

# Invenio and its Potential Role in DPHEP

Tibor Šimko

`<tibor.simko@cern.ch>`

Department of Information Technology  
CERN

DASPOS/DPHEP7 Workshop  
CERN, March 21–22, 2013

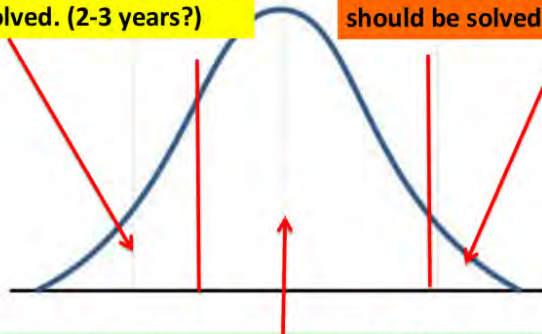
- 1 Introduction
- 2 Technology
- 3 Community
- 4 Projects
- 5 Conclusions

- 1 Introduction
- 2 Technology
- 3 Community
- 4 Projects
- 5 Conclusions

## Where to Invest – Summary (WLCG OB)

Tools and Services, e.g. Invenio:  
could be solved. (2-3 years?)

Archival Storage Functionality:  
should be solved. (i.e. "now")




Support to the Experiments for DPHEP Level 4:  
*must be solved – but how?*

# What is Invenio?

- modular **digital library** system
  - *“library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers”*
- free software, GNU General Public License
- mature, first release in 2002
- originated at CERN, sister project Indico
- nowadays co-developed by CERN, CfA, DESY, EPFL, SLAC. . .
- handles articles, books, notes, photos, videos. . .
- used by:
  - (a) institutional document repositories, e.g. CDS
    - pre-publication workflows e.g. approval
  - (b) subject-based information systems, e.g. INSPIRE
    - worldwide data e.g. citation analysis
  - (c) libraries and library networks, e.g. RERO

# Example: Invenio Demo Site

INVENIO  1 Search Personalize Administration Help admin

Search 109 records for: ellis Search

Any Collection  
Articles & Preprints (12)

Any Author  
Ellis, J (4)  
Bohrer, A (2)  
Ellis, R S (2)  
Markou, C (2)  
Abbarneo, D (1)  
More ...

Any Year  
2001 (3)  
1982 (2)  
1985 (2)  
2002 (2)  
1987 (1)  
More ...

Showing records 1 to 10 out of 12 results. Display Sort by

- 1 **Constraints on  $\Omega_{\Lambda}$  and  $\Omega_m$  from Distant Type 1a Supernovae and Cosmic Microwave Background Anisotropies** / Elstathiou, G P ; Lasenby, A N ; Hobson, M P ; Ellis, R S ; *et al* [astro-ph/9812226]  
We perform a combined likelihood analysis of the latest cosmic microwave background anisotropy data and distant Type 1a Supernova data of Perlmutter et al (1998a). [...]  
Published in **Mon. Not. R. Astron. Soc.:**  
**Fulltext:** [9812226](#) - [PDF](#); [9812226.fig1](#) - [PS.GZ](#); [9812226.fig3](#) - [PS.GZ](#); [9812226.fig5](#) - [PS.GZ](#); [9812226.fig6](#) - [PS.GZ](#); [9812226.fig7](#) - [PS.GZ](#);  
Detailed record - Similar records
- 2 **Grand unification with large supersymmetry breaking** / Ellis, J ; Ibanez, L E ; Ross, G G [PRE-25553]  
[RL-82-024]  
Detailed record - Similar records
- 3 **Search for R-Parity Violating Production of Single Sneutrinos in  $e^+e^-$  Collisions at  $\sqrt{s} = 189\text{--}209$  GeV** / Helster, A ; Schael, S ; Barate, R ; Bruneliere, R ; *et al* [CERN-EP-2001-094] [hep-ex/0201013]  
A search for single sneutrino production under the assumption that R-parity is violated via a single dominant  $\tilde{L}\bar{L}E$  coupling is presented. [...]  
Published in **Eur. Phys. J., C:**  
**Fulltext:** [PDF](#) [PS.GZ](#);  
Detailed record - Similar records

