



# Introduction to Invenio

Tibor Šimko

@tiborsimko

CERN-UNESCO School on Digital Libraries, Kumasi, Ghana · November 2016

# Outline

- 1 Introduction
- 2 Use cases
- 3 Invenio 1
- 4 Technology evolution
- 5 Invenio 3

1

# Introduction

# What is Invenio?

- digital library and document repository software
  - mature platform: first public release in 2002
  - rich data: articles, books, notes, photos, videos, data, code
- originated in high-energy physics
  - institutional repository: CERN Document Server
  - integrated library system: CERN Library
  - disciplinary repository: INSPIRE
- nowadays co-developed by an international collaboration



- participating in and collaborating with several EU projects



2

# Use cases

# Use cases

CERN Access and Records | Sign in | Overview

## CERN Document Server

Access articles, reports and multimedia content in HEP

Search Submit Help Personalize

Search 1,262,525 records for:

Advanced Search

<ul style="list-style-type: none"><li>Articles &amp; Preprints (1,086,543)</li><li>Published Articles (21,647) Preprints (27,136)</li><li>Theses (2,675) Reports (2,645)</li><li>CERN Notes (2,049)</li><li>Conference Documents (2,048)</li></ul>	<ul style="list-style-type: none"><li>CERN Articles &amp; Preprints (1,05,016)</li><li>CERN Published Articles (21,646) CERN Preprints (27,136)</li><li>CERN Theses (2,675) CERN Reports (2,645)</li><li>CERN Notes (2,049)</li><li>Conference Documents (2,048)</li></ul>
<ul style="list-style-type: none"><li>Books &amp; Proceedings (100,511)</li><li>Books (1,485) Proceedings (12,449)</li><li>Standards (12,019) Design Reports (273)</li></ul>	<ul style="list-style-type: none"><li>CERN Journals Reports (110) CERN Yellow Reports (1,350)</li><li>CERN Theory (13,330)</li><li>Academic Training Lectures (716) Summer Student Lectures (261)</li><li>General Talks (2,077)</li></ul>
<ul style="list-style-type: none"><li>Presentations &amp; Talks (19,143)</li><li>Books (1,485) Proceedings (12,449)</li><li>Standards (12,019) Design Reports (273)</li></ul>	<ul style="list-style-type: none"><li>CERN Departments (93,140)</li><li>Diplomate Services Unit (DSU) (2,014) Technology Department (TE) (211)</li><li>Finance and Procurement (FP) (1,316) Physics (PH) (60,145)</li><li>Accelerator Technology (AT) (1,316) Accelerators &amp; Technology Sector (ATS)</li><li>Technical Support (TS) (1,316) Information Technology (IT) (1,616)</li><li>Human Resources (HR) (1,316) Engineering Department (EP) (116)</li><li>Beams Department (BS) (116)</li></ul>

## institutional repositories & ILS

zenodo Research. Shared.

Search Communities Browse Upload Get started Sign in Sign up

Filter by types: Datasets, Images, Posters, Presentations, Publications, Software, Video/Audio

Recent Uploads: keyerson: Pretest (with friends, students) November 16th

it's work! (but results are not meaningful, because there are too few participants) Uploaded by maabre83 on 13 November 2014.

Using GitHub? Check out our GitHub integration. Software Preservation Made Simpler

New to ZENODO? Research, Shared. — all research outputs from across all fields of science are welcome. Citable. Discoverable. —

## research data

INSPIRE HEP

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

HEP · He-Names · INSTITUTIONS · CONFERENCES · JOBS · EXPERIMENTS · JOURNALS · HELP

## HEP Search

### High-Energy Physics Literature Database

Use Find or SPRES-style search (other tips)

Use Find

url: [hep-th/9711054](http://hep-th/9711054)

#### HOW TO SEARCH

**SPRES syntax is (mostly) supported (requires 'Find')**  
find a quark, b and 1 quark and data > 1994  
find physics rev (JPL, LBL or JHEP) 1993, 112  
find eprint arxiv:1007.5048 (Note the plots available on the detailed record)  
find fulltext "quark-gluon plasmas" (note new 'fulltext' operator)  
find a site and refer to a writer (Note 'referator')  
find a name and cite by title 'SU5Y and topic 2001' (Note 'citeby')

**New techniques:**  
1993 paper quark multiplicity  
arxiv:1007.5048  
citeby/author refs - refer to author/writer  
author/keyword | author/sundrum cited 450->1350

**Additional Help:**  
More search tips and full help

#### HEP

- Search Tips
- INSPIRE Help
- Corrections
- Addresses
- Email Us

#### INSPIRE

- About INSPIRE
- Recent topics
- HEP Reviews
- SPRES-HEP (?)

