A Storage Model for Supporting Figures and Other Artefacts in Scientic Libraries: the Case Study of Invenio

Piotr Praczyk *(1,2)*, Javier Nogueras-Iso *(2)*, Samuele Kaplun *(1)*, Tibor Simko *(1)*

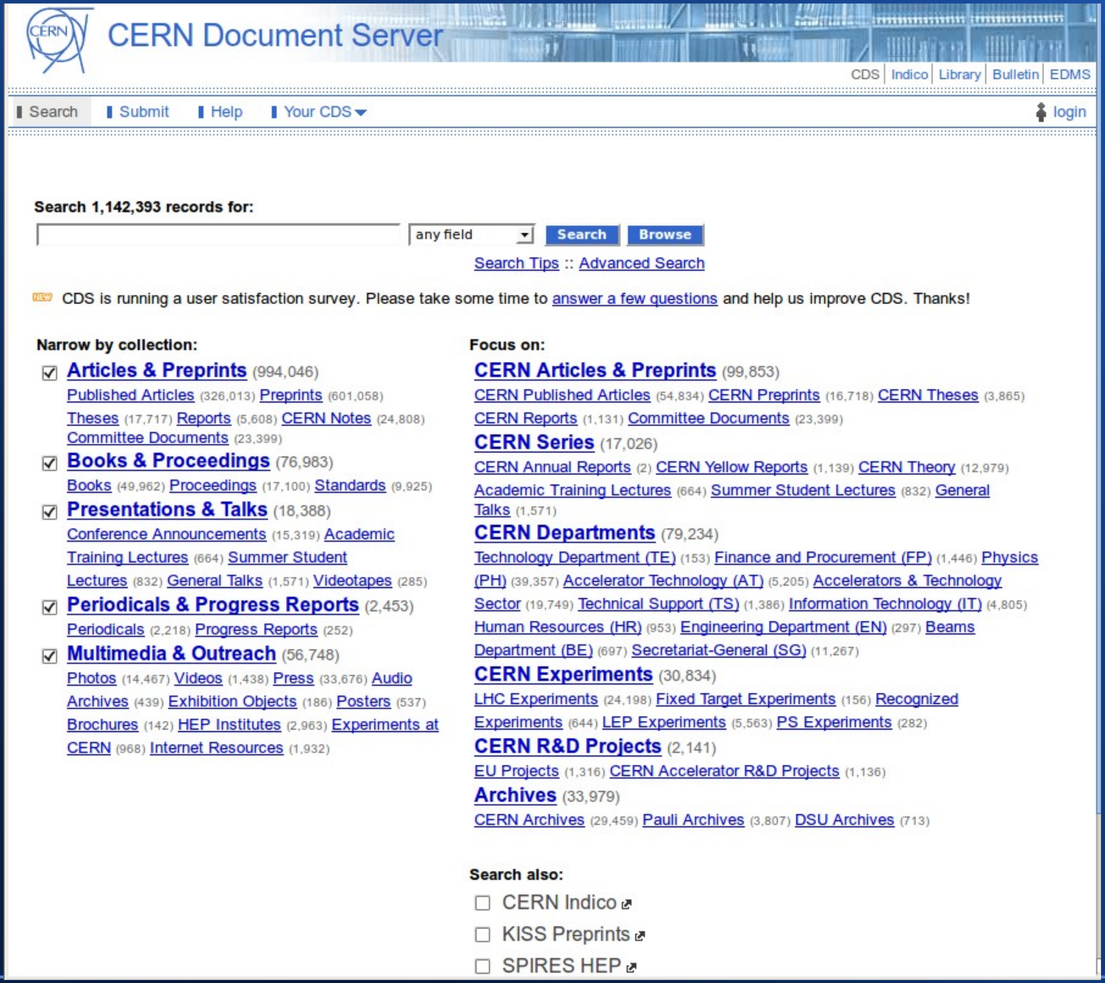
(1) CERN, Geneva, Switzerland
(2) Universidad de Zaragoza, Zaragoza, Spain

Berlin, 29.09.2011

Outline

- The Invenio and Inspire projects
- Old data model
- New use cases of storin data
    - Figures
    - Data preservation
- New data model
- Uploading data into the repository

Invenio



Search 1,142,393 records for:

Search Tips :: Advanced Search

CDS is running a user satisfaction survey. Please take some time to answer a few questions and help us improve CDS. Thanks!

Narrow by collection:

☑ **Articles & Preprints** (994,046)
Published Articles (326,013) Preprints (601,058) Theses (17,717) Reports (5,608) CERN Notes (24,808) Committee Documents (23,399)

☑ **Books & Proceedings** (76,983)
Books (49,962) Proceedings (17,100) Standards (9,925)

☑ **Presentations & Talks** (18,388)
Conference Announcements (15,319) Academic Training Lectures (664) Summer Student Lectures (832) General Talks (1,571) Videotapes (285)

☑ **Periodicals & Progress Reports** (2,453)
Periodicals (2,218) Progress Reports (252)

☑ **Multimedia & Outreach** (56,748)
Photos (14,467) Videos (1,438) Press (33,676) Audio Archives (439) Exhibition Objects (186) Posters (537) Brochures (142) HEP Institutes (2,963) Experiments at CERN (968) Internet Resources (1,932)

Focus on:

CERN Articles & Preprints (99,853)
CERN Published Articles (54,834) CERN Preprints (16,718) CERN Theses (3,865) CERN Reports (1,131) Committee Documents (23,399)

CERN Series (17,026)
CERN Annual Reports (2) CERN Yellow Reports (1,139) CERN Theory (12,979) Academic Training Lectures (664) Summer Student Lectures (832) General Talks (1,571)

CERN Departments (79,234)
Technology Department (TE) (153) Finance and Procurement (FP) (1,446) Physics (PH) (39,357) Accelerator Technology (AT) (5,205) Accelerators & Technology Sector (19,749) Technical Support (TS) (1,386) Information Technology (IT) (4,805) Human Resources (HR) (953) Engineering Department (EN) (297) Beams Department (BE) (697) Secretariat-General (SG) (11,267)

CERN Experiments (30,834)
LHC Experiments (24,198) Fixed Target Experiments (156) Recognized Experiments (644) LEP Experiments (5,563) PS Experiments (282)