## CERN Document Server

Access articles, reports and multimedia content in HEP

Search

Submit

Help

Personalize

Search 1,262,525 records for:

Search

Search Tips

Advanced Search 

### Articles & Preprints (1,086,540)

Published Articles (334,677) Preprints (671,836)

Theses (18,873) Reports (6,913)

CERN Notes (36,649)

Committee Documents (24,269)

### Books & Proceedings (100,511)

Books (71,080) Proceedings (17,639)

Standards (11,639) Design Reports (171)

### CERN Articles & Preprints (105,016)

CERN Published Articles (57,284) CERN Preprints (17,706) CERN Theses (4,690)

CERN Reports (1,180) Committee Documents (24,269)

### CERN Series (18,729)

CERN Annual Reports (115) CERN Yellow Reports (1,150) CERN Theory (13,390)

Academic Training Lectures (706) Summer Student Lectures (951) General Talks (2,577)

### CERN Departments (95,148)

## CERN Document Server

Search

Submit

Help

Personalize

Home > CERN Experiments > LHC Experiments > ATLAS > ATLAS Internal

## ATLAS Internal

Search 10,734 records for:

Search

Search Tips

Advanced Search 

### Narrow by collection:

- ATLAS Communications** (8,639) [restricted]  
ATLAS Communications Physics (5,985) ATLAS Communications General (2,691)
- ATLAS Internal Notes** (786) [restricted]  
ATLAS Internal Notes Physics (548) ATLAS Internal Notes General (242)
- ATLAS Publication Drafts** (18) [restricted]



# Example: INSPIRE 1/2



Welcome to [INSPIRE](#), the High Energy Physics information system. Please direct questions, comments or concerns to [feedback@inspirehep.net](mailto:feedback@inspirehep.net).

HEP :: HEPNAMES :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

## HEP Search

### High-Energy Physics Literature Database

Use "find " for SPIRES-style search ([other tips](#))

Brief format Search [Easy Search](#) [Advanced Search](#)

[find | "Phys.Rev.Lett.,105"](#) :: [more](#)

#### HOW TO SEARCH

SPIRES syntax is (mostly) supported (requires "find")

find a richter, b and t quark and date > 1984  
find j phys.rev.,D50,1140 or jjhep,0903,112  
find eprint arxiv:1007.5048 (Note the plots available on the detailed record)  
find fulltext "quark-gluon plasma" (Note new "fulltext" operator)  
find a ellis and refersto a witten (Note "refersto")  
find a kane and citedby title SUSY and topcite 200+ (Note "citedby")

New techniques:

1985 richter quark multiplicity  
arXiv:1007.5048  
citedby:author:ellis -refersto:author:witten  
author:randall | author:sundrum cited:450->1350

Additional Help:

[More search tips](#) and [full help](#)

#### HEP

[Search Tips](#)  
[INSPIRE Help](#)  
[Corrections](#)  
[Additions](#)  
[Email Us](#)

#### INSPIRE

[About INSPIRE](#)  
[Recent topcites](#)  
[HEP Reviews](#)  
[SPIRES-HEP \[?\]](#)

#### RESOURCES

[ADS](#)  
[arXiv](#)  
[HepData](#)  
[PDG](#)

#### INSPIRE NEWS

2013-03-14 Use INSPIRE to generate your TeX bibliography, see [http://inspirehep.net](#)

# Example: INSPIRE 2/2

Co-cited with: 22512 records

(1264) [Neutrino Mass and Spontaneous Parity Violation](#) - Mohapatra, Rabindra N. *et al.* Phys.Rev.Lett. 44 (1980) 912 MDDP-TR-80-060, MDDP-PP-80-105, CCNY-HEP-79-10

(568) [Neutrino Masses in SU\(2\) x U\(1\) Theories](#) - Schechter, J. *et al.* Phys.Rev. D22 (1980) 2227 SU-4217-167, COO-3533-167

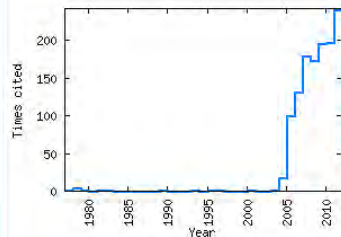
(538) [Baryogenesis Without Grand Unification](#) - Fukugita, M. *et al.* Phys.Lett. B174 (1986) 45 RIFP-641