#### RESOURCES

- ADS
- arXiv
- HEData
- FDG

#### INSPIRE News

2013-01-14 Use INSPIRE to generate your TALK Bibliography, see [INSPIRE-HEP](#)

## subject-based repositories

Labordoc Search Resources Help

Search:  any field  Ask a librarian

Search Tips:  Advanced Search

#### Focus on:

- Global Wage Report: Global Wages Report 2012-13
- Migrant workers
- Disability in the workplace
- HEALTHS
- Youth employment
- Fundamental principles and rights at work

#### Narrow by collection:

- ILD publications (97,430)
- ILD books (41,441)
- ILD journal articles (1,180)
- ILD working papers (2,403)
- ILD training materials (1,361)

#### Other publications (119,181)

- books (108,133)
- journal articles (95,300)

#### Latest publications by ILD (9)

- ILD meetings (21,220)
- International Labour Conference (7,227)
- Geneva Daily (13,440)

#### ILD Journals (7,543)

- International Labour Review (7,700)
- International Journal of Labour Research (87)

## multimedia

# Example: cds.cern.ch

CERN Accelerating science Sign in Directory

## 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](#) (534,677) [Preprints](#) (671,836)

[Theses](#) (18,873) [Reports](#) (6,913)

[CERN Notes](#) (56,649)

[Committee Documents](#) (24,269)

**Books & Proceedings** (100,511)

[Books](#) (71,080) [Proceedings](#) (17,659)

[Standards](#) (11,659) [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](#) (13,590)

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

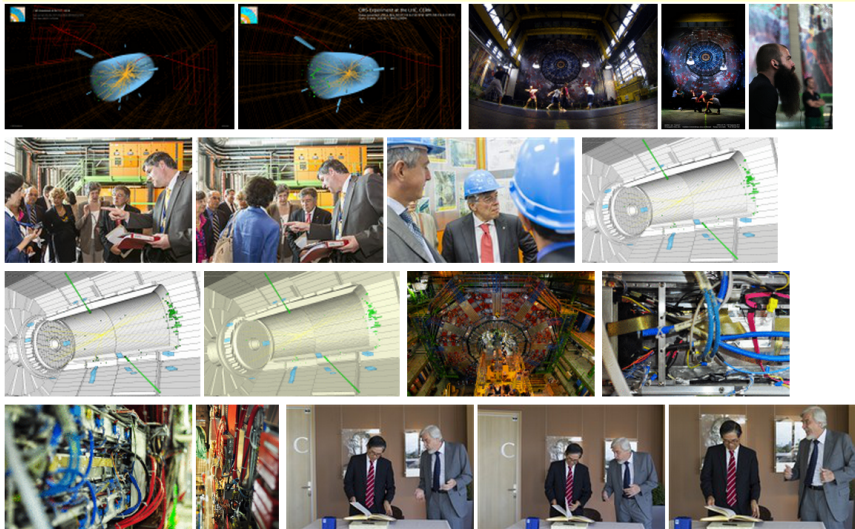
[CERN Departments](#) (95 148)

# E.g. CERN photo search

Photos

1,545 records found 1 - 10 ▶▶ jump to record:

Search took 0.48 seconds.






# E.g. CERN books and loans

**Results overview:** Found 2 records in 0.48 seconds.

**Books, 1 records found**

**Book Proposals, 1 records found**

**Books** 1 records found

1.  **Programming in Python 3 : a complete introduction to the Python language** / Summerfield, Mark

Upper Saddle River, NJ : Prentice-Hall, 2009. - 525 p.