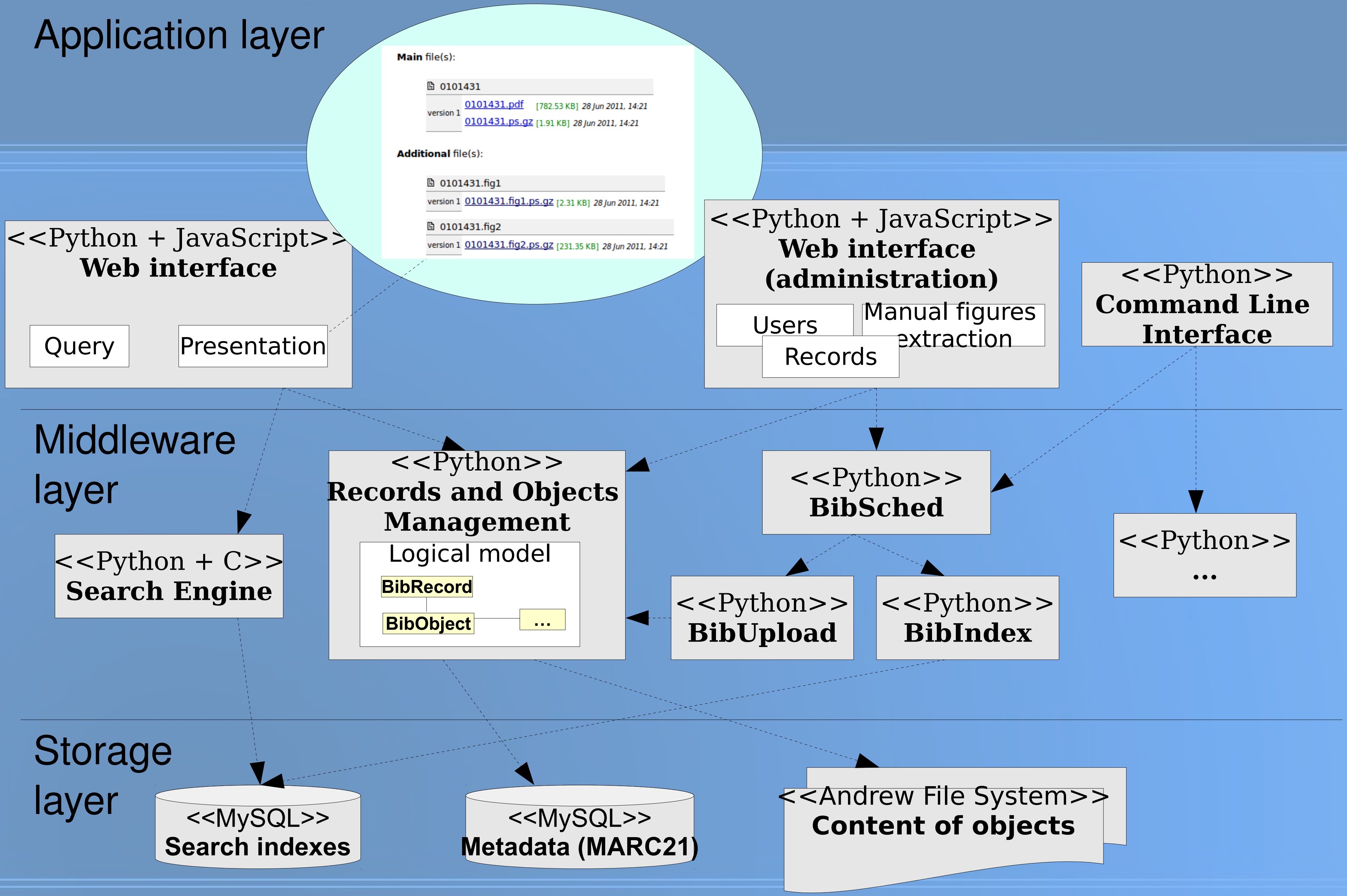
CERN R&D Projects (2,141)
EU Projects (1,316) CERN Accelerator R&D Projects (1,136)

Archives (33,979)
CERN Archives (29,459) Pauli Archives (3,807) DSU Archives (713)

Search also:

□ CERN Indico
□ KISS Preprints
□ SPIRES HEP

- Created to be a basis for CERN Document Server
- Meta-data represented in MARC



Main file(s):

0101431

version 1	0101431.pdf [782.53 KB] *28 Jun 2011, 14:21* 0101431.ps.gz [1.91 KB] *28 Jun 2011, 14:21*

Additional file(s):

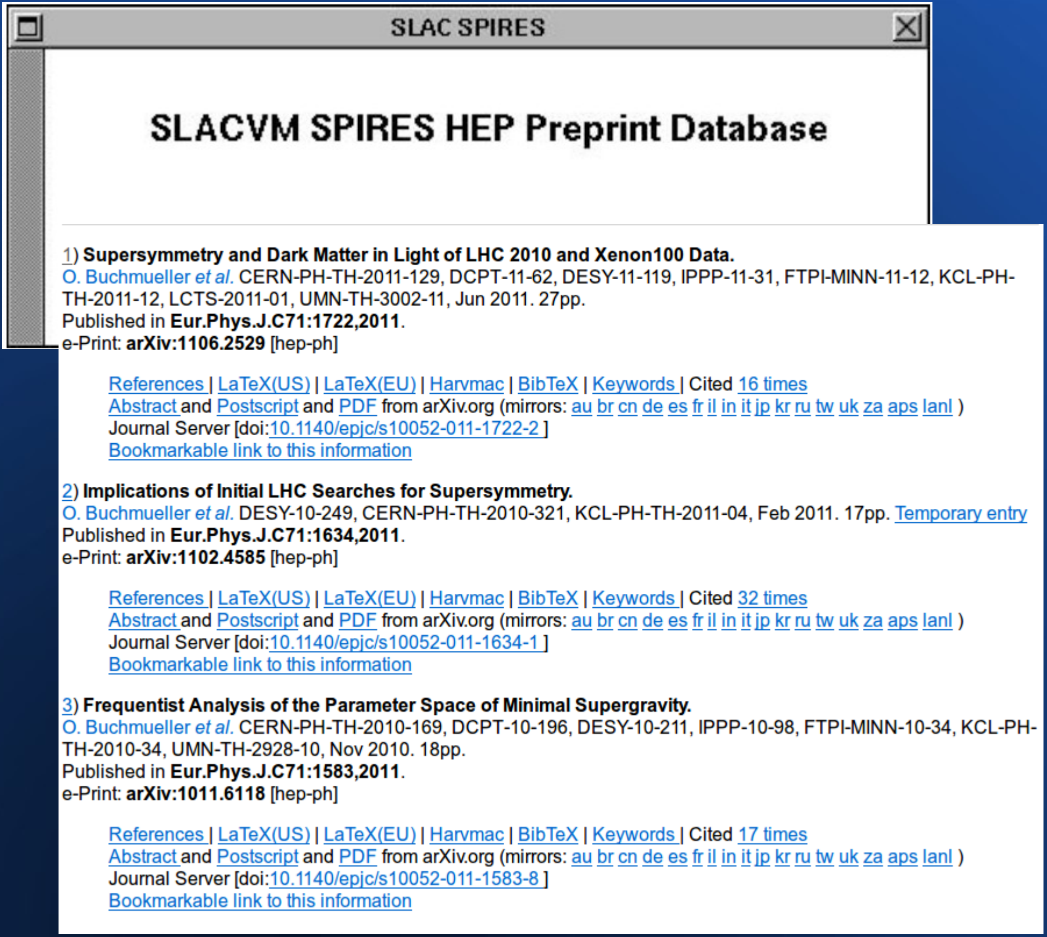
0101431.fig1

version 1	0101431.fig1.ps.gz [2.31 KB] *28 Jun 2011, 14:21*

0101431.fig2

version 1	0101431.fig2.ps.gz [231.35 KB] *28 Jun 2011, 14:21*

SPIRES



SLACVM SPIRES HEP Preprint Database

1) **Supersymmetry and Dark Matter in Light of LHC 2010 and Xenon100 Data.**
O. Buchmueller *et al.* CERN-PH-TH-2011-129, DCPT-11-62, DESY-11-119, IPPP-11-31, FTPI-MINN-11-12, KCL-PH-TH-2011-12, LCTS-2011-01, UMN-TH-3002-11, Jun 2011. 27pp.
Published in **Eur.Phys.J.C71:1722,2011**.
e-Print: **arXiv:1106.2529** [hep-ph]

