
INSPIRE Technology

Tibor Šimko

`<tibor.simko@cern.ch>`

2nd HEP Information Resource Summit, Hamburg, Germany

May 20-21, 2008

Part I

Invenio

CDS Invenio History

▷ **pre-2000**: web interface to the CERN library server

▷ **2000**: CERN Document Server



- ★ digital library: articles, books, photos, videos

- ★ digital conference: sister web app for meetings

▷ **2002**: first public **CDSware** release (GNU GPL)

- ★ SDSC, San Diego, USA

- ★ HBZ NRW, Cologne, Germany

- ★ AUTH, Thessaloniki, Greece (and more)



▷ **2004**: first public **CDS Indico** release (sister product)

▷ **2006**: CDSware becomes **CDS Invenio**



- ★ SW targeting large repositories (1M+)

- ★ SW used by 20+ institutions and libraries worldwide

Invenio Key Features

- ▷ **navigable collection tree** (regular, virtual)
 - ▷ **powerful search engine**
 - ★ Google-like speed for up to 2M records
 - ★ combined metadata, reference and fulltext search
 - ▷ **flexible metadata** (MARC, OA)
 - ★ handling any kind of document (e.g. multimedia)
 - ★ customizable input, formatting and linking
 - ▷ **personalization** and **collaborative** features:
 - ★ groups, reviews, comments, baskets, alerts
 - ▷ **GNU General Public Licence**, Apache/Python/MySQL
 - ▷ **organic-growth** software development model
-

Example: Collection Tree



[Home](#)

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 886830 records for:

any field Search Browse

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

Narrow by collection:

- ☒ [Articles & Preprints](#) (729550)
 - [Published Articles](#) (276359) [Preprints](#) (368927)
 - [Theses](#) (42786) [Reports](#) (5452) [CERN Internal Notes](#) (10832) [Committee Documents](#) (26856)
- ☒ [Books & Proceedings](#) (60662)
 - [Books](#) (36573) [Proceedings](#) (16539) [Standards](#) (7550)
- ☒ [Presentations & Talks](#) (13948)
 - [Conference Announcements](#) (12913) [Academic Training Lectures](#) (539) [Summer Student Lectures](#) (386) [General Talks](#) (33) [Videotapes](#) (299)
- ☒ [Periodicals & Progress Reports](#) (3451)
 - [Periodicals](#) (2783) [Progress Reports](#) (668)

Focus on:

[CERN Articles & Preprints](#) (90546)
[CERN Published Articles](#) (47788) [CERN Preprints](#) (12430)
[CERN Theses](#) (2512) [CERN Reports](#) (1059) [Committee Documents](#) (26856)
[CERN Series](#) (1996)
[CERN Yellow Reports](#) (1104) [Academic Training Lectures](#) (539) [Summer Student Lectures](#) (386) [General Talks](#) (33)
[CERN Departments](#) (63144)
[Accelerator Technology \(AT\)](#) (4677) [Accelerators & Beams \(AB\)](#) (15149) [Finance \(FI\)](#) (725) [Human Resources \(HR\)](#) (0) [Information Technology \(IT\)](#) (2196) [Physics \(PH\)](#) (36281) [Secretariat-General \(SG\)](#) (7096) [Technical](#)

Example: Flexible Metadata

Photos

Search:

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

Search collections:

Sort by:

Display results:

Output format:

Photos

178 records found 1 - 12 ►► jump to record:

Search took 0.23 seconds.



Photos : 178 records found 1 - 12 ►► jump to record:

Example: Reviewing

People who viewed this page also viewed:

- (3) [The Feynman lectures on physics](#) - [Feynman, Richard Phillips](#) *et al*
- (3) [Learning Windows server 2003 2nd ed. ;](#) - [Hassell, Jonathan](#)
- (2) [With the unveiling of its new sign, the CERN Control Centre was officially inaugurated on Thursday 16 March.](#) - [IT-UDS-AVC Team](#) - CERN-VIDEOCLIP-2006-08
- (2) [Liability hedging and portofolio choice](#) - [Scherer, Bernd](#)
- (2) [Conduite de projet Web2e éd. ;](#) - [Bordage, Stephane](#)

Rate this document:

Average review score: ★★★★★ based on 1 reviews
Readers found the following reviews to be most helpful.