(462) [Neutrino Masses and Mixings in Gauge Models with Spontaneous Parity Violation](#) - Mohapatra, Rabindra N. *et al.* Phys.Rev. D23 (1981) 165 FERMILAB-PUB-80-061-THY, FERMILAB-PUB-80-061-T

(418) [Proton Lifetime and Fermion Masses in an SO\(10\) Model](#) - Lazarides, George *et al.* Nucl.Phys. B181 (1981) 287 FREIBURG-THEP-80-2

[more](#)

Citation history:



P.Minkowski, Phys.Lett. B67 (1977) 421

### Search:

any field ▾

[Search Tips](#) :: [Advanced Search](#)[Ask a librarian](#)  **e-docs only:** 

### Focus on:



Migrant workers



Disability in the  
workplace



HIV/AIDS



Youth  
employment



Fundamental  
principles and  
rights at work

### Narrow by collection:

- ILO publications** (82,410)
  - ILO books (43,441)
  - ILO journal articles (14,182)
  - ILO working papers (2,401)
  - ILO training manuals (4,381)
- Other publications** (369,360)

### Latest publications by ILO (86)

- ILO meetings** (26,848)
  - ▶ International Labour Conference (7,227)
  - ▶ Governing Body (19,680)

**ILO Conventions: background and preparatory work** (2,902)



YOU ARE ▾

BY SCHOOL ▾

ABOUT EPFL ▾

EPFL • *InfoScience*

## INFOSCIENCE

Search Browse Deposit / Publish Help




RSS


## Search publications

Any field ▾ 

▶ *Advanced Search*

*Search Tips*

 [\*Access to the 96,816 scientific publications\*](#)

 [\*Access to the 10,657 documentary resources\*](#)

EPFL

EPFL supports Open Access

Search

Submit

Personalize

Help

Home &gt; Search Results

Search:

any field ▾

Search

Browse

[Search Tips](#) :: [Advanced Search](#)

Search collections:

\*\*\* any public collection \*\*\* ▾

Sort by:

latest first ▾

desc. ▾

- or rank by - ▾

Display results:

10 results ▾

single list ▾

Output format:

HTML brief ▾

ILC Document Server 33,007 records found 1 - 10 ▶▶▶ jump to record:  Search took 0.00 seconds.

- 1. **ILC Higgs White Paper**  
-author info not available- / -info not available- / Fri, 22 Mar 2013  
[Detailed record](#) - [Similar records](#)
- 2. **Additional Higgs Studies for Snowmass**  
Barklow Timothy / ILC Higgs White Paper / Fri, 22 Mar 2013  
[Detailed record](#) - [Similar records](#)
- 3. **SiEcal plans**  
Balagura Vladislav / CALICE Collaboration Meeting in Hamburg / Fri, 22 Mar 2013  
[Detailed record](#) - [Similar records](#)
- 4. **DESY GridGEM Test Beam**  
Müller Felix / LCTPC WP meeting #171 / Thu, 21 Mar 2013  
[Testbeam\\_March\\_2013.pdf](#)  
[Detailed record](#) - [Similar records](#)