References | LaTeX(US) | LaTeX(EU) | Harvmac | BibTeX | Keywords | Cited 16 times
Abstract and Postscript and PDF from arXiv.org (mirrors: au br cn de es fr il in it jp kr ru tw uk za aps lanl)
Journal Server [doi:10.1140/epjc/s10052-011-1722-2]
Bookmarkable link to this information

2) **Implications of Initial LHC Searches for Supersymmetry.**
O. Buchmueller *et al.* DESY-10-249, CERN-PH-TH-2010-321, KCL-PH-TH-2011-04, Feb 2011. 17pp. Temporary entry
Published in **Eur.Phys.J.C71:1634,2011**.
e-Print: **arXiv:1102.4585** [hep-ph]

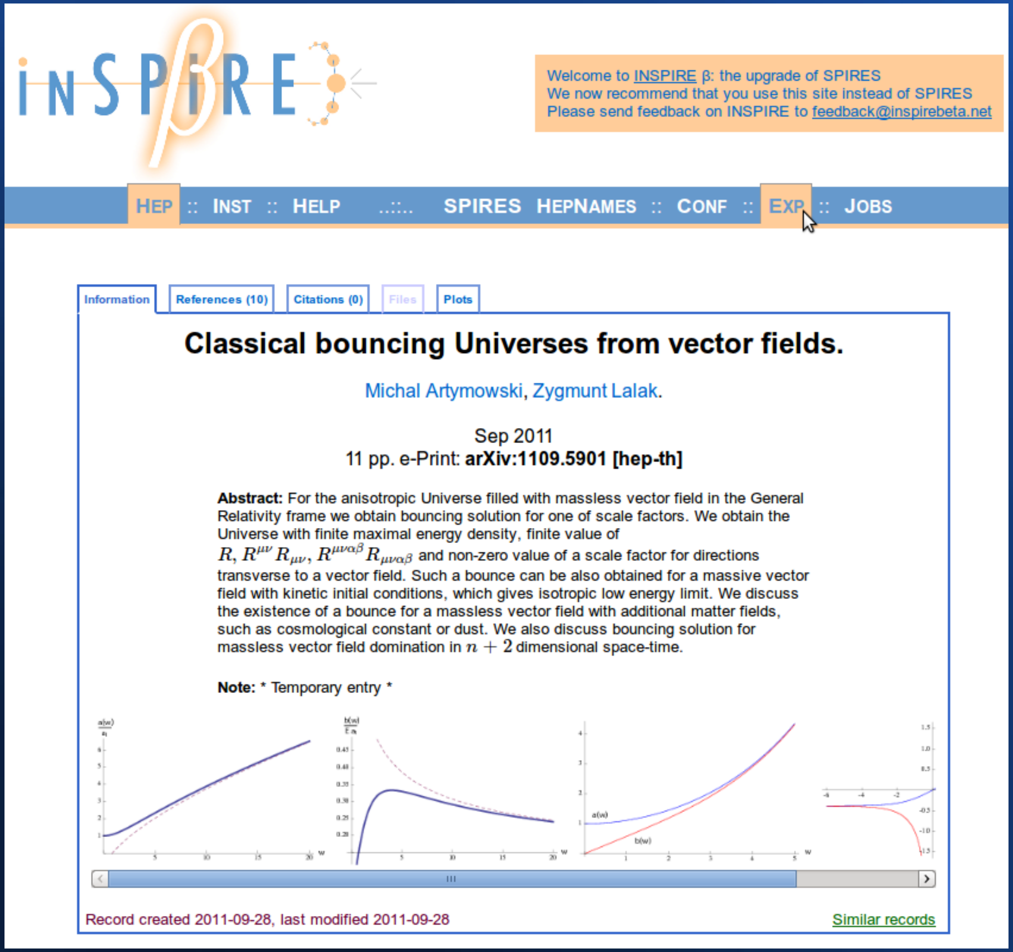
References | LaTeX(US) | LaTeX(EU) | Harvmac | BibTeX | Keywords | Cited 32 times
Abstract and Postscript and PDF from arXiv.org (mirrors: au br cn de es fr il in it jp kr ru tw uk za aps lanl)
Journal Server [doi:10.1140/epjc/s10052-011-1634-1]
Bookmarkable link to this information

3) **Frequentist Analysis of the Parameter Space of Minimal Supergravity.**
O. Buchmueller *et al.* CERN-PH-TH-2010-169, DCPT-10-196, DESY-10-211, IPPP-10-98, FTPI-MINN-10-34, KCL-PH-TH-2010-34, UMN-TH-2928-10, Nov 2010. 18pp.
Published in **Eur.Phys.J.C71:1583,2011**.
e-Print: **arXiv:1011.6118** [hep-ph]

References | LaTeX(US) | LaTeX(EU) | Harvmac | BibTeX | Keywords | Cited 17 times
Abstract and Postscript and PDF from arXiv.org (mirrors: au br cn de es fr il in it jp kr ru tw uk za aps lanl)
Journal Server [doi:10.1140/epjc/s10052-011-1583-8]
Bookmarkable link to this information

- Database of preprints started in 70s
- In 90s the first WEB page in USA
- Very difficult to maintain, extremely slow

Invenio + SPIRES = INSPIRE



Welcome to INSPIRE β: the upgrade of SPIRES
We now recommend that you use this site instead of SPIRES
Please send feedback on INSPIRE to feedback@inspirebeta.net

Classical bouncing Universes from vector fields.

Michal Artymowski, Zygmunt Lalak.

Sep 2011
11 pp. e-Print: **arXiv:1109.5901 [hep-th]**

Abstract: For the anisotropic Universe filled with massless vector field in the General Relativity frame we obtain bouncing solution for one of scale factors. We obtain the Universe with finite maximal energy density, finite value of $R, R^{\mu\nu}R_{\mu\nu}, R^{\mu\nu\alpha\beta}R_{\mu\nu\alpha\beta}$ and non-zero value of a scale factor for directions transverse to a vector field. Such a bounce can be also obtained for a massive vector field with kinetic initial conditions, which gives isotropic low energy limit. We discuss the existence of a bounce for a massless vector field with additional matter fields, such as cosmological constant or dust. We also discuss bouncing solution for massless vector field domination in $n+2$ dimensional space-time.

