N. Robinson; JY. Le Meur;
T. Simko;

CHEP 07

Managing an Institutional Repository with CDS

invenio

Contents

- What is CDS Invenio?
  - An overview of the main features
- The internals of CDS Invenio
  - An overview of the CDS Invenio architecture and technology
- Invenio licensing and support
- Conclusion

Overview

- An Integrated Digital Library system
  - Suite of applications and tools enabling the operation and maintenance of:
    - electronic preprint server, digital library catalogue, document archive
  - Web platform
- Used to manage CERN's institutional scientific repository:
  - About 1 million records; 550 collections; 10,000 searches per day
  - Wide range of content:
    - Documents (articles, preprints, etc)
    - Multimedia (photos, videos)
    - More…
  - Designed to cope with new dissemination channels of scientific results of LHC (Open Access)
- Makes use of existing standards:
  - US Library of Congress standards for bibliographic information description (MARC 21, MARCXML)
  - Unicode
  - OAI-PMH

Powerful search engine

- Specially designed indexes to provide "Google-like" search speeds for repositories of a few million records
- Customizable simple and advanced search interfaces
- Combined metadata, full-text and citation search in one go
- Results clustering by collection

CERN Document Server's Invenio

N. Robinson; JY. Le Meur;
T. Simko;
CHEP 07

CERN Document Server

Search | Submit | Convert | Agenda | Webcast | Bulletin | Library

guest :: login

Home > Search Results

CERN Document Server

Search:

lhc | any field | Search | Browse

Search Tips :: Advanced Search

Search collections:

*** any collection ***

Sort by: latest first | desc. | - or rank by - **Display results:** 10 results | split by collection **Output format:** HTML brief

Results overview: Found **29,797** records in 0.05 seconds.

Articles & Preprints, **19,586** records found
Books & Proceedings, **184** records found
Presentations & Talks, **197** records found
Periodicals & Progress Reports, **50** records found
Multimedia & Outreach, **7,100** records found
Archives, **2,680** records found

Articles & Preprints **19,586** records found 1 - 10 ▶ ▶▶ jump to record: 1

1. **Coordinate-space picture and $x \to 1$ singularities at fixed $k_\perp$** / Hautmann, F
We discuss ongoing progress towards precise characterizations of parton distributions at fixed transverse momentum, focusing on matrix elements in coordinate space and the treatment of endpoint singularities. [...]
arXiv:0708.1319; 10 Aug 2007 . - mult. p Fulltext

Navigable collection tree

- Documents organized in collections
- Customizable interfaces for each collection:
  - Flexible metadata to represent all type of objects
  - Flexible output formats (display and linking)
- Collections organized with regular and virtual trees
  - Regular collections are organized according to type of document
  - Virtual collections can be created according to any query (keywords, subject, author, etc)
- Integrated external collection searching

CERN Document Server's Invenio

N. Robinson; JY. Le Meur;
T. Simko;
CHEP 07

CERN Document Server

Over **800,000** bibliographic records, including **360,000** fulltext documents,
of interest to people working in particle physics and related areas.
Covers preprints, articles, books, journals, photographs, and much more.

Search 911,677 records for:

any field | Search | Browse

Search Tips :: Advanced Search

CERN Library News (5th July 2007)
Please give your comments on the list of scientific journals proposed for cancellation.

Narrow by collection:

- **Articles & Preprints** (746,930)
  Published Articles (289,778) Preprints (399,136) Theses (15,478) Reports (5,474) CERN Internal Notes (11,200) Committee Documents (27,623)
- **Books & Proceedings** (61,953)
  Books (38,118) Proceedings (16,187) Standards (7,648)
- **Presentations & Talks** (15,705)
  Conference Announcements (14,623) Academic Training Lectures (540) Summer Student Lectures (437) General Talks (33) Videotapes (295)
- **Periodicals & Progress Reports** (3,566)
  Periodicals (2,894) Progress Reports (672)
- **Multimedia & Outreach** (37,988)
  Photos (10,041) Videos (447) Press (19,589) Audio Archives (311) Exhibition Objects (179) Brochures (37) Posters (352) HEP Institutes (1,506) Experiments at CERN (841) Internet Resources (4,685)
- **Archives** (55,685)
  CERN Archives (49,421) Pauli Archives (3,780) DSU Archives (713) SL Archives (1,026) AB Archives (745)

