

Digital Library and Conferencing update

HEPiX at Cornell

Tim Smith – CERN/IT Department

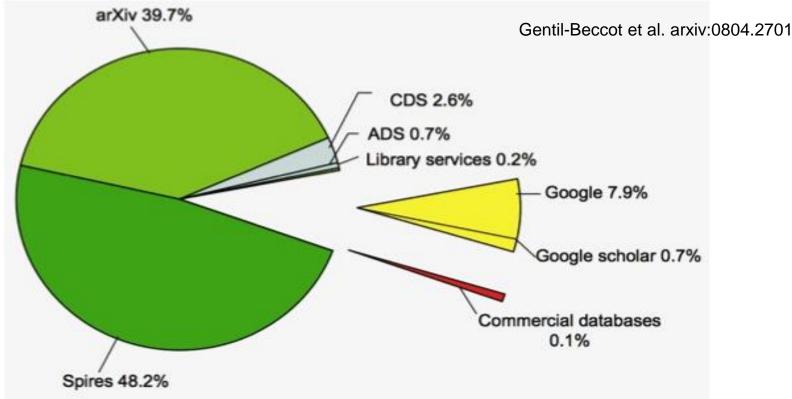
Contents



- INSPIRE Information System for HEP
 - Creation, Launch
 - Functionality
- Indico
 - Recent and planned developments
- Previous HEPiX presentations
 - INSPIRE: May 2008
 - Indico: May 2004, Oct 2005

Where do HEP scientists look for info?





- Survey of 2'000+ scientists (10% of community)
- OA tools answer scientists' information needs
- Google as proxy of arXiv, SPIRES, publishers

Building Blocks: A Quick Recap



SPIRES

- 1974 Birth of SPIRES HEP DataBase
 - IBM mainframe at SLAC run with DESY
- 1980s email interface
- 1991 Web interface to SPIRES
 - First US web server
- 1994 Citation services
- 2000 summary formats

Invenio

- 1993 CERN preprint server on web
- 1996 CERN Library on Web
- 2000 CERN Document server
- 2002 CDSWare released GPL
 - Python, Apache application server, MySQL
- 2006 CDSWare became Invenio
- 2010 http://invenio-software.org/







INSPIRE: A Recipe for Collaboration



- Meta-data from SPIRES
 - Noise free, high quality
 - Manually validated affiliation and citation data
- Technology from Invenio
 - Modern Open Source digital library SW
 - Good performance, scalability and feature set
- Feeds from information sources
 - arXiv, Publishers
 - Institutional repositories
- Digital Librarians from the labs