Note: * Temporary entry *

Record created 2011-09-28, last modified 2011-09-28

- Large community of users
- Multiple sources of data (SPIRES, arXiv, diect submissions, publishers)
- Nearly 1000000 records

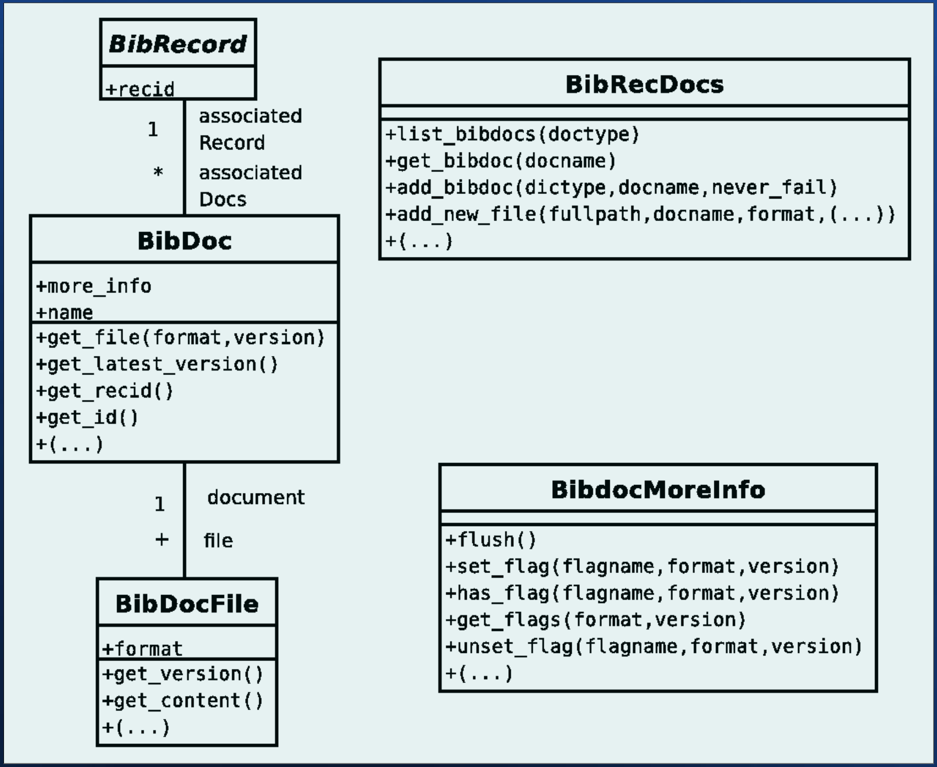
Invenio/INSPIRE

- Invenio – digital library software developed at CERN to manage the repository of documents created in the institution

- SPIRES – The digital library of preprints created at SLAC.

- **In**venio + **SPIRE**S = INSPIRE

Non-bibliographical data in Invenio



BibRecord

+recid

1 associated Record

* associated Docs

BibDoc

+more_info
+name

+get_file(format,version)
+get_latest_version()
+get_recid()
+get_id()
+(...)

1 document

+ file

BibDocFile

+format

+get_version()
+get_content()
+(...)

BibRecDocs

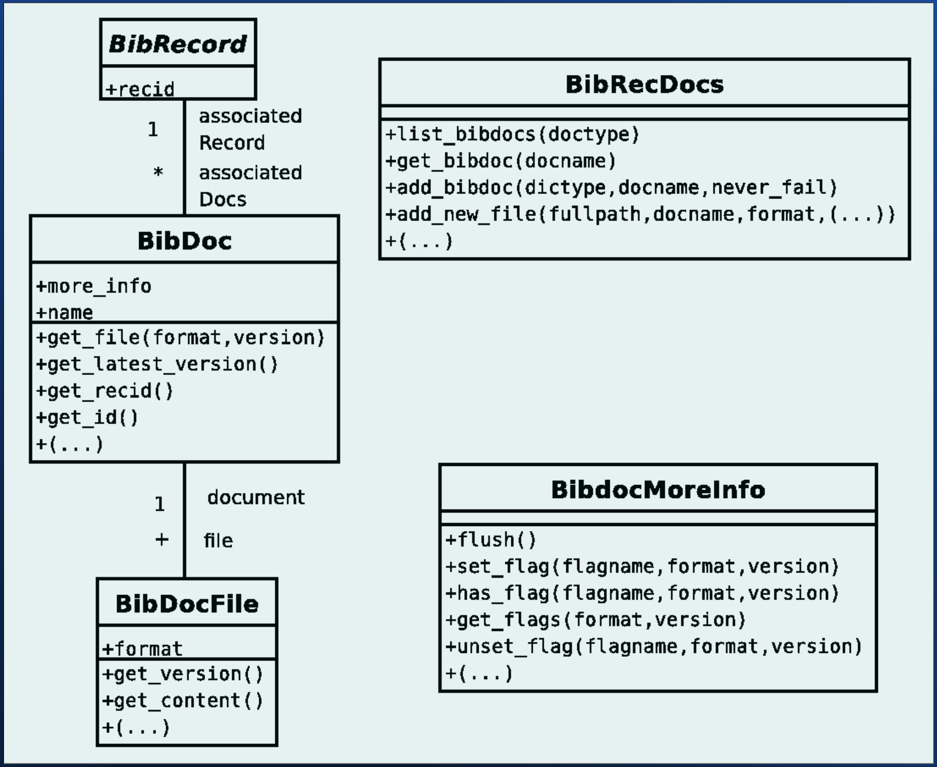
+list_bibdocs(doctype)
+get_bibdoc(docname)
+add_bibdoc(dictype,docname,never_fail)
+add_new_file(fullpath,docname,format,(...))
+(...)

BibdocMoreInfo

+flush()
+set_flag(flagname,format,version)
+has_flag(flagname,format,version)
+get_flags(format,version)
+unset_flag(flagname,format,version)
+(...)

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

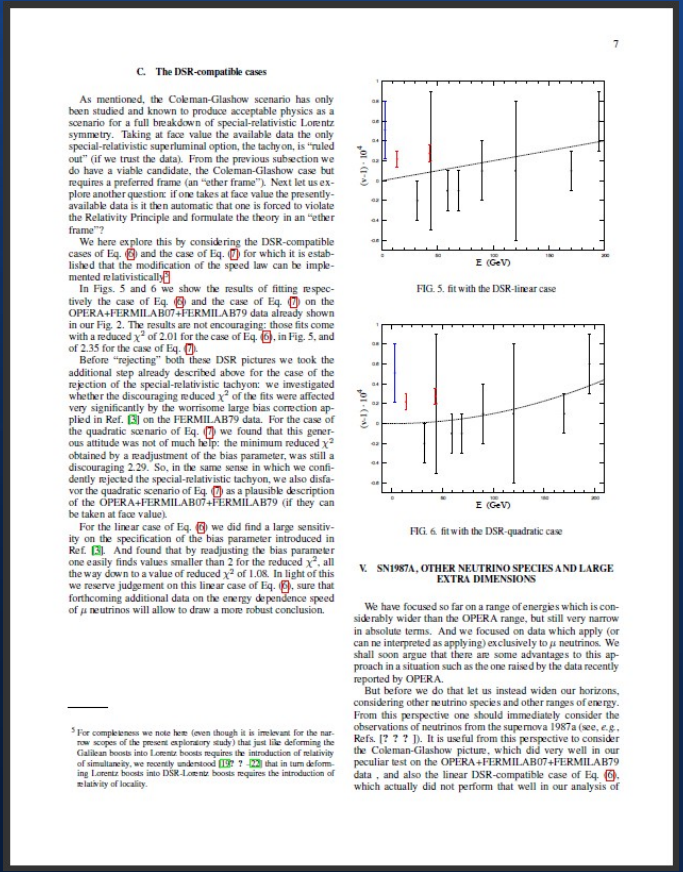
Non-bibliographical data in Invenio



```
BibRecord
+recid

1   associated Record
*   associated Docs

BibDoc
+more_info
+name
+get_file(format,version)
+get_latest_version()
+get_recid()
+get_id()
+(...)

1   document
+   file

BibDocFile
+format
+get_version()
+get_content()
+(...)

BibRecDocs
+list_bibdocs(doctype)
+get_bibdoc(docname)
+add_bibdoc(dictype,docname,never_fail)
+add_new_file(fullpath,docname,format,(...))
+(...)

BibdocMoreInfo
+flush()
+set_flag(flagname,format,version)
+has_flag(flagname,format,version)
+get_flags(format,version)
+unset_flag(flagname,format,version)
+(...)
```

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

Figures from scientific publications



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[5]

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. [3] 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. [3]. 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 ne 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. [? ? ?]). 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

[5] 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 [19? ? -22] that in turn deforming Lorentz boosts into DSR-Lorentz boosts requires the introduction of relativity of locality.