★★★★★ **A wonderful (and fun) guide to Common Lisp**

Reviewed by [tsi](#) on 14 Nov 2006, 17:48

0 out of 0 people found this review useful

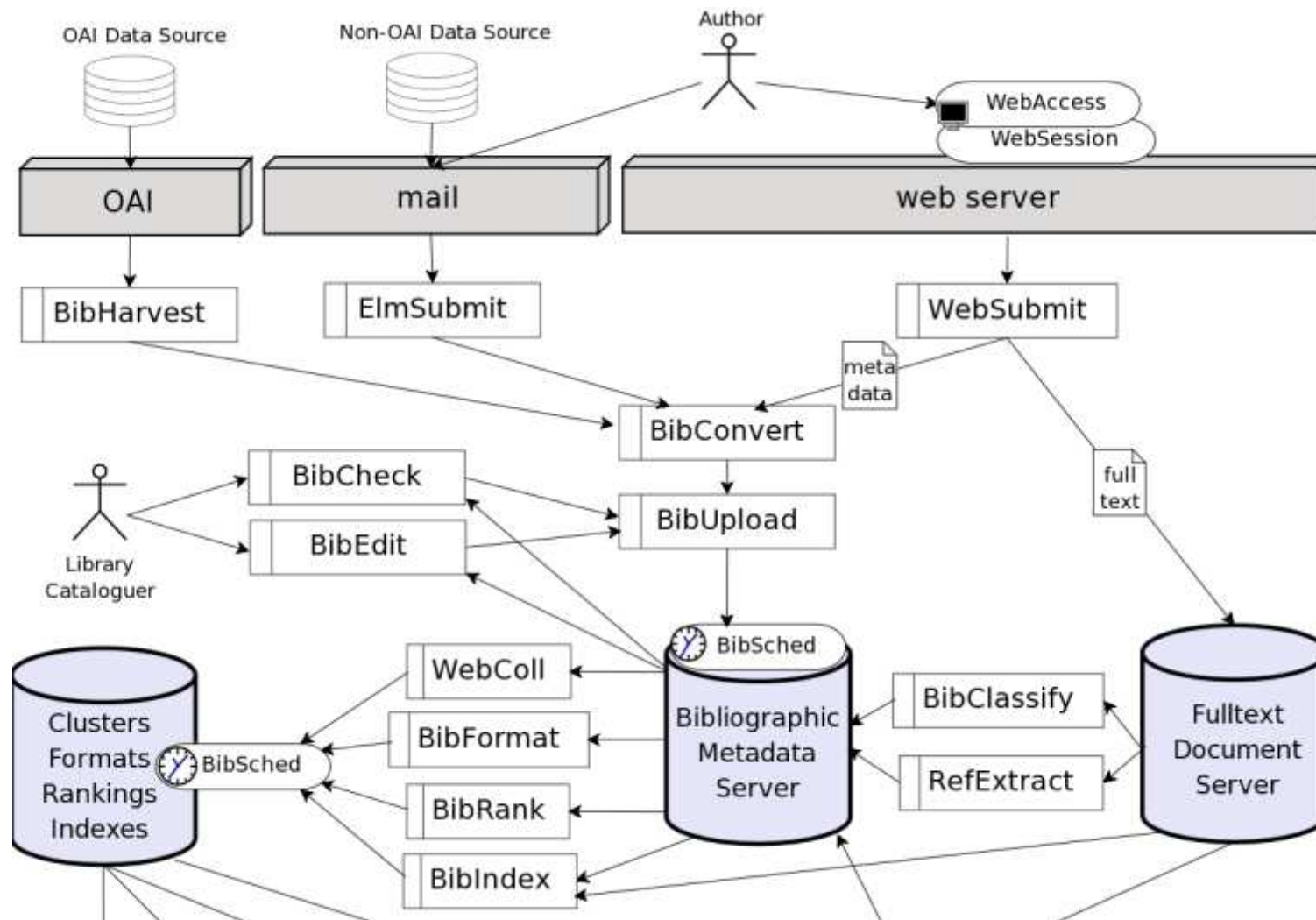
(Test.) I've been recommending this text to people who want to start learning Common Lisp since it was first available in draft form on the author's web site. Now that it's out in print I can enthusiastically recommend that anybody who is interested in learning Common Lisp - or even curious about how the language can improve your productivity - purchase it.