[ebook - CERN library copies](#)




This  
book at  
Amazon

[Detailed record - Similar records](#)

Add to basket

**Book Proposals** 1 records found

1.  **Python in practice: create better programs using concurrency, libraries, and patterns** / Summerfield, Mark
- Python in Practice presents Python programming technologies and techniques to help programmers write code that's elegant, maintainable, and fast. [...]
- Addison-Wesley Professional, 2013. - 360 p.

This  
book at  
Amazon

[Purchase it for me! - Suggest for library](#)

[Detailed record - Similar records](#)

Add to basket

# Example: inspirehep.net



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 j "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 j hep,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 topics](#)
- [HEP Reviews](#)
- [SPIRES-HEP \[?\]](#)

#### RESOURCES

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

#### INSPIRE NEWS

2013-03-14 Use INSPIRE to generate your TeX bibliography, see [http://cds.cern.ch/record/1250000](#)

# E.g. citation summary

standard model Citesummary Search [Easy Search](#)  
[Advanced Search](#)  
[find j "Phys.Rev.Lett.,105"](#) :: [more](#)

Sort by:

Display results:

latest first desc. - or rank by - 25 results single list

## Citations summary

Generated on 2014-11-13

42915 papers found, 38790 of them citeable (published or arXiv)

### Citation summary results

**Total number of papers analyzed:**

[38,790](#)

Published only

[23,739](#)

**Total number of citations:**

1,150,182

1,036,646

**Average citations per paper:**

29.7

43.7

**Breakdown of papers by citations:**

Renowned papers (500+)

[160](#)

[153](#)

Famous papers (250-499)

[368](#)

[350](#)

Very well-known papers (100-249)

[1,821](#)

[1,706](#)

Well-known papers (50-99)

[3,163](#)

[2,891](#)

Known papers (10-49)

[12,253](#)

[10,176](#)

Less known papers (1-9)

[14,084](#)

[7,027](#)

Unknown papers (0)

[6,941](#)

[1,436](#)

$h_{\text{HEP}}$  index [\[?\]](#)

327

322

# E.g. citation history

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

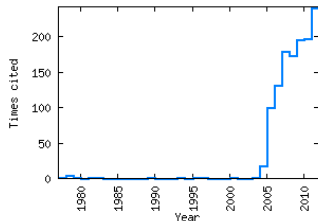
(538) [Baryogenesis Without Grand Unification](#) - Fukugita, M. *et al.* Phys.Lett. B174 (1986) 45 RIFP-641

(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:



# E.g. INSPIRE author page

Dixon, Lance Jenkins

[View Profile](#)

[Manage Profile](#)

[Manage Publications](#)

[Help](#)

## PERSONAL INFORMATION

### Personal Details (HepNames)

**Name** Lance J. Dixon

**Current Institution** SLAC

**E-mail** lance@slac.stanford.edu

**Links** <http://www.slac.stanford.edu/~...>

**Fields** HEP-PH  
HEP-TH

**Identifiers** BAI: [L.J.Dixon.1](#)  
INSPIRE: INSPIRE-00077352

Period	Rank	Institution
1978 – 1982	UG	Caltech
1982 – 1986	PHD	Princeton U.
1986 – 1987	PD	SLAC

## PUBLICATIONS AND OUTPUT

### Publications Datasets External

1. Universality in W+Multijet Production
2. Bootstrapping an NMHV amplitude through three loops
3. Universality in W+Multijet Production
4. Bootstrapping six-gluon scattering in planar N=4 super-Yang-Mills theory
5. Next-to-Leading Order Gamma Gamma + 2-Jet Production at the LHC
6. The four-loop remainder function and multi-Regge behavior at NNLLA in planar N = 4 super-Yang-Mills theory
7. Next-to-leading order diphoton+2-jet production at the LHC
8. Ntuples for NLO Events at Hadron Colliders
9. A brief introduction to modern amplitude methods
10. The BlackHat Library for One-Loop Amplitudes

[Click here to see all](#)