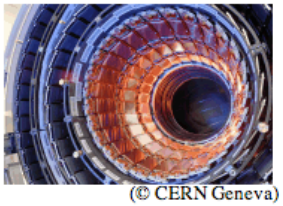
Focus on:

CERN Articles & Preprints (94,489)
CERN Published Articles (48,248) CERN Preprints (14,785) CERN Theses (2,813) CERN Reports (1,131) Committee Documents (27,623)
CERN Series (2,055)
CERN Yellow Reports (1,113) Academic Training Lectures (540) Summer Student Lectures (437) General Talks (33)
CERN Departments (66,793)
Accelerator Technology (AT) (4,882) Accelerators & Beams (AB) (15,766) Finance (FI) (818) Human Resources (HR) (4) Information Technology (IT) (3,352) Physics (PH) (37,298) Secretariat-General (SG) (7,788) Technical Support (TS) (1,207)
CERN Experiments (16,216)
LEP Experiments (5,563) LHC Experiments (10,287) Recognized Experiments (373)
CERN R&D Projects (462)
CERN Accelerator R&D Projects (462)

Search also:

- CERN Indico
- KISS Preprints
- SPIRES HEP

Photos



19 Oct 2006. *Detail of the sensor from the first half tracker inner barrel (TIB).*

(© CERN Geneva)

SEE ALSO...

CERN Press Office selection
Pictures of the Week
DESY Photos
FERMILab Photos
SLAC Photos
ESA Photos
NASA Gallery

Search 10,041 records for:

any field | Search | Browse

Search Tips :: Advanced Search

CERN Library News (5th July 2007)
Please give your comments on the list of scientific journals proposed for cancellation.

search also CERN PhotoLab Archive of unscanned pictures (1952-)

Latest additions:

2007-08-29 12:19

Israeli-Palestinian Party organized by the Summer Student of CERN the 22th August 2007.
22 Aug 2007
Keyword: Summer Student
Photo number: CERN-GE-0708008

Focus on:

CERN PhotoLab (5,829)
PhotoLab Archives (18,100)
Press Office Photo Selection (502)
ALICE Photos (326)
ATLAS Photos (1,358)
CMS Photos (423)

Conference	
Conference title	**International Conference on Computing in High Energy and Nuclear Physics**
Related conference title(s)	CHEP 07 CHEP 2007
Date(s), location	2 - 7 Sep 2007, Victoria, BC, Canada
Conference contact	CHEP2007: c/o Elly Driessen, Conference Coordinator: TRIUMF: 4004 WesbrookMall: Vancouver, BC V6T 2A3 CANADA email: chep07@triumf.ca
Imprint	2007


Conference home page

Other contributions to this conference in CDS

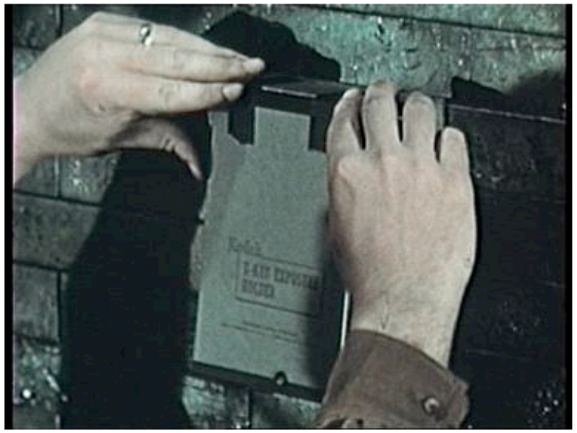
Record created 2006-05-12, last modified 2006-11-07

Similar records

ADD TO BASKET

Format: HTML | BibTeX | DC | EndNote | NLM | MARC | MARCXML

High Energy Particle Accelerators



View Movie
(Choose quality)

Low	Medium	High
(120 kbps)	(480 kbps)	(1000 kbps)

How to view .wmv videos?