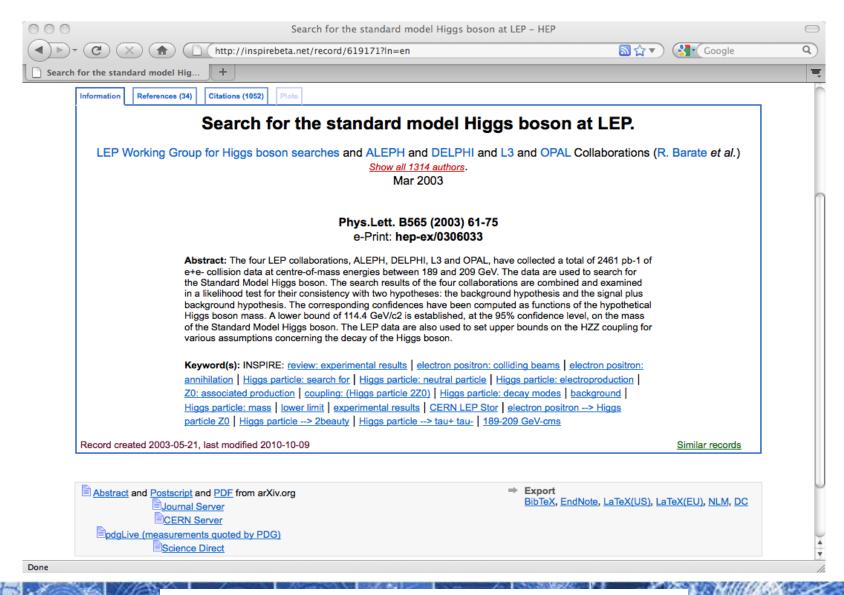
INSPIRE Launched





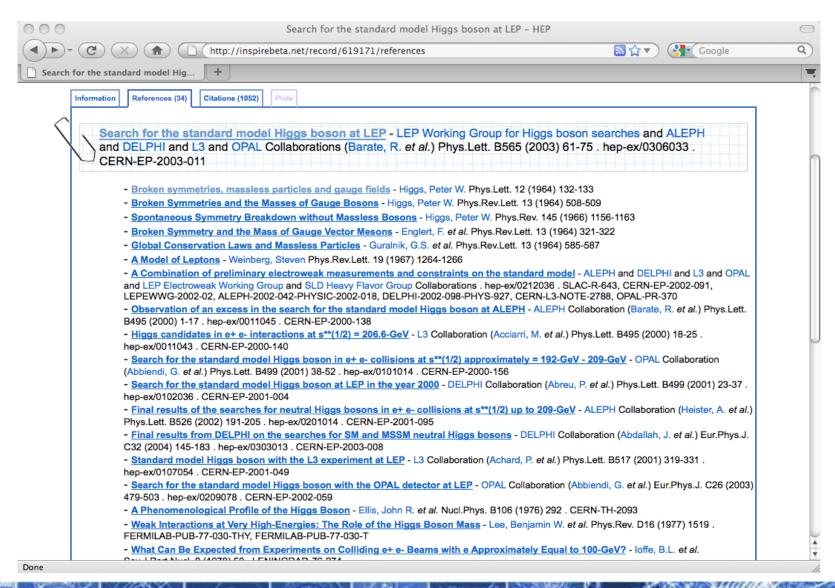
INSPIRE Detailed Records





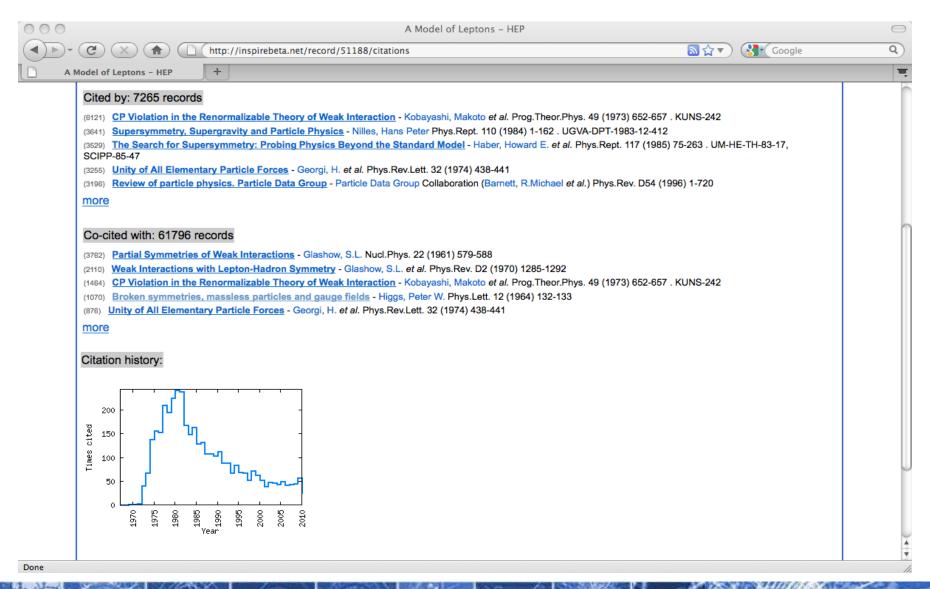
INSPIRE References





INSPIRE Citations





INSPIRE Author Page



Weinberg, Steven

Papers:

All papers (220) Published (156) Review (23) Conference (20)

Introductory (16) Book (11) Lectures (8)

Affiliations:

Texas U. (100)

Harvard U. (51)

MIT, LNS (26)

UC, Berkeley (20)

unknown (12)

Harvard-Smithsonian Ctr. Astrophys. (8)

Columbia U. (4)

Stanford U., Phys. Dept. (3)

Imperial Coll., London (2)

LLNL, Livermore (1)

Stanford U., ITP (1)

Frequent keywords:

spontaneous symmetry breaking (31)

gravitation (21)

renormalization (21)

unified field theory (20) quantum chromodynamics (17)

S-matrix (16)

FIELD THEORY: HIGHER-DIMENSIONAL (15)

strong interaction (15)

electroweak interaction (14)

field theory: scalar (14)

Princeton U. (2)

MIT (1)

Frequent co-authors:

Glashow, S.L. (3)

Schnitzer, Howard J. (3)

Biorken, J.D. (2)

Flauger, Raphael (2)

Gerstein, I.S. (2)

Hall, Lawrence J. (2)

Jackiw, R. (2)

Lane, Kenneth D. (2)