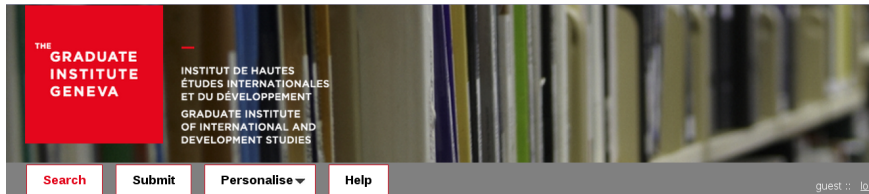
### Co-Authors

[Z.Bern.1 \(90\)](#)  
[D.A.Kosower.1 \(64\)](#)  
[F.Febres.Cordero.1 \(25\)](#)  
[H.Ita.1 \(23\)](#)  
[D.Maitre.1 \(21\)](#)  
[D.Seymour.1 \(16\)](#)

### Papers

	All papers	Single authored
<b>All papers</b>	174	22
Book	0	0
ConferencePaper	55	14
Introduction	2	2

# Example: IHEID



THE GRADUATE INSTITUTE GENEVA

INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT  
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

Search Submit Personalise ▾ Help

guest :: log

## The Graduate Institute's Publications

Welcome to the institutional repository of the Graduate Institute of International and Development Studies. It contains research produced by its members, including books, book chapters, journal articles, working papers, theses and more.

### Publications from the Faculty

Select a Professor...

Search 7,393 records for:

any field

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

### Recent Publications



**Cassese's international criminal law**  
Cassese, Antonio ; Gaeta, Paola. - 3rd ed. -  
Oxford : Oxford University Press, 2013. - 414 p.




**European security in NATO's shadow: party ideologies and institution building**  
Stephanie C. Hofmann. - Cambridge :  
Cambridge University Press, 2013. - 264 p

# Example: RERO

Submit Personalize Help | DE EN FR IT | guest :: Login

**réro doc**  
Digital Library

See all... 

> Advanced Search

**28'741 objects**



### Working papers SES

The Working Papers SES collection is a series of research papers authored by members of the Faculty of Economics and Social Sciences of the University of Fribourg (Switzerland). This series...

[Collection Page »](#)



## Latest news

-  RERO DOC : new interface  
2013/11/02

 RERO DOC & Multivo: improved integration

## Latest collections

-  Journal « La Gruyère »  
2012/09/12

 SDC - SDC Deposit Library  
2013/06/05

## Content by document type

- > Articles (14'694)
- > Periodicals (5'463)
- > Theses (2'958)
- > Books (2'731)
- > Dissertations (2'369)

# Example: EPFL



YOU ARE ▾

BY SCHOOL ▾

ABOUT EPFL ▾

Directory ▾ 🔍

EPFL > **infoscience**

English / français

## INFOSCIENCE

Search Browse Deposit / Publish Help



### Search publications

Any field 🔍

▶ [Advanced Search](#)

[Search Tips](#)

▶ [Access to the 99,567 scientific publications](#)

▶ [Access to the 9,940 documentary resources](#)

### Infoscience, EPFL's scientific publications

[Directive for Research Integrity](#)

[SNSF's Directive for open access publications](#)



#### Open Access conferences, Thursday October 24, 2013

**17.09.13** In the context of the Open Access Week, the EPFL library is pleased to invite you to the event « Open Access: authors, publishers, investors and institutions for the dissemination ... »



#### Collecting research data sets: a worldwide movement.

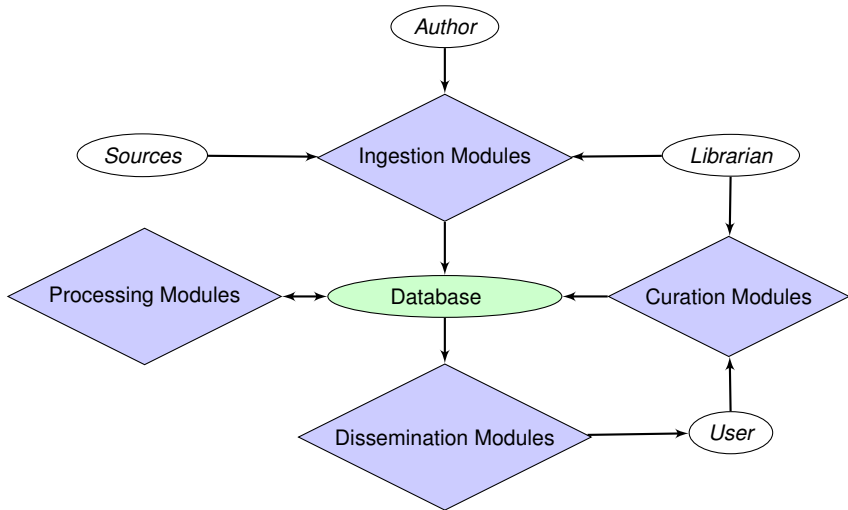
**03.06.13** One often hears that the principle of open access is to make research results available online for free as scientific publications (mostly articles). However, scientific research produces increasingly large amounts ... »