Film about the different particle accelerators in the US. Nuclear research in the US has developed into a broad and well-balanced program.Tour of accelerator installations, accelerator development work now in progress and a number of typical experiments with high energy particles. Brookhaven, Cosmotron. Univ. Calif. Berkeley, Bevatron. Anti-proton experiment. Negative k meson experiment. Bubble chambers. A section on an electron accelerator. Projection of new accelerators. Princeton/Penn. build proton synchrotron. Argonne National Lab. Brookhaven, PS construction. Cambridge Electron Accelerator; Harvard/MIT. SLAC studying a linear accelerator. Other research at Madison, Wisconsin, Fixed Field Alternate Gradient Focusing. (FFAG) Oakridge, Tenn., cyclotron. Two-beam machine.
Comments : Interesting overview of high energy particle accelerators installations in the US in these early years. .

Produced by: Audio Productions, Inc, New York
Director: Atomic Energy Commission
35:00 min. / 1960 / AEC Copyright
Keywords: accelerators, particles, cosmotron, cylotron, proton synchrotron, linear accelerator

Original Source: P104
Language: eng
Source Medium: BETACAM PAL (Master)
Note: Original film : 16 mm, optical sound

Reference: CERN-MOVIE-1960-005

Experiments and Tracks

OUTREACH	CMS-PHO-OREACH-2006-027

Lowering of the YE+3 endcap disc on 30th November

Note: CMS Collection.;
Photographer: Max Brice, CERN;
Date: 30 Nov 2006

Gigantic disc of CMS detector travels 100 m under the Earth It's an amazing engineering challenge - the lowering of the first tremendous endcap disc, known as YE+3, of the CMS particle detector slowly and carefully 100 m underground into the experimental cavern. The disc is one of 15 large pieces to make the grand descent. It's a uniquely shaped slice, 16 m high, about 50 cm thick and weighing 400 tonnes. The solid steel structure of the disc forms part of the magnet return yoke and is equipped on both sides with muon chambers. A special gantry crane will lower the element, with just 20 cm of leeway between the edges of the detector and the walls of the shaft. CMS is one of the four main experiments that will take data at the world's highest energy particle accelerator, CERN's Large Hadron Collider (LHC). The LHC is a 27 km circular ring 100 m underground. The CMS detector weighs a total of 12 500 tonnes and is constructed on the surface. Once all of the pieces are fully equipped, lowered underground and re-tested, they will be pushed together in preparation for the LHC start-up in November 2007.

Keywords: cms ; lowering ; endcap ; disc ; disk ; yoke ; magnet ;

Related links:
CERN Courier vol 47 no 1 : January/February 2007

You can look at these photographs in the following formats:

JPEG Image

oreach-2006-027.jpg
1241479 bytes
[1531 x 1018]

Download half-size version

oreach-2006-027_01.jpg
1339923 bytes
[1531 x 1018]

Download half-size version

oreach-2006-027_02.jpg
1381681 bytes
[1531 x 1018]

Download half-size version

oreach-2006-027_03.jpg
1477685 bytes
[1531 x 1018]

Download half-size version

oreach-2006-027_04.jpg
1352544 bytes
[1531 x 1018]

Download half-size version

Document acquisition from multiple sources

- Document submission via Web interface
  - Submission forms highly configurable
  - Customizable approval workflows - configured to suit a variety of user needs
- Automated imports
  - E-mail
  - OAI-PMH and other custom protocols
  - Synchronization with external databases

Collaborative tools

- User-defined document baskets & automated email notification alerts
- Basket-sharing within user groups
- User comments and reviews of documents
- External authentication (e.g. LDAP) and use of all resulting information (user's groups, affiliations, etc)
- Multilingual interfaces
  - Currently 20 languages supported

Home > Your Account > Personal baskets > Physics > Display baskets

Display baskets

Personal baskets Group baskets

Physics (1)

Create new basket

Edit basket

Papers I like
2 records - last update: 30 Aug 2007, 22:40

Magnetic Heisenberg-chain/pp-wave correspondence / Harmark, T; Kristjansson, K R; Orselli, M
We find a decoupling limit of planar N=4 super Yang-Mills (SYM) on R x S^3 in which it becomes equivalent to the ferromagnetic XXX_{1/2} Heisenberg spin chain in an external magnetic field. [...]
hep-th/0611242; 22 Nov 2006 . - 35 p Fulltext - Published in: *J. High Energy Phys.* 02 (2007) 085