Peter has a very enjoyable and easy-to-understand writing style, and he starts early with practical examples that show how Common Lisp can be used to solved problems. Chapter 3, "A Simple Database", is a great explanation of how programs are grown from pieces in Common Lisp to solve large problems. It's presented early and draws people in to the problem solving techniques used when programming in Lisp.

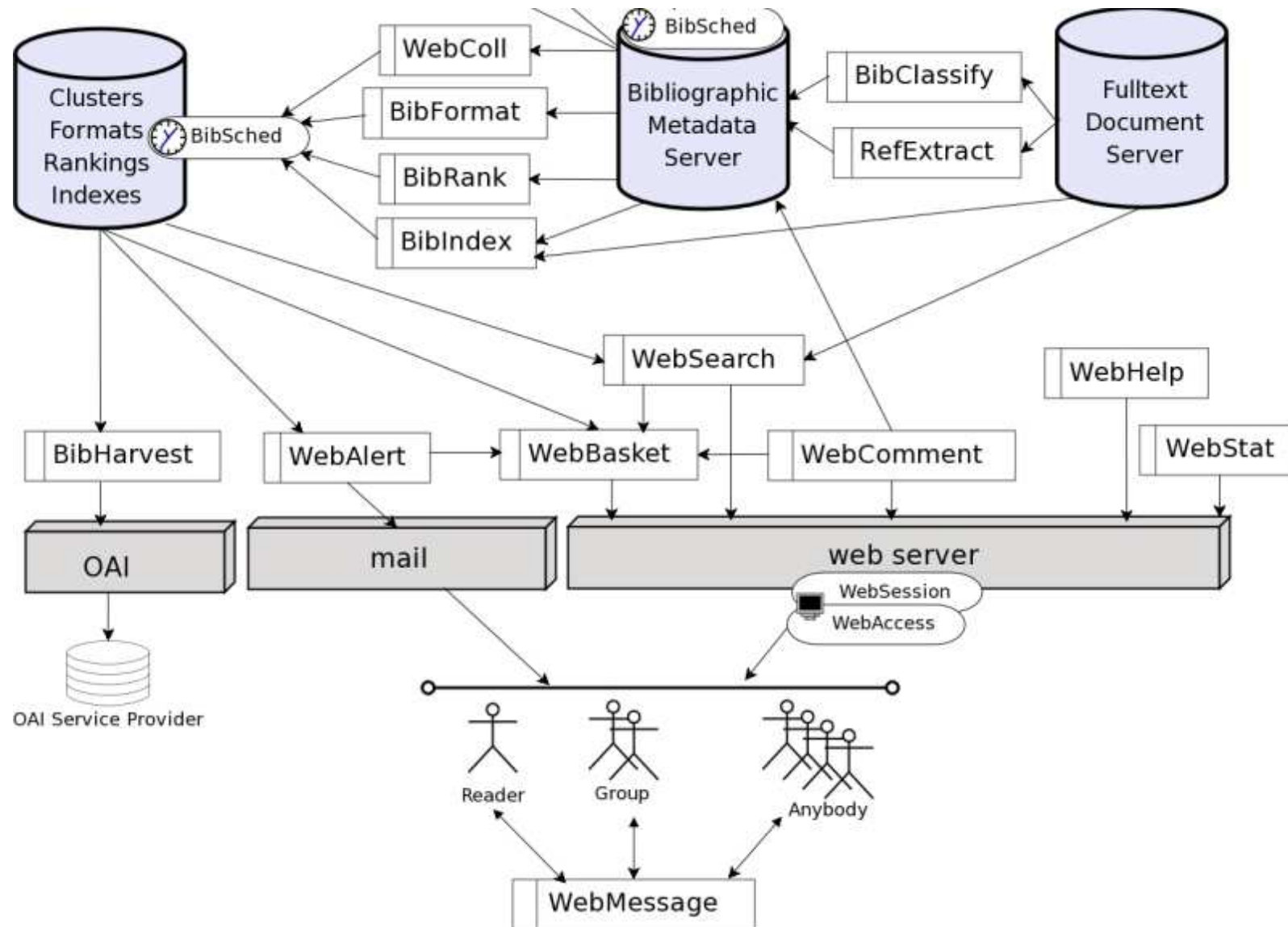
[Report abuse](#)