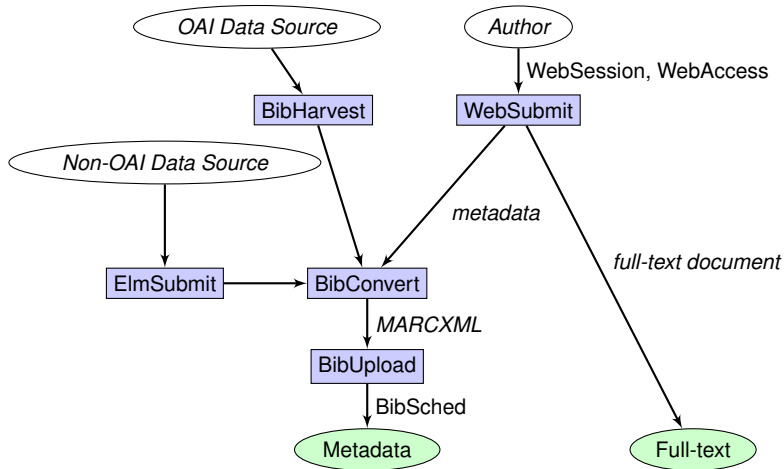
3

# Invenio 1

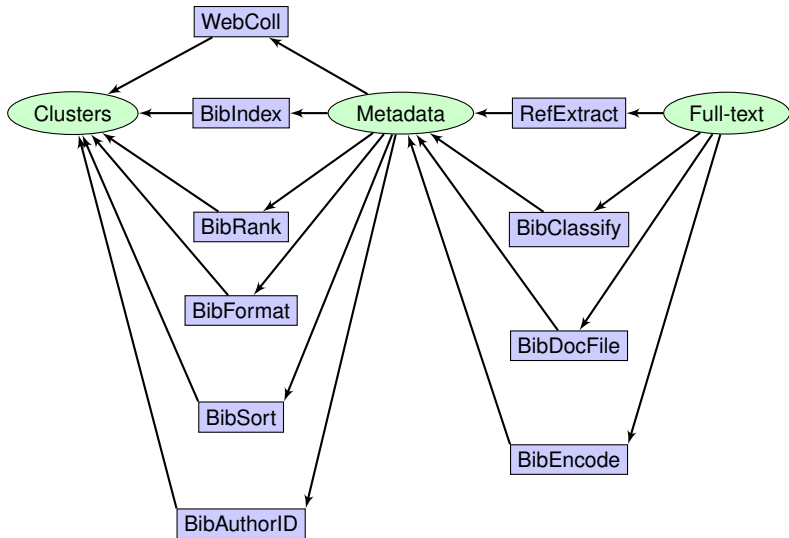
# Modular Architecture



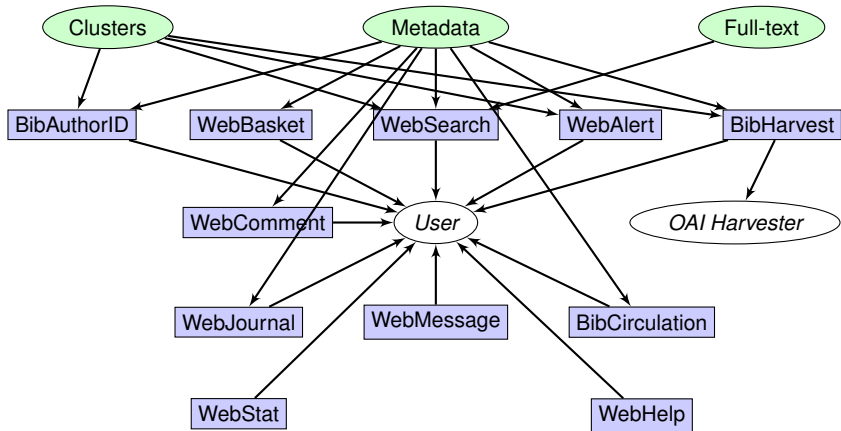
# Ingestion Modules



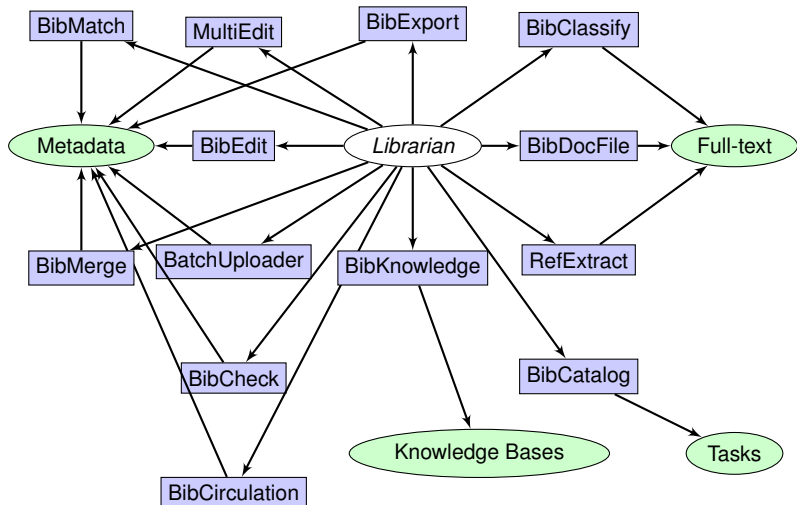
# Processing Modules



# Dissemination Modules



# Curation Modules



# Module overview

- 60+ modules
- 450,000+ lines of code
- 160+ authors and contributors since 2002

4

# Technology evolution



# Community



UNIVERSITY OF MINNESOTA



24 non-CERN

19 CERN-IT

11 CERN-Lib

54 contributors opening 1170 issues or pull requests on GitHub in the past 6 months

# Community

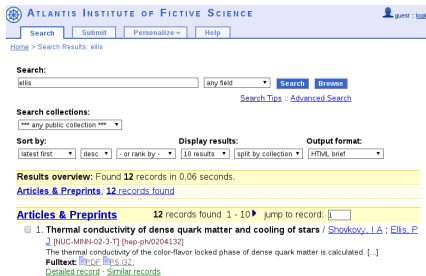


*Invenio User Group Workshop 2012*

- developer community growing
  - 4 developers and contributors in 2002
  - 48 developers and contributors in 2012
- adapting tools and processes
  - custom processes → mainstream processes
  - custom MVC → mainstream MVC

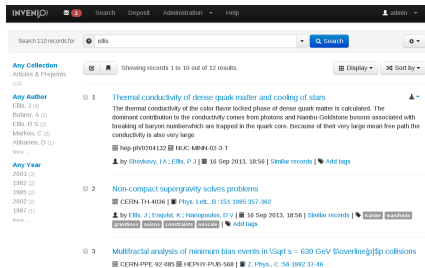
# Example: UI evolution

## Invenio v1



The screenshot shows the Invenio v1 search interface. At the top, it says 'ATLANTIS INSTITUTE OF FICTIVE SCIENCE' with a user profile 'guest - login'. Below that are search buttons: 'Search', 'Submit', 'Personalize', and 'Help'. A search bar contains 'ellis' and a dropdown menu is set to 'any field'. There are 'Search' and 'Browse' buttons. Below the search bar, there are options for 'Search collections' (set to 'any public collection') and 'Sort by' (set to 'latest first'). A 'Results overview' section states 'Found 12 records in 0.06 seconds.' and 'Articles & Preprints, 12 records found'. A list of results is shown, with the first item being 'Thermal conductivity of dense quark matter and cooling of stars' by Shovkovy, I.A. & Ellis, P.