Automatic extraction

Manual extraction

FIG. 5. fit with the DSR-linear case

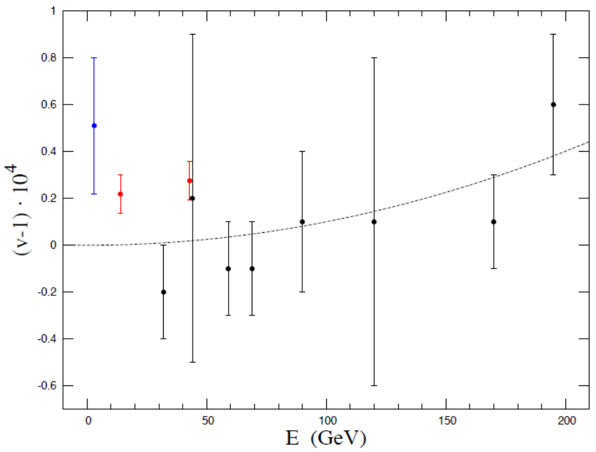
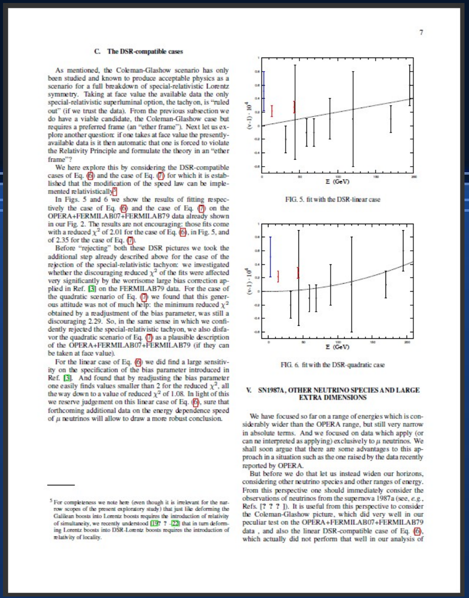


FIG. 6. fit with the DSR-quadratic case



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[5]

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. [3] 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. [3]. 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 ne 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. [? ? ?]). 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

[5] 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 [19? ? -22] that in turn deforming Lorentz boosts into DSR-Lorentz boosts requires the introduction of relativity of locality.

extraction

FIG. 5. fit with the DSR-linear case

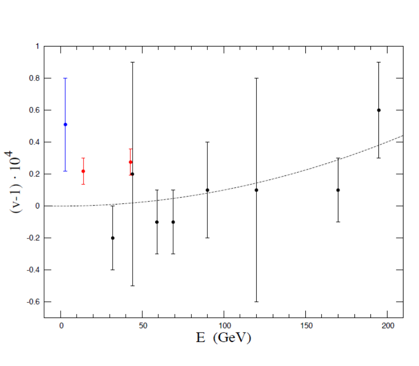
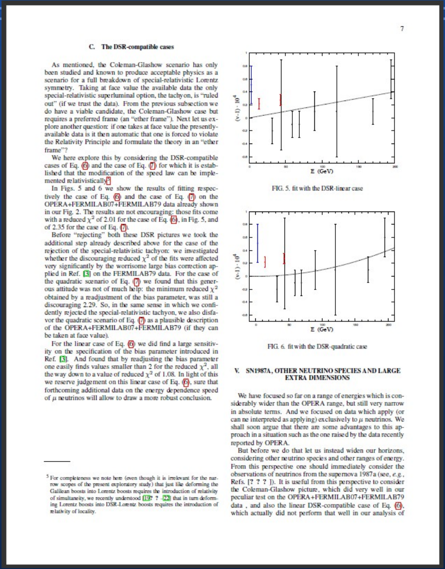


FIG. 6. fit with the DSR-quadratic case

- Figures should exist independently from articles

- Some type of meta data should not be presented to users directly

Figures from scientific publications



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[5]

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. [3] 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. [3]. 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 ne 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. [? ? ?]). 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

[5] 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 [19? ? –22] that in turn deforming Lorentz boosts into DSR-Lorentz boosts requires the introduction of relativity of locality.

Extracted from

FIG. 5. fit with the DSR-linear case

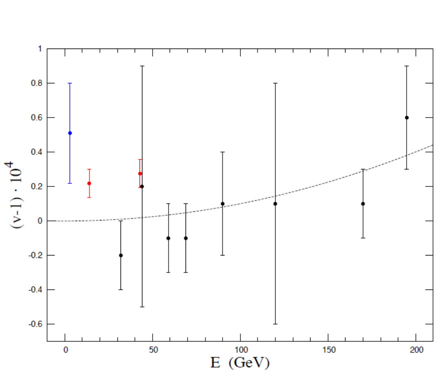
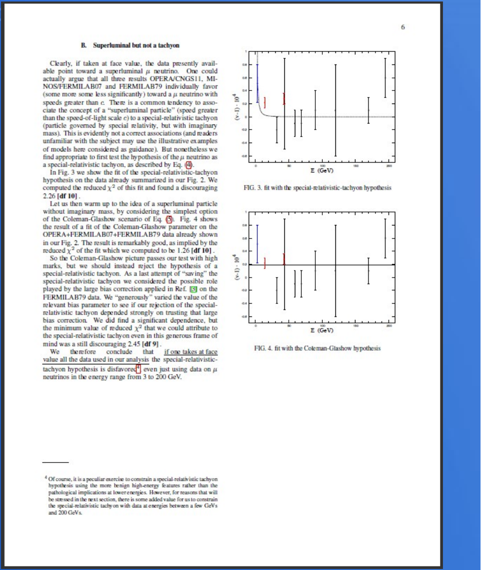


FIG. 6. fit with the DSR-quadratic case

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. (4).