- 1 Introduction
- 2 Technology**
- 3 Community
- 4 Projects
- 5 Conclusions

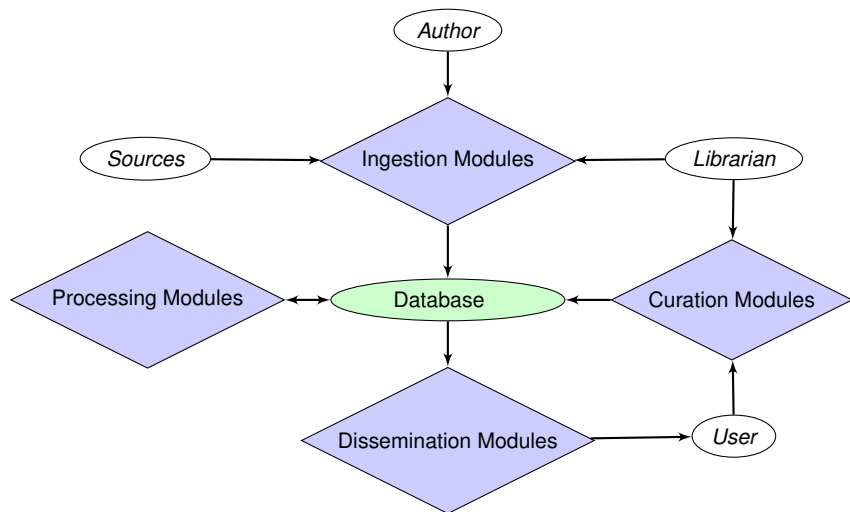
## ■ key features:

- **navigable collection tree** (regular, virtual)
- **powerful search engine**
  - Google-like speed for up to 5M records
  - combined metadata, reference and fulltext search
- **flexible metadata model** (MARC)
  - handling any kind of document (multimedia)
  - customizable input, formatting and linking
- **personalization** and **collaborative** features:
  - alerts, baskets, groups, reviews, comments
  - internationalisation (28 languages)

## ■ **modular architecture**: four module families

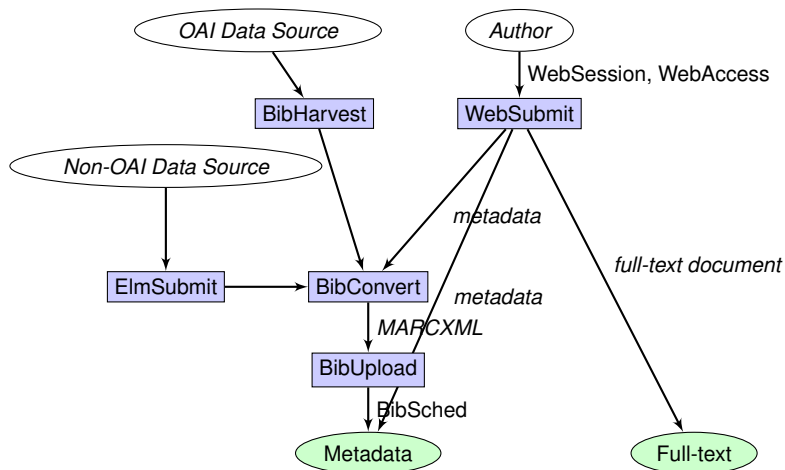
- (1) ingestion modules, e.g. periodical harvesting
- (2) processing modules, e.g. keyword extraction and indexing
- (3) dissemination modules, e.g. personalised RSS alerts
- (4) curation modules, e.g. batch corrections