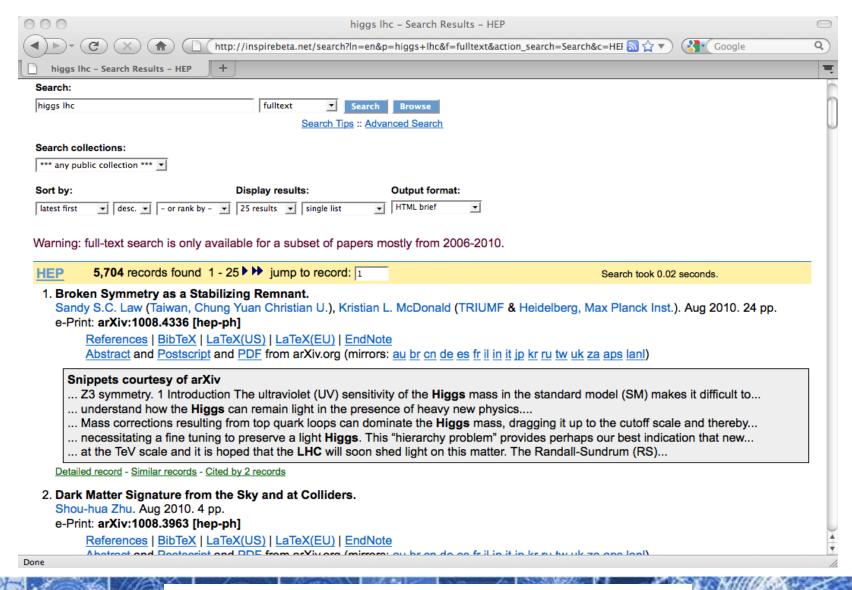
Lee, Benjamin W. (2)

Citations:

Citation summary results	All papers	Published only
Total number of citable papers analyzed:	<u>170</u>	<u>156</u>
Total number of citations:	47,121	46,594
Average citations per paper:	277.2	298.7
Breakdown of papers by citations:		
Renowned papers (500+)	<u>25</u>	<u>25</u>
Famous papers (250-499)	<u>20</u>	<u>20</u>
Verv well-known papers (100-249)	45	43

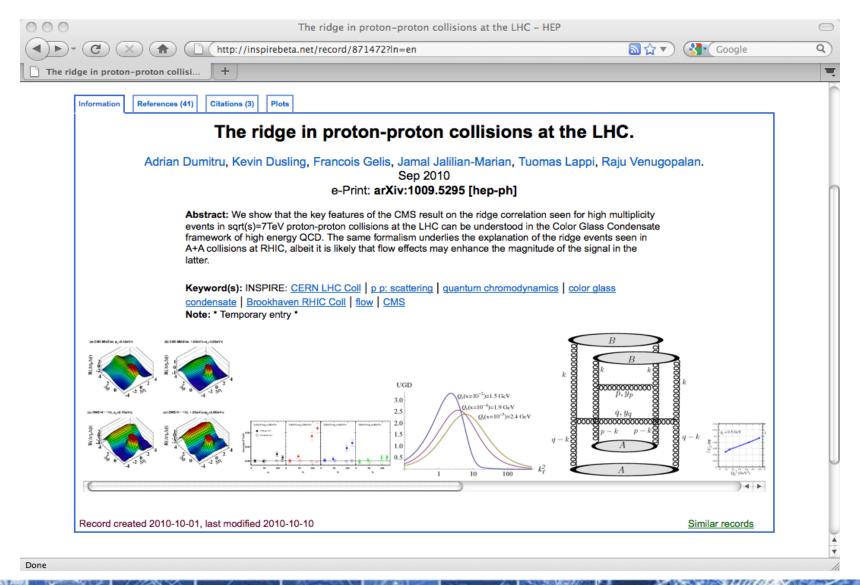
INSPIRE Fulltext Search





INSPIRE Plot Extraction





INSPIRE Machinery



- Parallelization and dispatching of intensive tasks
- Data Mining
- EU D4ScienceII project
 - OCR processing
 - Ocropus (open source)
 - De-skewing, text recognition, hOCR production (html), PDF
 - Fulltext indexing
 - Solr/Lucene, Hadoop, Process Execution Engine
 - Extension to semantic indexing with HEP ontology
 - Bibliometrics
 - Scientific impact e.g. H-Index