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. (5). 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. [3] 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

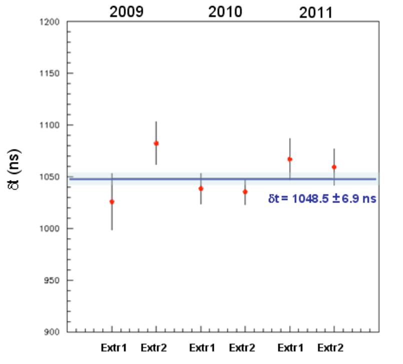


Figure 2 (extracted from different publication)

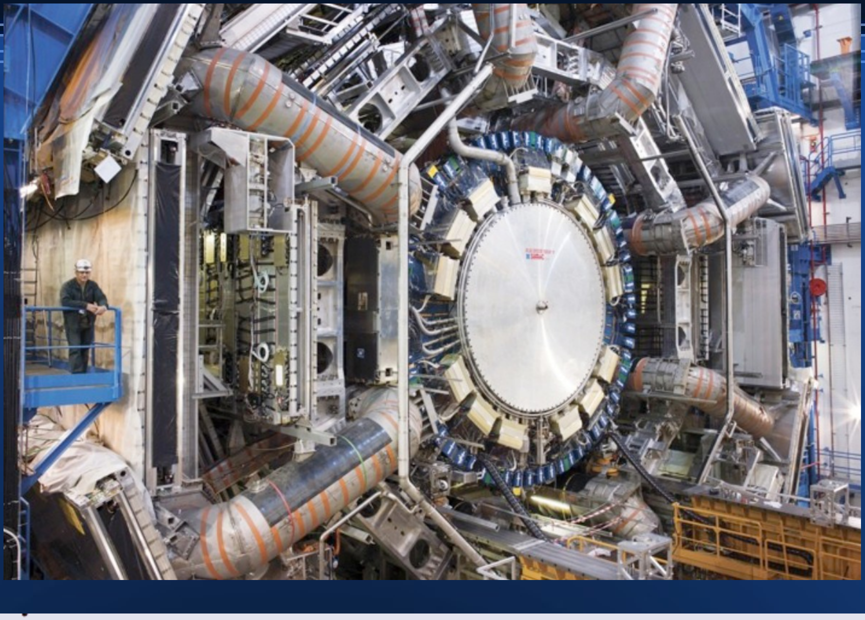
Data-model requirements

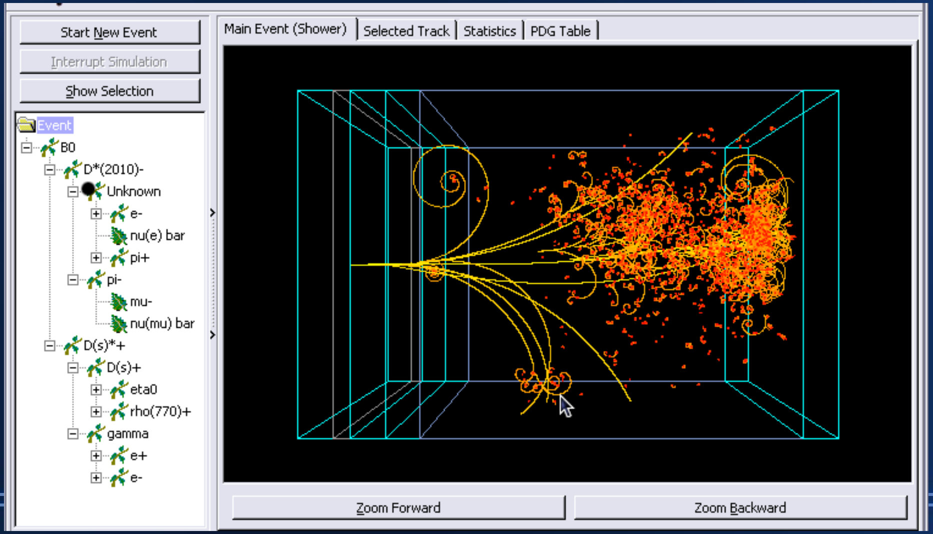
- Different types of relations between figures (illustrates the same data, is subfigure of...)
- Relation of being extracted from a document
- Meta-data of figures, different versions of figure, relations between figures, links between figure and document it is extracted from
- Storage of more complicated data-types

Examples of meta-data associated with different entities

Figure *(the most general meta-data)*	**Figure version** *(Appearance related meta-data)*	**Figure storage format**	**Relation between figure and document**	**Relation between two figures**
Type of figure	Semantics of a figure	Access permissions	Position of a figure within the original document	Type of relation
Quantities presented on axis (in the case of plots)	Data extracted from figure		Caption of a figure within a particular document	<additional type-dependent fields>
Units and scales of axis (in the case of plots)			References from within text	
...			Figure identifier within a document	

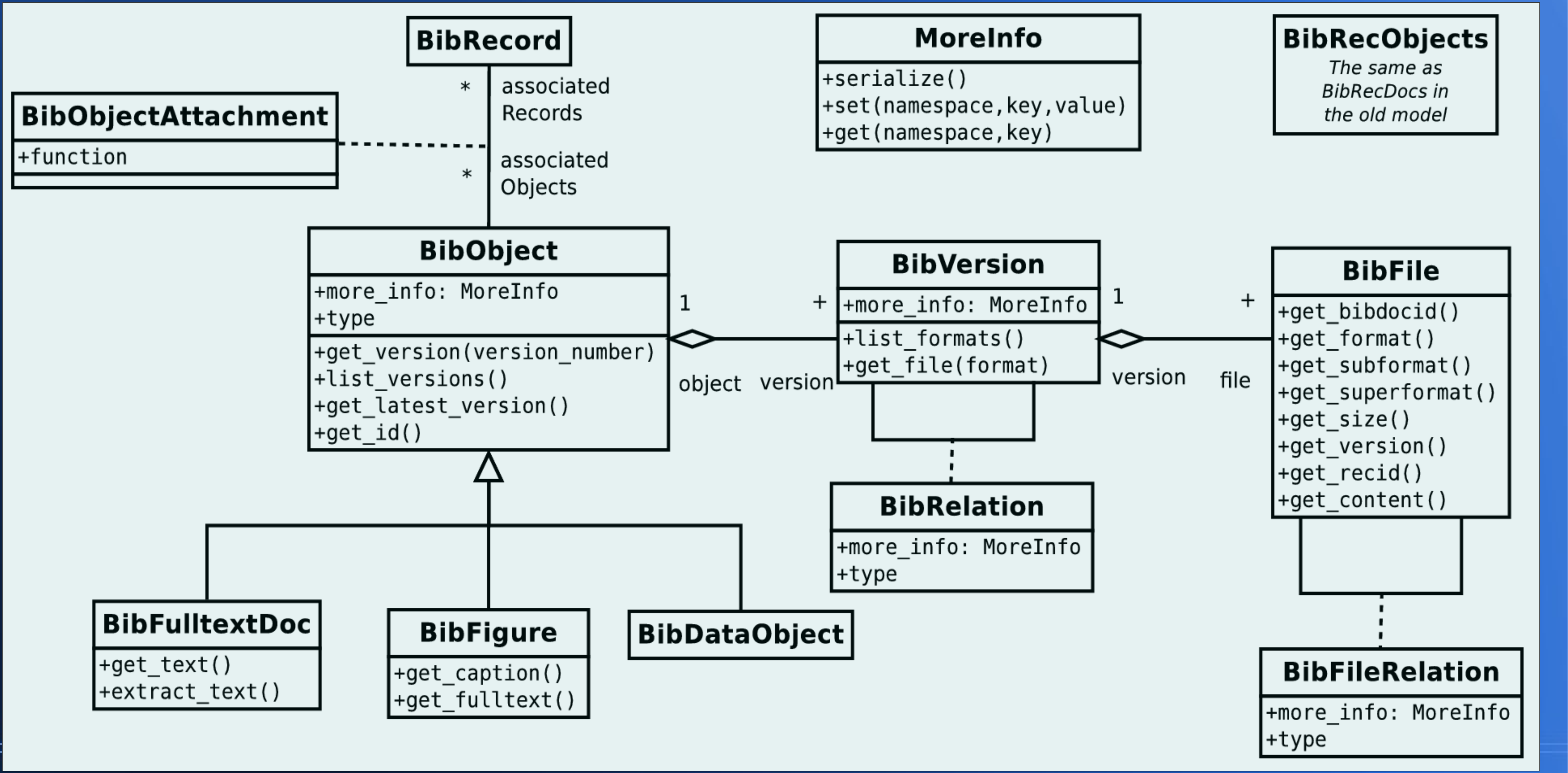
HEP data preservation





- Storing raw data
- Storing intermediate analysis
- Storing additional documentation
- Assigning Digital Object Identifiers

New architecture for storing non-bibliographical objects



BibObject

- Abstract representation of a document (not
- Document-type specific functionalities are implemented by subclasses (defined by modules of Invenio and loaded dynamically)
- Identified by a globally-unique identifier and by a name unique in the scope of a bibliographical record

BibVersion, BibFile

- Represent increasing specialisation of a document
- BibVersion represents a particular revision of an object (corresponding for example to correction of mistakes)
- BibFile describes a particular encoding of a version of an object (encapsulates the real file, remembers the format of a file)

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...)