Was this review helpful? [Yes](#) / [No](#)

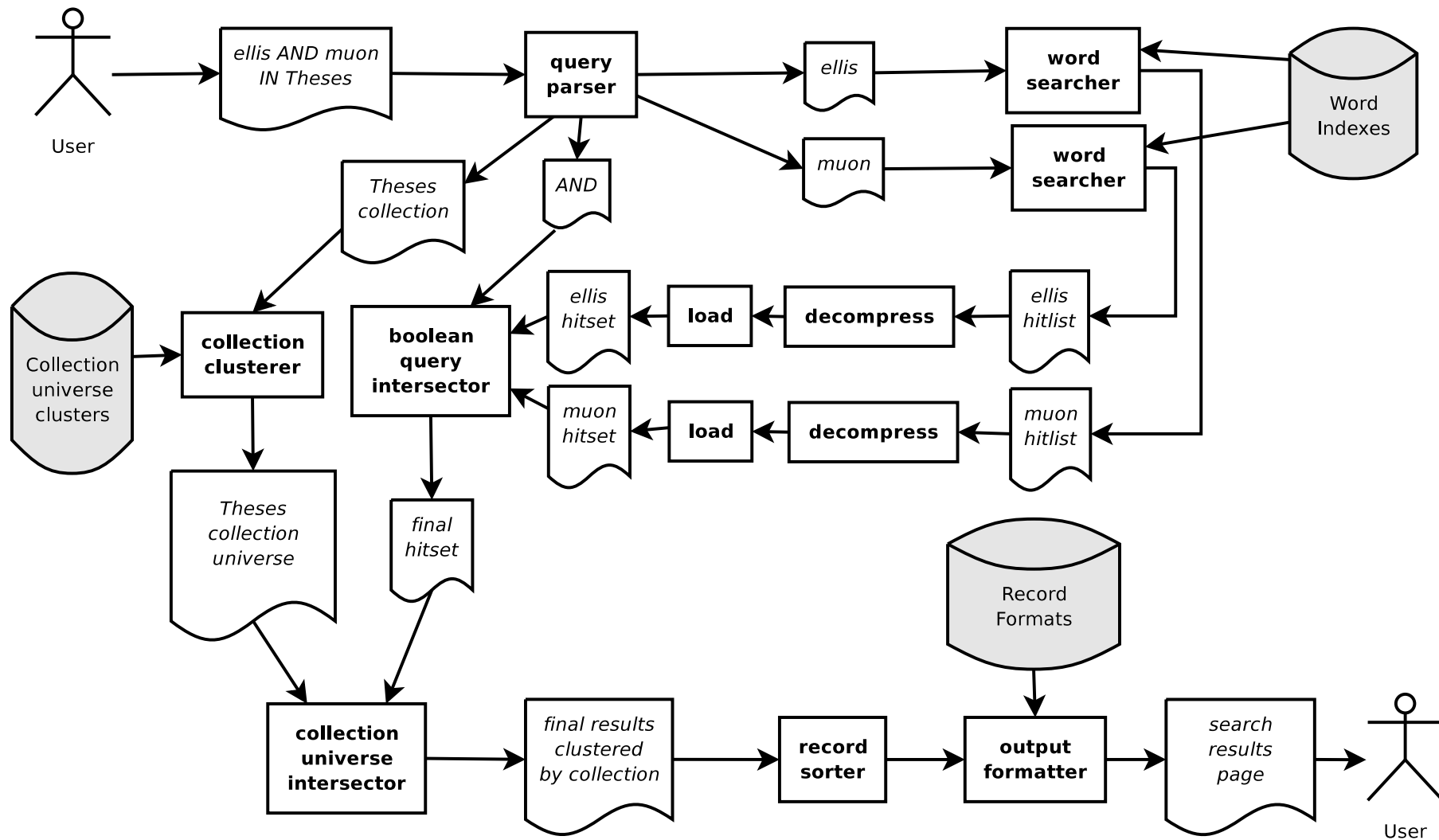
Invenio Module Overview I



Invenio Module Overview II



Example: Search Optimization I



Example: Search Optimization II

- ▷ four important **speed factors** to consider:
 - ★ speed of finding sets (DB Server)
 - ★ speed of uncompressing sets (DB \leftrightarrow Web App)
 - ★ speed of unmarshaling sets (Web App Server)
 - ★ speed of intersecting sets (Web App Server)
- ▷ **data structures** tested:
 - ★ 'sorted' (lists, Patricia trees)
 - ★ 'unsorted' (hashed sets, binary vectors)
- ▷ **fast prototyping**: (Lisp, Python)
 - ★ throw-away coding to test ideas
 - ★ typical search time gain: 4.0 sec \rightarrow 0.2 sec [2002]
- ▷ *binary vectors* found the best compromise!?

Part II

INSPIRE

INSPIRE Phase I (Q3 2007)

