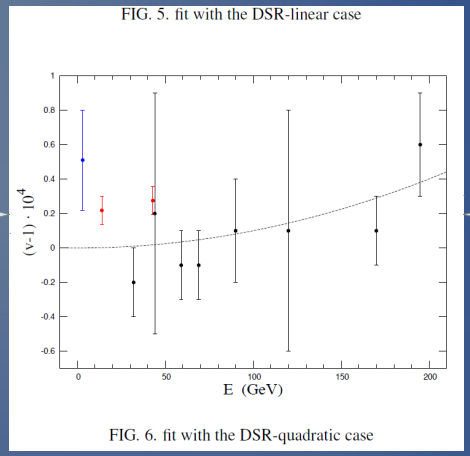
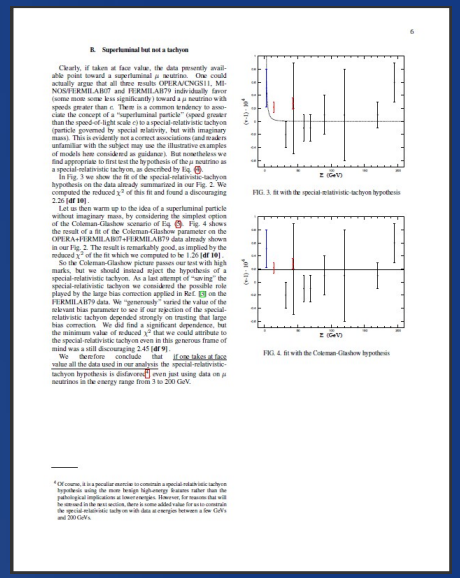


FIG. 6. fit with the DSR-quadratic case

Extracted from



B. Superluminal but not a tachyon

Clearly, if taken at face value, the data presently available point toward a superluminal μ neutrino. One could actually argue that all three results OPERA/CNGS11, MINOS/FERMILAB07 and FERMILAB79 individually favor (some more some less significantly) toward a μ neutrino with speeds greater than c. There is a common tendency to associate the concept of a "superluminal particle" (speed greater than the speed-of-light scale c) to a special-relativistic tachyon (particle governed by special relativity, but with imaginary mass). This is evidently not a correct associations (and readers unfamiliar with the subject may use the illustrative examples of models here considered as guidance). But nonetheless we find appropriate to first test the hypothesis of the μ neutrino as a special-relativistic tachyon, as described by Eq. (5).

In Fig. 3 we show the fit of the special-relativistic-tachyon hypothesis on the data already summarized in our Fig. 2. We computed the reduced χ^2 of this fit and found a discouraging 2.26 [**df 10**].

Let us then warm up to the idea of a superluminal particle without imaginary mass, by considering the simplest option of the Coleman-Glashow scenario of Eq. (6). Fig. 4 shows the result of a fit of the Coleman-Glashow parameter on the OPERA+FERMILAB07+FERMILAB79 data already shown in our Fig. 2. The result is remarkably good, as implied by the reduced χ^2 of the fit which we computed to be 1.26 [**df 10**].

So the Coleman-Glashow picture passes our test with high marks, but we should instead reject the hypothesis of a special-relativistic tachyon. As a last attempt of "saving" the special-relativistic tachyon we considered the possible role played by the large bias correction applied in Ref. [8] on the FERMILAB79 data. We "generously" varied the value of the relevant bias parameter to see if our rejection of the special-relativistic tachyon depended strongly on trusting that large bias correction. We did find a significant dependence, but the minimum value of reduced χ^2 that we could attribute to the special-relativistic tachyon even in this generous frame of mind was a still discouraging 2.45 [**df 9**].

We therefore conclude that if one takes at face value all the data used in our analysis the special-relativistic-tachyon hypothesis is disfavored[4] even just using data on μ neutrinos in the energy range from 3 to 200 GeV.

FIG. 3. fit with the special-relativistic-tachyon hypothesis

FIG. 4. fit with the Coleman-Glashow hypothesis

[4] Of course, it is a peculiar exercise to constrain a special relativistic tachyon hypothesis using the more benign high-energy features rather than the pathological implications at lower-energies. However, for reasons that will be stressed in the next section, there is some added value for us to constrain the special-relativistic tachyon with data at energies between a few GeVs and 200 GeVs.