Details and comments

A string field theoretical description of (p,q) minimal superstrings / Fukuma, M; Irie, H
A string field theory of (p,q) minimal superstrings is constructed with the free-fermion realization of 2-component KP (2cKP) hierarchy, starting from 2-cut ansatz of two-matrix models. [...]
hep-th/0611045; KUNS-2047.- Kyoto : Kyoto Univ., 6 Nov 2006 . - 53 p Fulltext - Published in: *J. High Energy Phys.* 01 (2007) 037

Details and comments

Home > Your Account > Your Groups

Your Groups

You are an administrator of the following groups:

Group	Description		
nichtest	This is my test	Edit group	Edit members

Create new group

You are a member of the following groups:

Group	Description
cds	cds team

Join new group | Leave group

You are a member of the following external groups:

Group	Description
All Exchange People [CERN (external)]	All Exchange People (Group)
All Exchange People [CERN]	All Exchange People (Group)

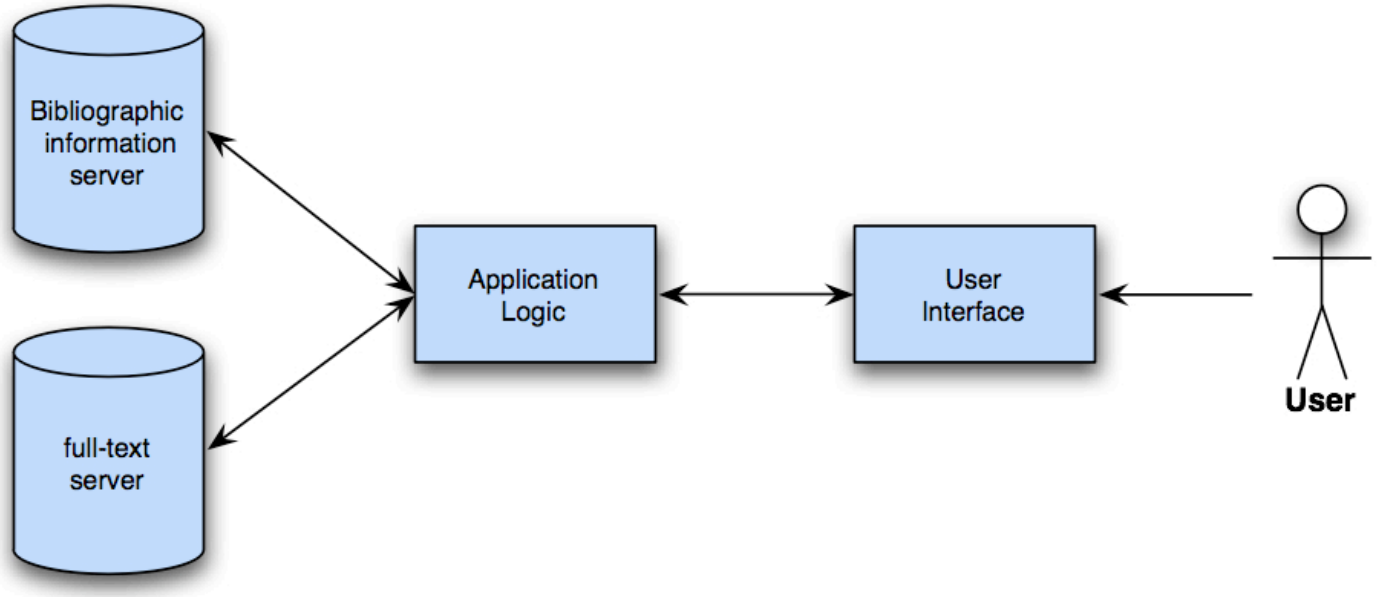
Additional applications running alongside Invenio at CERN