# Invenio Modules: Overview

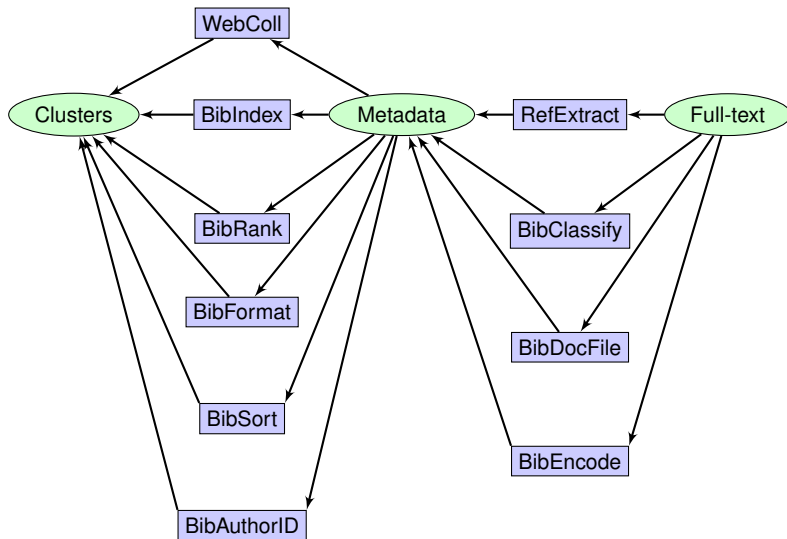




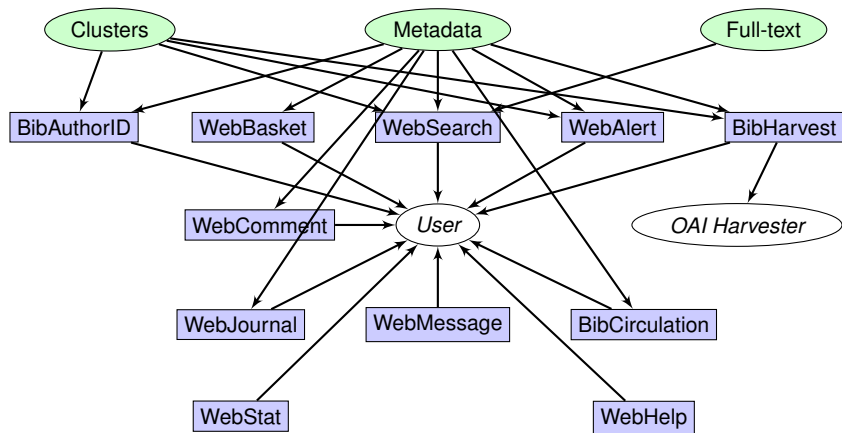
# Invenio Modules: Ingestion



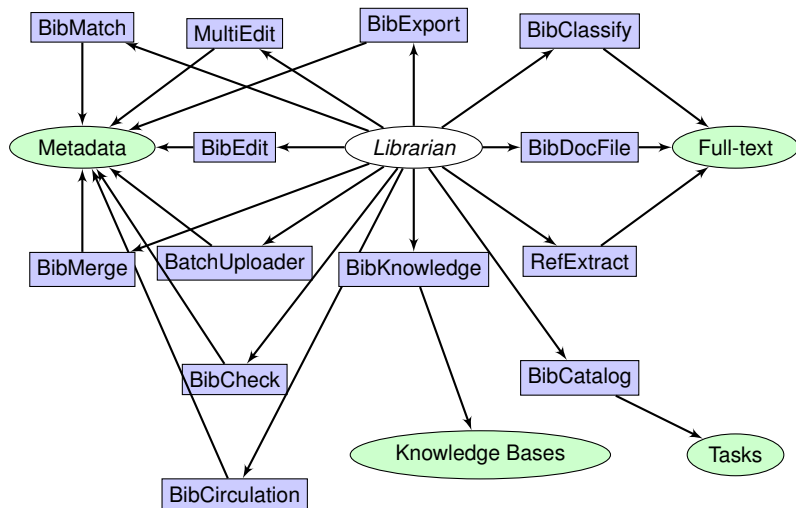
# Invenio Modules: Processing



# Invenio Modules: Dissemination

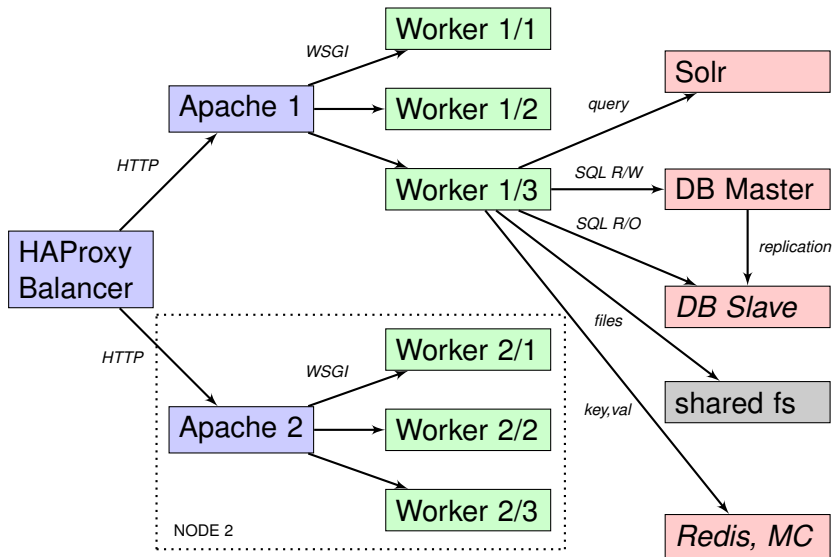


# Invenio Modules: Curation



# Scalable Architecture

- 800 hits per second on CDS during Higgs seminar July 4th



## ■ **technology overview:**

- load balancing: HAProxy
- web application: Apache, WSGI, Python, Flask, Jinja
- database: SQLAlchemy, MySQL/PostgreSQL/SQLite, MongoDB
- caching: Memcached, Redis
- UI: Twitter Bootstrap, jQuery
- project tools: Git, Trac, Jenkins, Selenium