Figure 1

Describes the same data

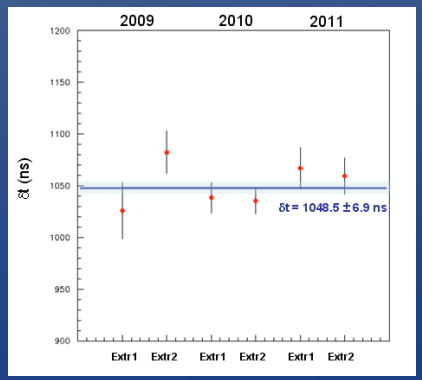
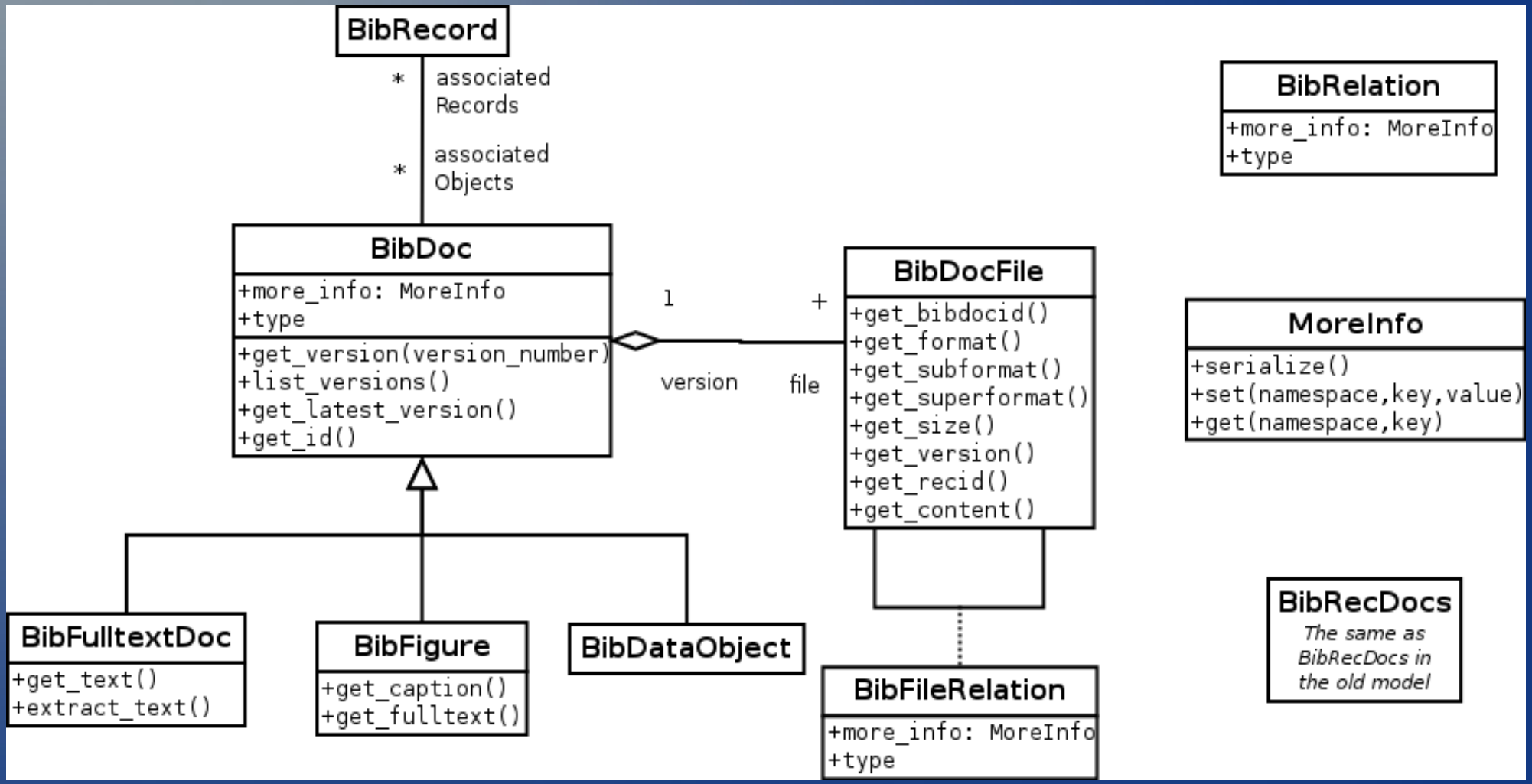
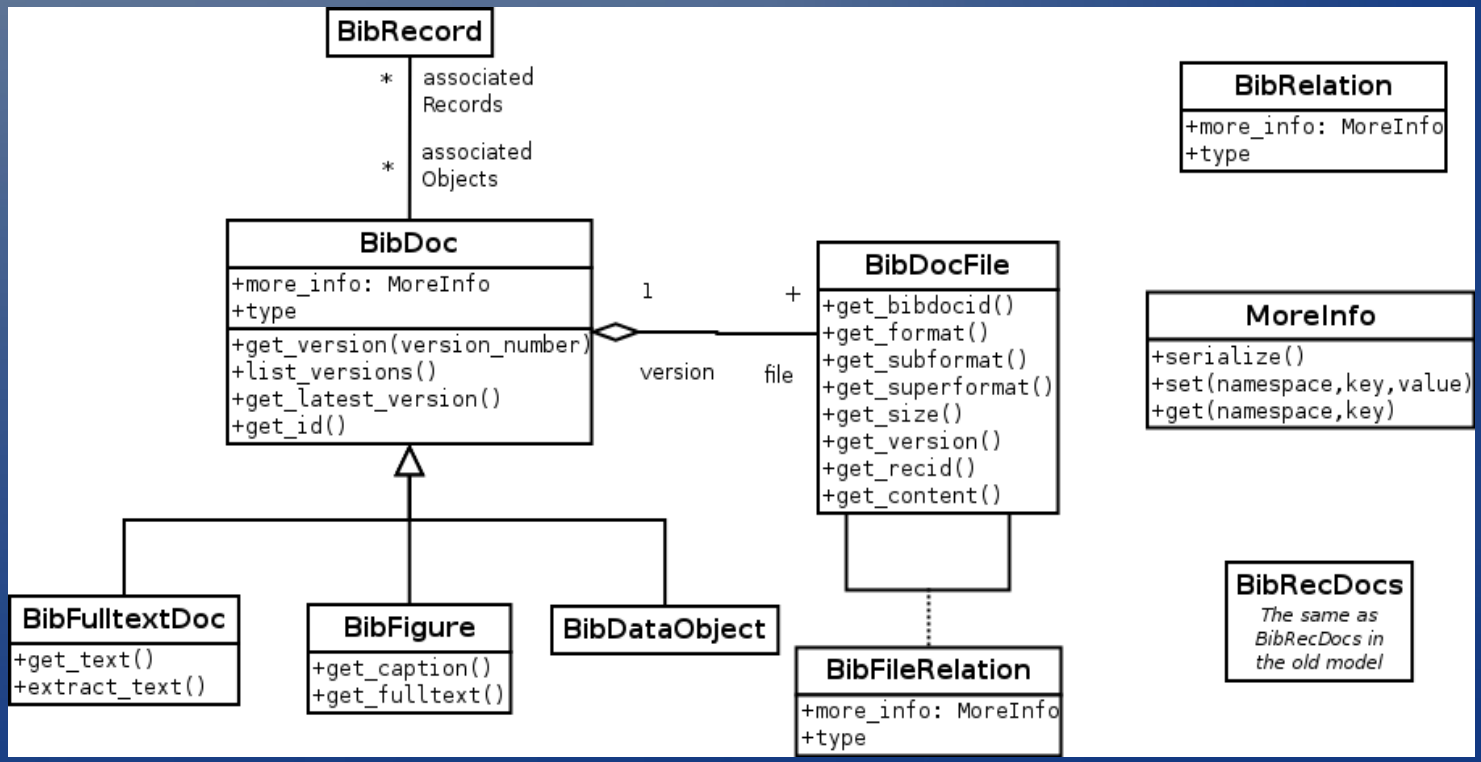