- ▶ *goal*: study of technical feasibility
- ▶ comparison of existing SPIRES and Invenio systems
- ▶ comparison of SLAC, Fermilab, DESY, CERN workflows
- ▶ ... concluded positively

ΙΝΣΤΙΤΟΥΤΟ
ΦΑΝΤΑΣΤΙΚΩΝ
ΕΠΙΣΤΗΜΩΝ
ΑΤΛΑΝΤΙΔΟΣ

επισκέπτης :: είσοδος

Αναζήτηση Υποβολή Ρυθμίσεις Βοήθεια

Αρχική Σελίδα

Ινστιτούτο Φανταστικών Επιστημών Ατλαντίδος

Αναζήτηση 94 εγγραφών για:

οποιοδήποτε πεδίο [Αναζήτηση](#) [Φυλλομέτρηση](#)

[Παραδείγματα αναζήτησης](#) :: [Σύνθετη αναζήτηση](#)

Περιορισμός με συλλογή:

- ☒ [Άρθρα & Προδημοσιεύσεις](#) (61)
- [Άρθρα](#) (26) [Προδημοσιεύσεις](#) (35)
- ☒ [Βιβλία & Αναφορές](#) (24)
- [Βιβλία](#) (14) [Διατριβές](#) (8) [Αναφορές](#) (2)

Επικέντρωση σε:

- [Τομείς του CERN](#) (5)
- [Πειραματική Φυσική \(EP\)](#) (1) [Θεωρητική Φυσική \(TH\)](#) (4)
- [Πειράματα του CERN](#) (2)
- [ALEPH](#) (1) [ISOLDE](#) (1)

ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΣΕΛΙΔΑ

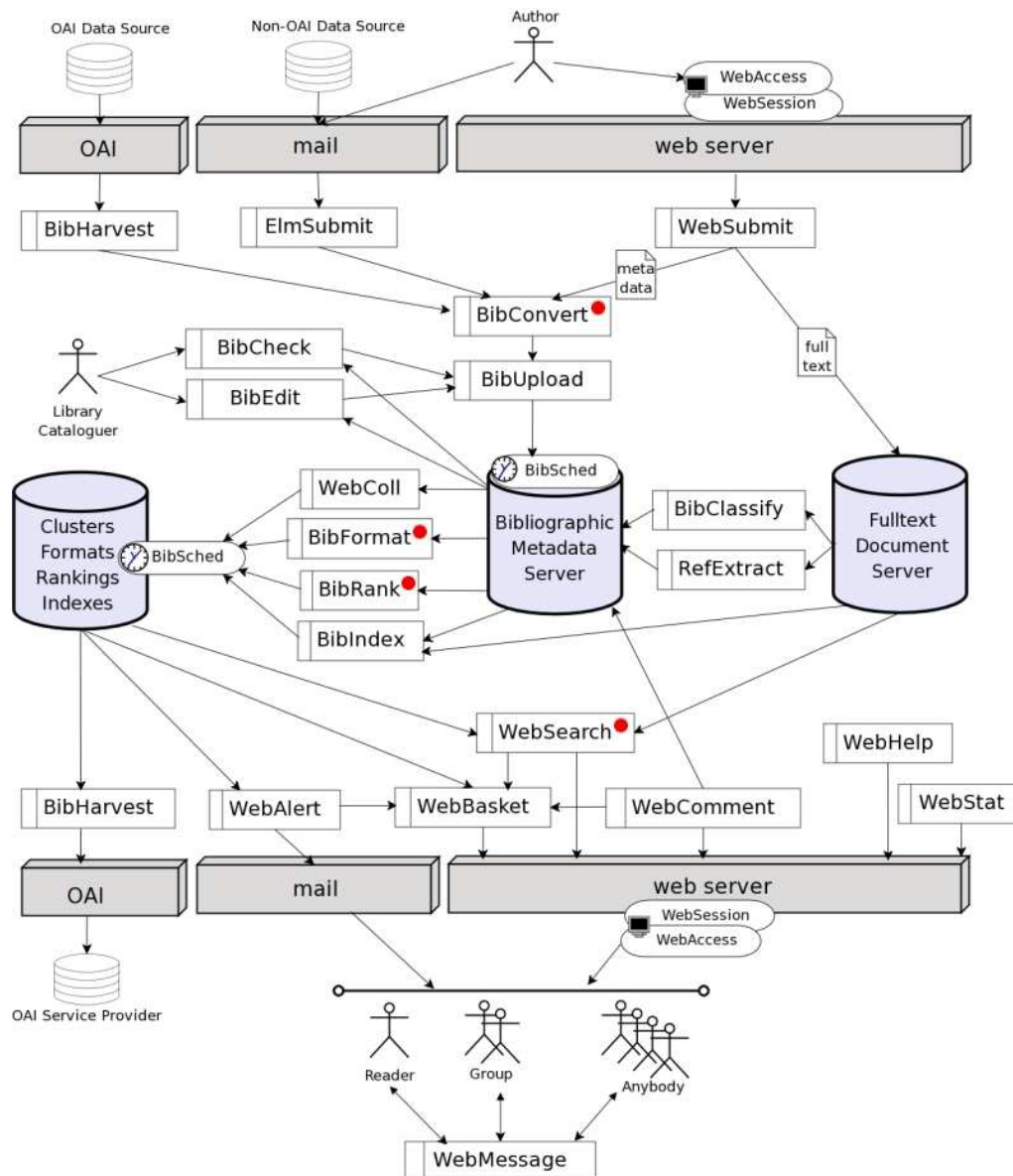
Καλώς ήλθατε στον δικτυακό τόπο του CDS Invenio, ενός δωρεάν εξυπηρετητή για έγγραφα προερχόμενο από το CERN. Είστε ευπρόσδεκτοι να εξερευνήσετε σε βάθος τις δυνατότητες που σας παρέχει ο δικτυακός αυτός τόπος.