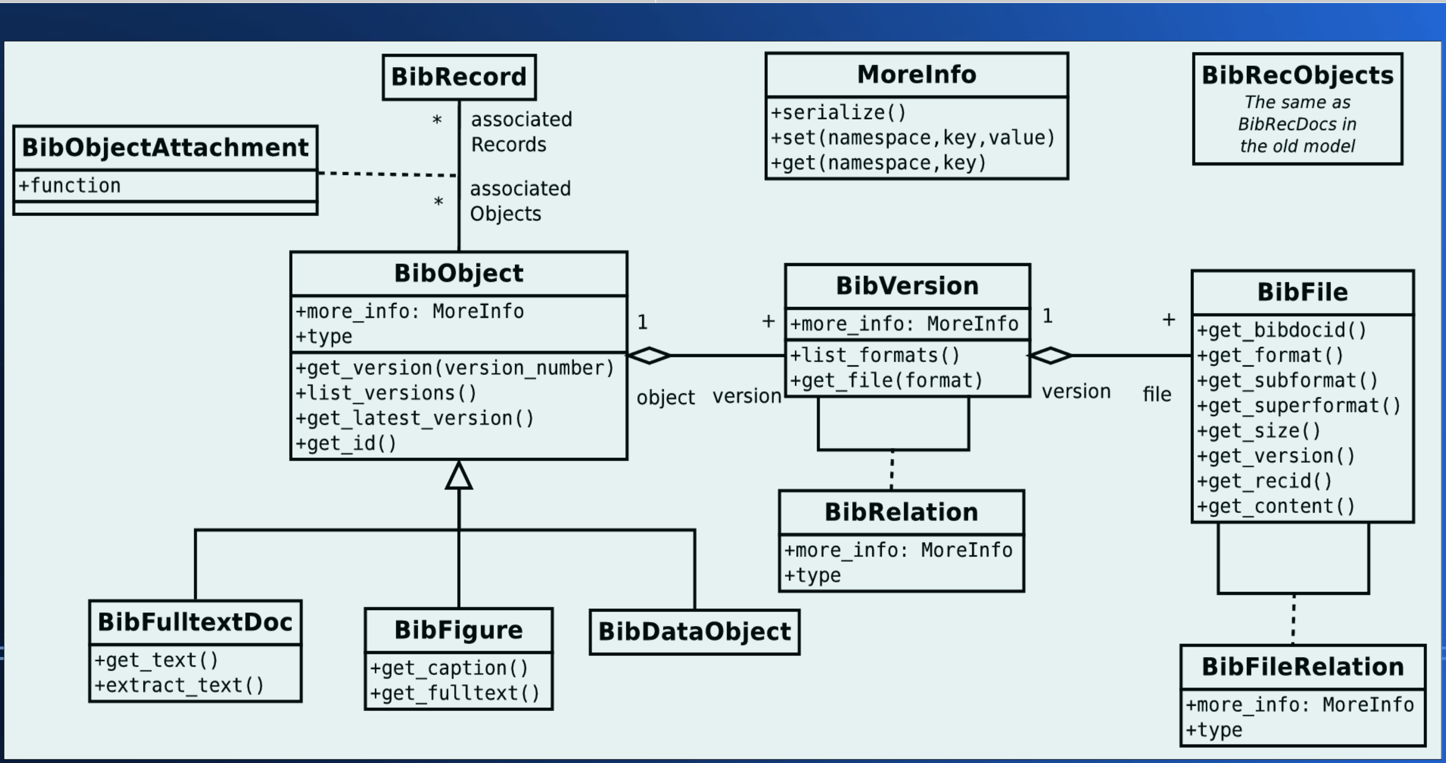
MoreInfo: custom meta-data container

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)

Data model and figures storage

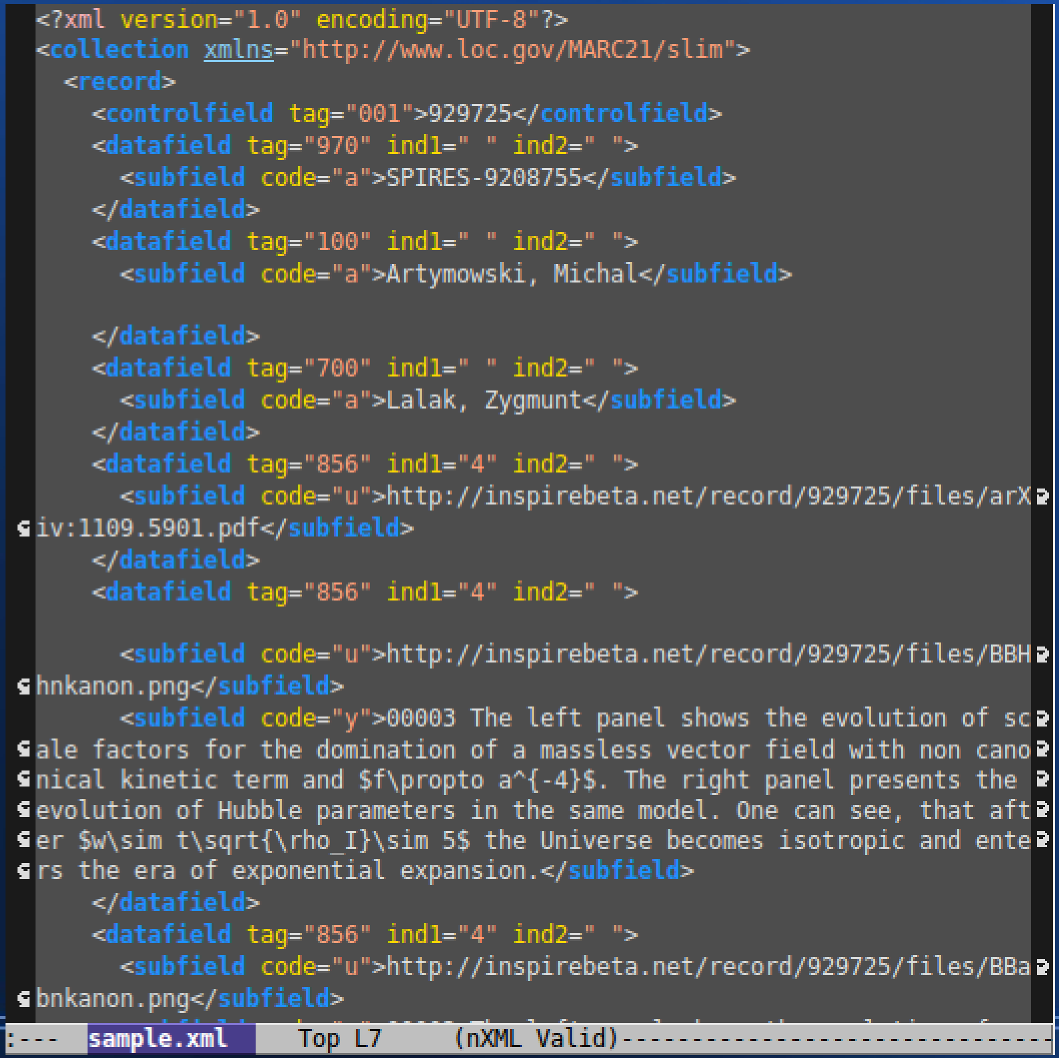
Data model	Figures
BibObject	Figure
BibObjectVersion	Figure version
BibFile	Particular encoding of a figure
BibRelation	Relation between figure and original document
BibRelation	Relation between two figures