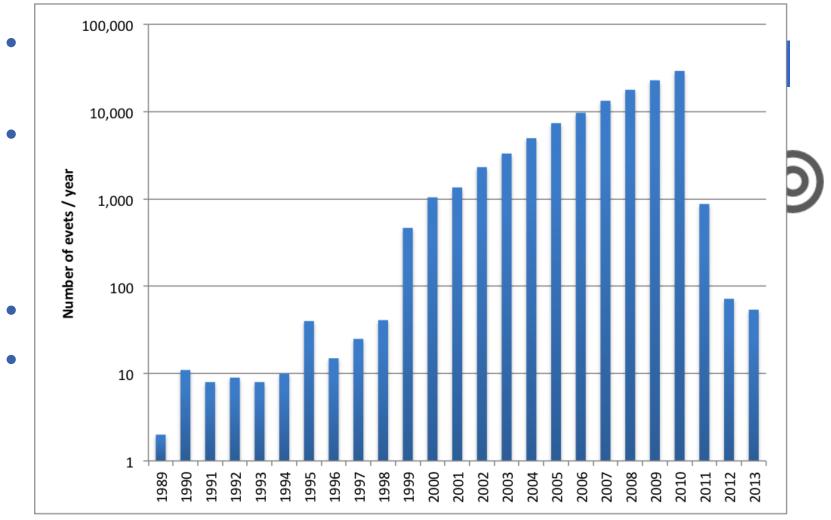
Next Steps: INSPIRE Users



- Personal libraries, alerts
- Claim-my-papers (with arXiv and ORCID)
- Submit theses and old non-arXiv material
- Attach non-text material (high level data files)
- OCR of library holdings (with D4Science-II)
- Advanced feeds (with ADS, arXiv, Publishers)
- Then
 - Crowdsourcing of keywording (tagging)
 - Semantic layer (did-you-mean and classification)
 - (Semantic) image search
 - Platform for high-level data preservation

Indico: A Quick Recap





Scale at CERN: 115k events; 699k files

Indico: Technology Advances



- New interface (summer 2008)
 - From REST to AJAX
- mod_python migration to mod_wsgi
 - Had to
 - Development stopped for 2 years
 - Officially discontinued June 2010
 - Wanted to
 - Improvements in performance and flexibility
 - Robustness and scalability
 - URL mapping (e.g. http://indico.cern.ch/event/BigEvent)
- Conversion server
 - Home grown to Neevia professional

Indico: Feature Advances



- Timezone aware (FNAL)
- ePayment
- Room booking
- Collaboration tools plugin framework
 - Integration (create / modify / delete / start)

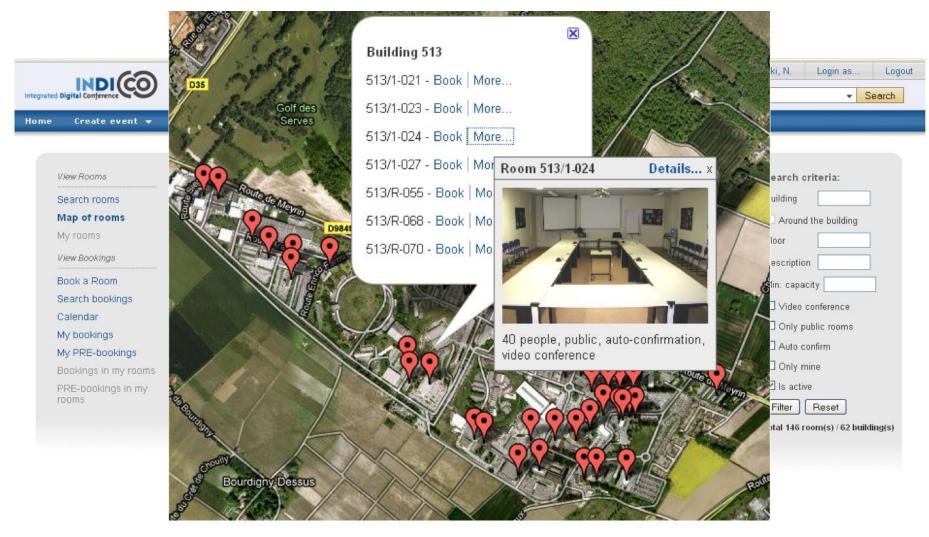




- Video conference: EVO, H.323/CERN MCU
- Web Lectures
- Requests
 - Lecture recording and Webcast
- Chat Rooms
 - Booked, launched and archived with event
- Eagerly anticipated: Drag and drop

Indico: Map of Rooms





Indico: Releases



- v0.97 (prod now), v0.98 (Beta Nov 2010)
- v1.0 (2H2011)
- Testing Framework
 - Source code analysis
 - Unit testing
 - Code coverage
 - Functional testing Selenium grid
- http://indico-software.org
- External instances
 - 100 known instances in 33 countries



Questions?

http://inspirebeta.net/

http://invenio-software.org
http://indico-software.org