- Document format conversion
  CERN Conversion Server http://cdsconv.cern.ch
- Multimedia conversion & analysis (R&D)
- Search engine used as a back-end platform for Web front-end applications
  - Electronic Bulletins http://bulletin.cern.ch
  - Generation of Lists (publications, events, etc)
  - Conference & meetings (Indico) search

The internals of CDS Invenio

Architecture - simple view

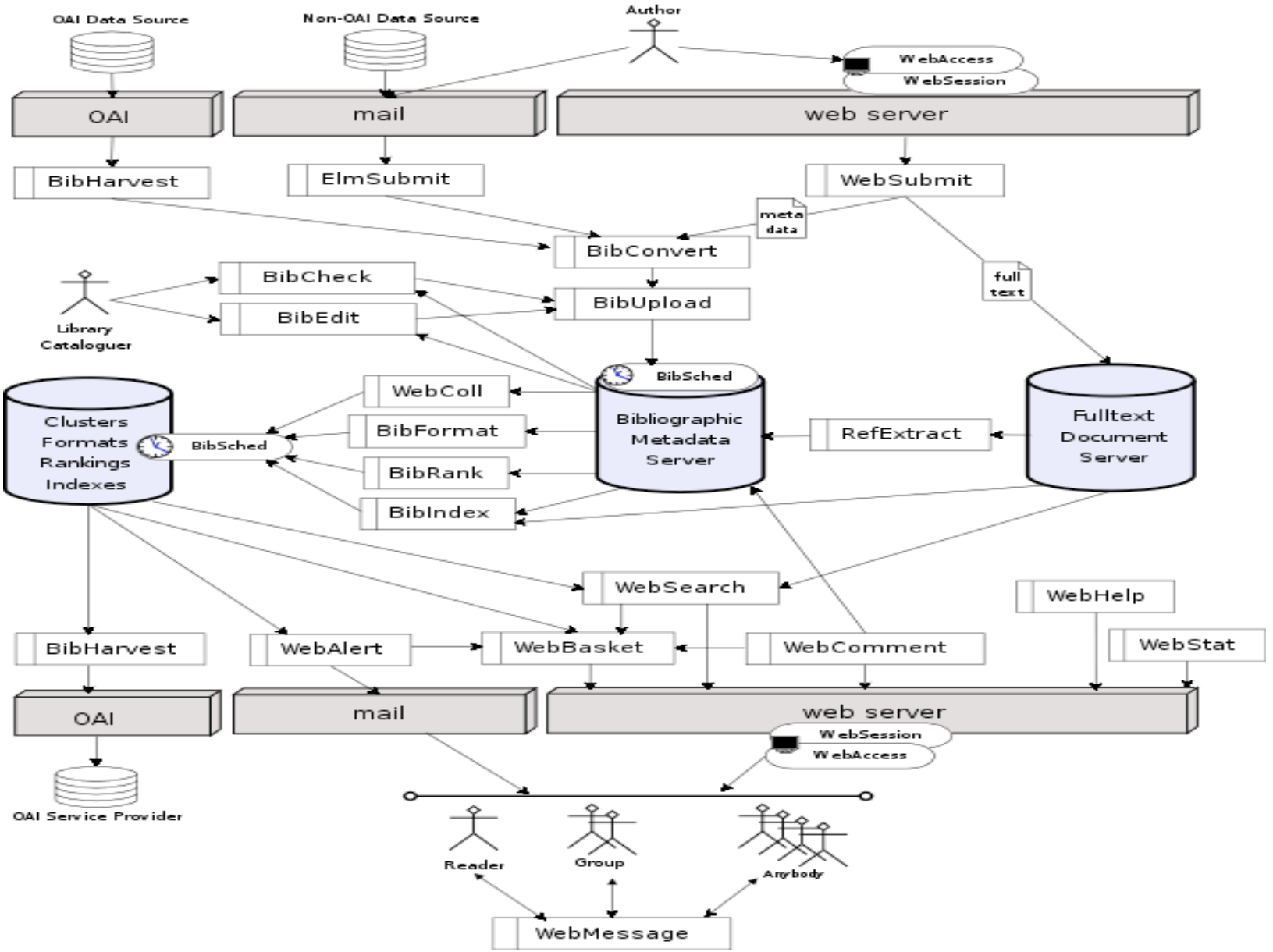
- Application and DB Servers



- Main load on the application server

CERN Document Server's Invenio

N. Robinson; JY. Le Meur;
T. Simko;
CHEP 07



Technologies used:

- Main programming language: *Python*
- Apache Web server; Python integration with *mod_python*
- Uses MySQL RDBMS
  - Take advantage of fully featured query language
- Internal metadata representation with MARCXML
- Export gateways:
  - Multiple output formats: HTML, XML, MARC, OAI, DC, etc.
- Home made indexes
  - Native RDBMS (MySQL) indexes have been tested:
    - 500,000 records ! 25+ Mrows ! 5+ sec searches
    - Google-like speed for up to 100,000 records only

Index space design

- **Performance-driven design** assumptions:
  - low number of updates, high number of selects
  - fast searching, slow indexation
  - put load on Web App Server, free DB Server
  - cache everything cacheable
- **Search modes:**
  - search for words
  - search for phrases (exact, partial)
  - search for regular expressions
- **Index types:**
  - forward : *term1* → [*rec1*, *rec2*, . . .]
  - reverse : *rec1* → [*term1*, *term2*, . . .]

Performance stats

- Dual Xeon(HT) 3.06 GHz, SCSI Ultra320
- 650,000+ records, 450+ collections
- **Indexing**: total index size 11 GB, indexing time 2 days
  - global words index: 3,000,000+ words
  - global words index growth rate: 2.8 words/record
  - title words index growth rate: 0.1 words/record
- **Searching**: typical search speed

query	*no. hits*	*search time*
ellis	1,797	0.07 sec
cern	223,843	0.07 sec
of	439,793	0.07 sec
of cern	109,635	0.10 sec
of cern the this	11,940	0.17 sec

(Stats from 2004)

- 2007: More than 1,000,000 records
  - developed new ultra fast bit vector library -> much better performance

The advantages of Python for Invenio

- Clean aesthetical language
- Easy to learn, important for many internship students and temporary members working on the project
- Very good for rapid prototyping & organic-growth development
- Plenty of ready-to-be-used modules
  - Heavier tasks make use of C modules
- Bytecode-compiled only, but speed okay for our needs

Invenio licensing and support

- Free: GNU GPL
- Regular public releases of software
  - CVS tree for developers and testers
- Support modes
  - Free via mailing lists
  - Paid support possible
- CDS Software Development Consortium
  - Main partners: EPFL, EIF; exchanging students, code, strategy
  - World wide contributions; internationalization
  - Open to newcomers!

Worldwide installations of Invenio

- **CERN Document Server** *(CERN, Geneva, Switzerland)*
- **International Linear Collider DOC** *(http://ilcdoc.linearcollider.org/)*
- **MeIND** *(HBZ NRW Koln Germany),*
- **INFOSCIENCE** *(EPFL, Lausanne, Switzerland)*
- **Aristotle University of Thessaloniki** *(Thessaloniki, Greece)*
- **RERO DOC** *(Martigny, Switzerland)*
- **PADIS** *(Università La Sapienza, Rome, Italy)*
- **CAB UNIME** *(University of Messina, Italy)*
- **FYNU UCL** *(Université catholique de Louvain, Belgium)*
- **McCammon Group** *(UCSD, San Diego, USA)*
- **University of Applied Sciences of Fribourg** *(Fribourg, Switzerland)*
- **DDD** *(Universitat Autònoma de Barcelona, Spain)*
- **EDUDOC** *(Swiss Education Server)*
- […]

Conclusions 1/2

- CDS Invenio: a powerful, flexible solution suitable for the management of very large collections of full text documents
- Deployed at CERN as the long-term institutional electronic archive
- Aims to enrich user experience by combining the best of the traditional library world with modern information retrieval technology
- OAI-compliant: maximum dissemination of scientific literature and results from LHC

Conclusions 2/2

- Open source system, used world-wide
- Collaboration with SLAC, Fermilab and DESY to study the possible use of Invenio by SPIRES
- On-going R&D:
  - ranking of scientific documents
  - automatically creating document clusters
  - increasing the size of repository up to 10 million records