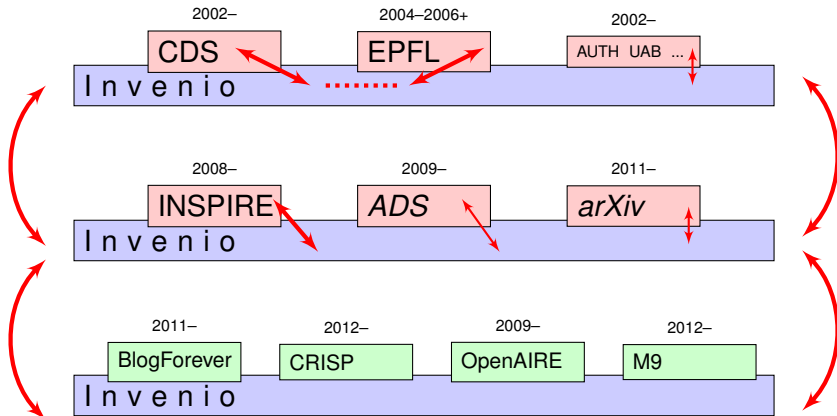
## ■ **development and collaboration model:**

- organic-growth software development model
- pull-on-demand collaboration model
- rapid prototyping
- unit tests, regression tests, web tests



- 1 Introduction
- 2 Technology
- 3 Community**
- 4 Projects
- 5 Conclusions

# Community



350k LOC - Invenio core sources

11k LOC - INSPIRE overlay sources





*Invenio User Group Workshop 2012.*

- 50 developers and contributors in 2012 (22 new):
  - 20 CERN-IT (7 new), 14 CERN-GS (5 new)
  - 4 SLAC, 2 ADS
  - 2 JUELICH (2 new), 1 GSI (1 new)
  - 1 UAB, 1 AUTH, 1 UNIZAR (1 new), 1 RERO (1 new)
  - 1 RNRT (1 new)
  - 2 upstream others (1 new)

- 1 Introduction
- 2 Technology
- 3 Community
- 4 Projects**
- 5 Conclusions

## ■ BlogForever

- OAIS, ingestion store
- preservation, METS, PREMIS



## ■ CRISP

- DOI minting via DataCite
- persistent identifier store
- “data continuum”



## ■ M9

- non-MARC master formats (EAD, UNIMARC)
- virtual fields, calculated or derived
- data model definition and checks



## ■ OpenAIRE, OpenAIREplus

- going beyond papers: multidisciplinary data archive
- community curated collections





*“Think about how great it would have been [...] to see the last minute “blogs” from disappearing Pompei...”*

–BlogForever video at CeBIT 2013

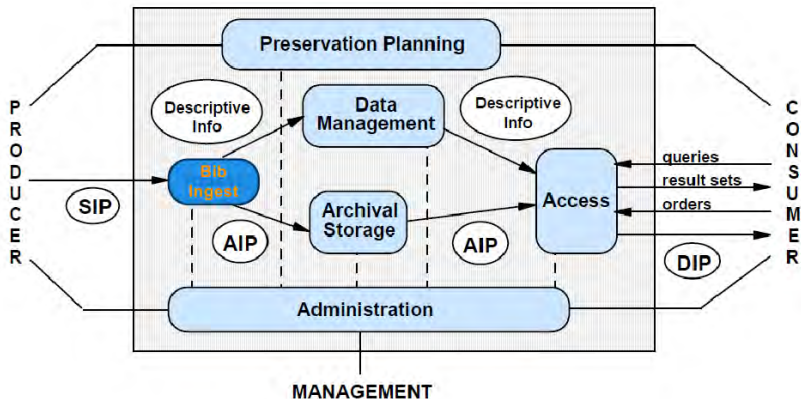
<<http://youtu.be/Gzp5ug4i7r4>>

- preserving blogs: taking care of today’s stories for tomorrow’s history
- EC funded, 2011–2013
- using Invenio for repository
  - blog posts, digital objects (images, videos), user comments
  - harvest, preserve, manage, disseminate blog content
  - ensure authenticity, integrity, completeness, usability