## Invenio v2



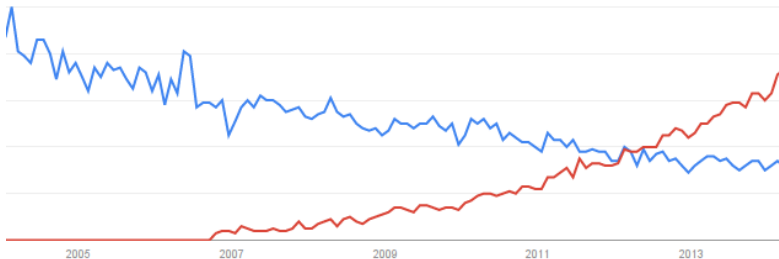
The screenshot shows the Invenio v2 search interface. At the top, it says 'INVENIO' with a user profile 'admin'. Below that are search buttons: 'Search', 'Deposit', 'Administration', and 'Help'. A search bar contains 'ellis' and a dropdown menu is set to 'any field'. There are 'Search' and 'Browse' buttons. Below the search bar, there are options for 'Search collections' (set to 'any public collection') and 'Sort by' (set to 'latest first'). A 'Results overview' section states 'Found 12 records in 0.06 seconds.' and 'Articles & Preprints, 12 records found'. A list of results is shown, with the first item being 'Thermal conductivity of dense quark matter and cooling of stars' by Shovkovy, I.A. & Ellis, P.

- home-grown CSS (custom)
- fast but niche template engine (Python)

- mainstream CSS (Twitter Bootstrap)
- mainstream template engine (Jinja)

# Example: API evolution

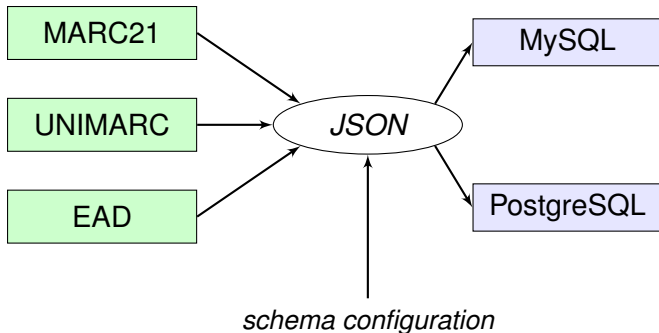
- Google search trends: XML API vs JSON API



- REST API

# Example: data model evolution

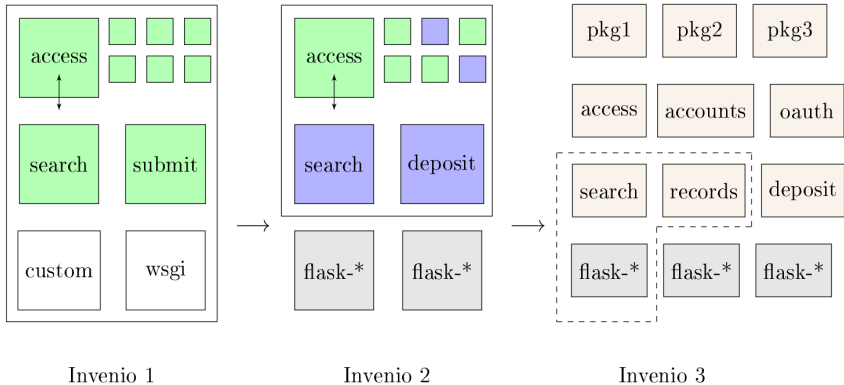
- abstraction of record fields:
  - metadata fields, e.g. author
  - derived fields, e.g. number\_of\_authors
  - virtual fields, e.g. number\_of\_citations
- abstraction of record formats:



4

# Invenio 3

# From Invenio 1 to Invenio 3



# Invenio watchmaking



Bootstrap

Jinja



BibSched  
BibRank  
WebSearch

...