Figure 2 (extracted from different publication)


BibRecord
* associated Records
* associated Objects
BibDoc
+more_info: MoreInfo
+type
+get_version(version_number)
+list_versions()
+get_latest_version()
+get_id()
1
version
+
file
BibDocFile
+get_bibdocid()
+get_format()
+get_subformat()
+get_superformat()
+get_size()
+get_version()
+get_recid()
+get_content()
BibRelation
+more_info: MoreInfo
+type
MoreInfo
+serialize()
+set(namespace,key,value)
+get(namespace,key)
BibFulltextDoc
+get_text()
+extract_text()
BibFigure
+get_caption()
+get_fulltext()
BibDataObject
BibFileRelation
+more_info: MoreInfo
+type
BibRecDocs
The same as BibRecDocs in the old model

Data model and figures storage

Data model	Figures
BibDoc	Figure
BibDocFile	Particular encoding of a figure
BibRelation	Relation between figure and original document
BibRelation	Relation between two figures



Making documents more standalone

- Docnames cannot be identified by their names outside of the scope of a record.
  - get_name() ----> get_storagename()
- ID numbers made visible to the external world (in the FFT fields and so on)

BibRelation – link between entities

- Allows to describe dependencies and connections between different entities of the data model
- Allows specifying an arbitrary type of the relation (for example „is extracted from”, „is the same as” etc...)

MoreInfo: custom meta-data container

Namespace → key → value

- Can be attached to any entity (BibObject, BibVersion, BibFile, BibRelation)
- Persistently stores a generic dictionaries (every module has their own identified by the namespace)
- Stored with smaller granularity

BibUpload extensions

- BibUpload = local replacement of the database transactions (assures the data consistency)
- Currently documents uploaded to Invenio using artifficial FFT tag.

- Extended syntax of FFT, new tags for the new types of objects

Uploading data in new format

- New artifficlal MARC XML fields:
  - BRT (Uploading and modifying relations between documents)
  - MIT (Uploading MoreInfo fields)

  - BDR (Attaching existing objects to records)

Uploading MoreInfo

- Externally (MIT field) or internally (from within FFT/BRT)
- Values encoded in serialised Python objects

```
{
   "namespace": {
      "key": "value",
      "key2": "value2"
   }}
```

- Semantics decoupled from BibUpload modes (insert/replace/correct/...)

Temporary identifiers

- Internal Invenio identifiers are assigned during the execution of BibUpload.
- We need to be able to upload relations between BibDocs whose IDs might be not know yet.

  SOLUTION:
- Temporary identifier = identifier unique within the input MARC XML file

Example of TMP ID usage:

```
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <datafield tag="100" ind1=" " ind2=" ">
      <subfield code="a">This is a record of the publication</subfield>
    </datafield>
    <datafield tag="FFT" ind1=" " ind2=" ">
      <subfield code="a">http://somedomain.com/document.pdf</subfield>
      <subfield code="t">Main</subfield>
      <subfield code="n">docname</subfield>
      <subfield code="i">TMP:id_identifier1</subfield>
      <subfield code="v">TMP:ver_identifier1</subfield>
    </datafield>
  </record>
```

```xml
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <datafield tag="100" ind1=" " ind2=" ">
      <subfield code="a">This is a record of the publication</subfield>
    </datafield>
    <datafield tag="FFT" ind1=" " ind2=" ">
      <subfield code="a">http://somedomain.com/document.pdf</subfield>
      <subfield code="t">Main</subfield>
      <subfield code="n">docname</subfield>
      <subfield code="i">TMP:id_identifier1</subfield>
      <subfield code="v">TMP:ver_identifier1</subfield>
    </datafield>
  </record>
```

```xml
<record>
  <datafield tag="100" ind1=" " ind2=" ">
    <subfield code="a">This is a record of a dataset extracted from the publication</subfield>
  </datafield>

  <datafield tag="FFT" ind1=" " ind2=" ">
    <subfield code="a">http://sample.com/dataset.data</subfield>
    <subfield code="t">Main</subfield>
    <subfield code="n">docname2</subfielxd>
    <subfield code="i">TMP:id_identifier2</subfield>
    <subfield code="v">TMP:ver_identifier2</subfield>
  </datafield>

  <datafield tag="BRT" ind1=" " ind2=" ">
    <subfield code="i">TMP:id_identifier1</subfield>
    <subfield code="v">TMP:ver_identifier1</subfield>
    <subfield code="j">TMP:id_identifier2</subfield>
    <subfield code="w">TMP:ver_identifier2</subfield>

    <subfield code="t">is_extracted_from</subfield>
  </datafield>
</record>
```

Questions?