# OAIS Functional Entities



SIP = Submission Information Package · AIP = Archival Information Package · DIP = Dissemination Information Package

- WP17 following “data continuum” from ingestion through analysis to dissemination
- EC funded, 2012–2014
- ESRF and ILL, neutron physics use case
- Invenio use case
  - DOI minting of objects via DataCite 10.1000/182
  - Persistent Identifier store
  - citability and tracability throughout lifecycle





- M9: innovative cultural pole with a museum, exhibition spaces, médiathèque archives, educational activities, public services
- “Museum of the 20th Century” in Mestre, Italy
- privately funded, Fondazione di Venezia, 2011–2014
- Cini Foundation, CILEA
- using Invenio for repository
  - native support for formats ICCD, MAG, UNIMARC
  - native support of **archival formats** EAD, ISAD(G)
  - **data model abstraction** and logical fields
  - rich API to expose objects to apps





- EC funded, 2009–2014
- linking peer-reviewed literature to associated data and funding schemes
- using Invenio for “orphan” repository
  - launching Zenodo
  - multidisciplinary publication and data store
  - attractive submission and visualisation platform
  - collaborative user-defined user-curated collections
  - open access, embargoed access
- collaborating with EUDAT
  - EUDAT selected Invenio for SimpleStore component
  - iRODS storage backend





zenodo

Research. Shared.

[Search](#) [Upload](#) [Get started](#)

[Home](#) [Publications](#)

Improved photovoltaic performances by post-deposition acidic treatments on tetrapod shaped colloidal nanocrystal solids

10 July 2012

Journal article **Embargoed access**

## Improved photovoltaic performances by post-deposition acidic treatments on tetrapod shaped colloidal nanocrystal solids

Mastria, Rosanna ; Rizzo, Aurora ; Nobile, Concetta ; Kumar, Susmit ; Maruccio, Giuseppe ; Gigli, Giuseppe

[\(show affiliations\)](#)

The ligand exchange reaction with pyridine is the standard procedure for the integration of colloidal semiconductor nanocrystals (NCs) in photovoltaic devices; however, for large sized and irregularly shaped branched NCs, such as CdSe@CdTe tetrapods, this procedure can lead to a considerable waste of materials and the aggregation of NCs in the colloidal solution, therefore

### Publication date

10 July 2012

### Embargoed

Files available as **Open Access** after 10 July 2013

### DOI

[10.1088/0957-4484/23/30/305403](https://doi.org/10.1088/0957-4484/23/30/305403)

### Report number(s):

OpenAIRE-ESCORT-2012-008

### Published in:

Nanotechnology. 23 (2012) no. 30,

### Funded by:

ESCORT - Efficient Solar Cells based on Organic and hybrid Technology (261920)

### Collections:

[Publications](#) > [Journal articles](#)

- 1 Introduction
- 2 Technology
- 3 Community
- 4 Projects
- 5 Conclusions**

## ■ Invenio

- mature, robust, configurable digital library platform
- co-developed by CERN, CfA, DESY, EPFL, SLAC. . .
- supporting diverse objects beyond papers or multimedia
- supporting OAIS-inspired preservation features
- workflows from data ingestion to processing and visualisation
- primary use cases in HEP: CDS, INSPIRE
- participating in (EC funded) projects BlogForever, CRISP, OpenAIREplus, M9; collaboration with EUDAT

## ■ potential role in DPHEP

- already plays a role, via INSPIRE
- possible front-end for larger-scale data backends (IRODS, CASTOR)
- technology reuse: DOI minting, persistent identifier store, ingestion store, audit store, format migrations. . .