CDS Invenio demo site in Greek, one of 20 available UI languages

INSPIRE Phase IIa (Q2 2008)

- ▷ *goal*: reproducing existing *user-level* functionalities
- ▷ **SPIRES data conversion** and load
 - ★ mapping to MARC
 - ★ load 760,000 literature and institute records
- ▷ **SPIRES citation analysis**
 - ★ self-cite elimination
 - ★ cite summary and related formats
- ▷ **SPIRES search syntax**
 - ★ supporting SPIRES traditional syntax
 - ★ fuzzy author search
- ▷ **output formats and site layout**
- ▷ ... May 2008: public release for alpha testers?

Phase IIa: Invenio/INSPIRE modules



Invenio software:

- ▶ 24 modules
- ▶ 160,000+ Python LOCs
- ▶ 470+ test cases

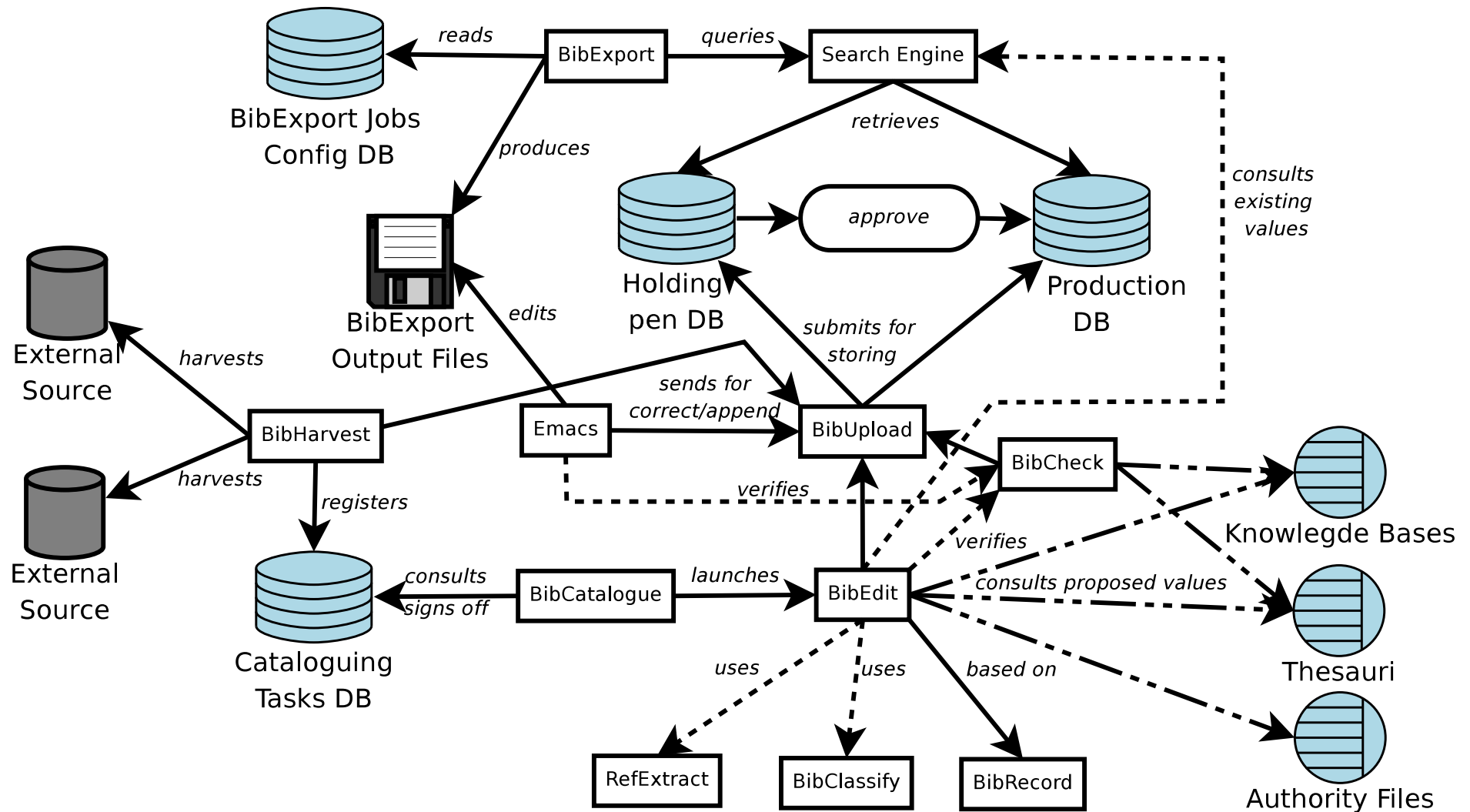
INSPIRE Phase IIa extensions:

- ▶ BibConvert extensions (SPIRES to MARC mapping)
- ▶ BibRank extensions (cite summary)
- ▶ WebSearch extensions (SPIRES search syntax)
- ▶ BibFormat, WebStyle (output style)

INSPIRE Phase IIb (Q2 2009)

- ▷ *goal*: reproducing existing *cataloguer-level* functionality
 - ▷ record editing interface
 - ▷ record checking tools
 - ▷ record maintenance tools
 - ▷ record inputting workflow
 - ▷ record harvesting workflow
 - ▷ knowledge bases
-
- ▷ ... Invenio traditionally relied on ALEPH cataloguing tool
 - ▷ ... build strong **native cataloguing tools** for INSPIRE
 - ▷ ... Q2 2009: full and complete replacement of SPIRES?

Phase IIb: Principal Developments



INSPIRE Phase III (from 2009)

- ▷ *goal: adding value* on top of reproduced SPIRES functionality
- ▷ ***user accounts** (even though most users are guests)
 - ★ *local accounts? authentication specific to every lab?
 - ★ *single sign-on and *certificates?
 - ★ e.g. CDS: 7,709 registered users, 67% non-CERN
- ▷ ***collaborative features**
 - ★ *user groups
 - ★ *sharable baskets
 - ★ *notification alerts
 - ★ *user recommendations
- ▷ ***keyword taxonomy, user tagging**
 - (* = already present in Invenio)

INSPIRE Phase II (from 2009) cont'd

- ▷ outspiring: *reaching outside* INSPIRE & HEP
 - ★ other community-based systems and publishers
 - ★ other related fields and open source communities
- ▷ **shared community author database**
- ▷ shared community inst/conf/expt databases
- ▷ ***full-text file** treatment
- ▷ advanced **linguistic context** indexing
- ▷ advanced content indexing of **plots** and **tables**
- ▷ ***conference contributions**, crawling static web sites
- ▷ extended **citation networks**, combined **impact metric**
- ▷ open **data mining** and **bibliometric** studies

(* = already present in Invenio)

Conclusions

▷ Invenio

- ★ more than 10 years of organic-growth development
- ★ software targeted for large repositories (1M+)
- ★ established open source community since 2002

▷ SPIRES

- ★ more than 30 years of high-quality data curation
- ★ established leading HEP information system

▷ INSPIRE

- ★ SPIRES and Invenio communities join forces
- ★ unique opportunity to build *the* single stop shop HEP information system
- ★ inspired by HEP, outspiring for other disciplines
- ★ a win-win situation