Flask  
web development,  
one drop at a time



# Invenio release series

## 1 Invenio v1 (2010–)

- stable release series
- recommended for production
- v1.0.10
- v1.1.7
- v1.2.2

## 2 Invenio v2 (2013–2015)

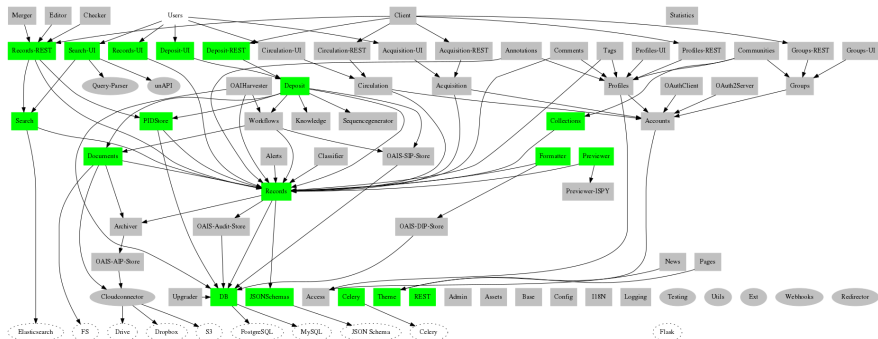
- transitional release series
- not recommended
- v2.0.6
- v2.1.1

## 3 Invenio v3 (2015–)

- latest and greatest
- bleeding-edge developer preview only
- v3.0.0a3

# Invenio 3 is a framework

- ecosystem of independent collaborative packages
- pick a pre-configured flavour. . .
- . . . or compose your own digital library solution



# Package ecosystem

This screenshot shows the GitHub repository page for Invenio digital library software. The page header includes navigation links for Pull requests, Issues, and Gist. The repository name is 'invenio', and the description is 'Invenio digital library software'. Below the repository name, there are statistics for Repositories, People (113), Teams (47), and Settings. A search bar contains 'invenio-records' and a 'New repository' button is visible. The main content area lists several packages:

- invenio-records-rest** (Python, 0 stars, 17 forks): Invenio records REST API module. Updated 23 hours ago.
- invenio-records** (Python, 2 stars, 20 forks): Invenio-Records is a metadata storage module. Updated 2 days ago.
- invenio-records-files** (Python, 0 stars, 5 forks): Invenio modules that integrates records and files. Updated 8 days ago.
- invenio-records-js** (JavaScript, 0 stars, 4 forks): Invenio-Records client library for building the web UI. Updated on 4 May.
- invenio-records-ui** (Python, 0 stars, 8 forks): Invenio records user interface module. Updated on 19 Apr.

Package	Build Status	Coverage	PyPI version
inveniosoftware/flask-breadcrumbs	build error	coverage 98%	pypi v0.3.0
inveniosoftware/flask-menu	build passing	coverage 100%	pypi v0.5.0
inveniosoftware/invenio-access	build passing	coverage 100%	pypi v1.0.0a5
inveniosoftware/invenio-accounts	build passing	coverage 100%	pypi v1.0.0a12
inveniosoftware/invenio-admin	build passing	coverage 100%	pypi v1.0.0a3
inveniosoftware/invenio-assets	build passing	coverage 98%	pypi v1.0.0a4
inveniosoftware/invenio-base	build passing	coverage 99%	pypi v1.0.0a11
inveniosoftware/invenio-celery	build passing	coverage 100%	pypi v1.0.0a4
inveniosoftware/invenio-collections	build passing	coverage 100%	pypi v1.0.0a2
inveniosoftware/invenio-communities	build passing	coverage 67%	pypi v1.0.0a6
inveniosoftware/invenio-config	build passing	coverage 100%	pypi v1.0.0a1
inveniosoftware/invenio-db	build passing	coverage 100%	pypi v1.0.0a9
inveniosoftware/invenio-documents	build passing	coverage 100%	pypi v1.0.0a1
inveniosoftware/invenio-files-rest	build failing	coverage 93%	pypi v1.0.0a3
inveniosoftware/invenio-groups	build passing	coverage 72%	pypi v0.1.3
inveniosoftware/invenio-i18n	build passing	coverage 98%	pypi v1.0.0a4
inveniosoftware/invenio-indexer	build passing	coverage 100%	pypi v1.0.0a6
inveniosoftware/invenio-jsonschemas	build passing	coverage 100%	pypi v1.0.0a3

# Technologies



front-end



Travis CI



tools and deployment



Jinja



elasticsearch.

back-end

SQLAlchemy

MySQL



MariaDB



mongoDB



redis

persistence

XML, OAI-PMH, JSON Schema,  
REST, Memento, DOI, ORCID,  
OpenAIRE...



<http://inveniosoftware.org/>

@inveniosoftware