Comparison between data models

	Old model	New model
Attaching document to a bibliographic record	YES	YES
Attaching the same document to many records	NO	YES
Storing custom data keys in MoreInfo dictionaries	NO	YES
Creating documents not attached to any records	NO	YES

Uploading data to Invenio

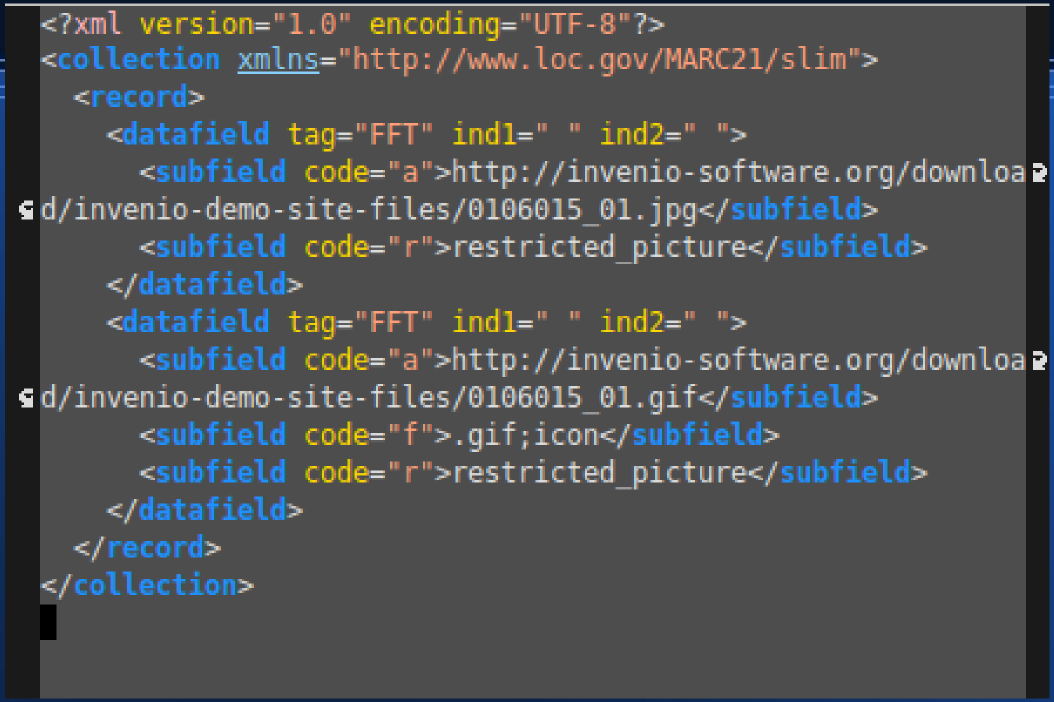


```xml
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <controlfield tag="001">929725</controlfield>
    <datafield tag="970" ind1=" " ind2=" ">
      <subfield code="a">SPIRES-9208755</subfield>
    </datafield>
    <datafield tag="100" ind1=" " ind2=" ">
      <subfield code="a">Artymowski, Michal</subfield>

    </datafield>
    <datafield tag="700" ind1=" " ind2=" ">
      <subfield code="a">Lalak, Zygmunt</subfield>
    </datafield>
    <datafield tag="856" ind1="4" ind2=" ">
      <subfield code="u">http://inspirebeta.net/record/929725/files/arX
iv:1109.5901.pdf</subfield>
    </datafield>
    <datafield tag="856" ind1="4" ind2=" ">

      <subfield code="u">http://inspirebeta.net/record/929725/files/BBH
hnkanon.png</subfield>
      <subfield code="y">00003 The left panel shows the evolution of sc
ale factors for the domination of a massless vector field with non cano
nical kinetic term and $f\propto a^{-4}$. The right panel presents the 
evolution of Hubble parameters in the same model. One can see, that aft
er $w\sim t\sqrt{\rho_I}\sim 5$ the Universe becomes isotropic and ente
rs the era of exponential expansion.</subfield>
    </datafield>
    <datafield tag="856" ind1="4" ind2=" ">
      <subfield code="u">http://inspirebeta.net/record/929725/files/BBa
bnkanon.png</subfield>
```

:--- sample.xml   Top L7   (nXML Valid)--------------------------------

- New record is encoded in MARC XML
- BibUpload is executed adding uploading task to the BibSched queue
- BibSched uploads data to the main database

FFT = Fulltext File Transfer



```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <datafield tag="FFT" ind1=" " ind2=" ">
      <subfield code="a">http://invenio-software.org/downloa
d/invenio-demo-site-files/0106015_01.jpg</subfield>
      <subfield code="r">restricted_picture</subfield>
    </datafield>
    <datafield tag="FFT" ind1=" " ind2=" ">
      <subfield code="a">http://invenio-software.org/downloa
d/invenio-demo-site-files/0106015_01.gif</subfield>
      <subfield code="f">.gif;icon</subfield>
      <subfield code="r">restricted_picture</subfield>
    </datafield>
  </record>
</collection>
```

subfield	explanation
a	URL of the file to upload
t	Function of the document within the record
f	Format of the file
...	

- Artificial, interpreted and removed during the BibUpload phase
- One entry represents one file
- Enforces documents to be attached to a record

Uploading data in new format

- New artifficial MARC XML fields:
  - BRT (Uploading and modifying relations between documents)
  - MIT (Uploading MoreInfo fields)
  - BDR (Attaching existing objects to records)

- Open for extension supporting METS

Uploading MoreInfo

- Externally (MIT field) or internally (from within FFT/BRT)

- Values encoded in Json or serialised Python objects

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

Conclusions & outlook

- The proposed a model is a flexible approach that facilitates the support of custom objects (figures, data files, software, ...)
    - Based on it, new applications for searching and accessing digital objects can be developed
- Issues to address in the future
    - Integration of extended objects within the INVENIO platform
        - Search and display information about custom objects
    - Need of assigning Digital Object Identiers (DOI) to stored data objects
        - To store the persistent state of a data object (management of versions)

Thank you !

http://invenio-software.org/
http://www.projecthepinspire.net/

piotr.praczyk@cern.ch