Invenio @ INSPIRE

Samuele Kaplun

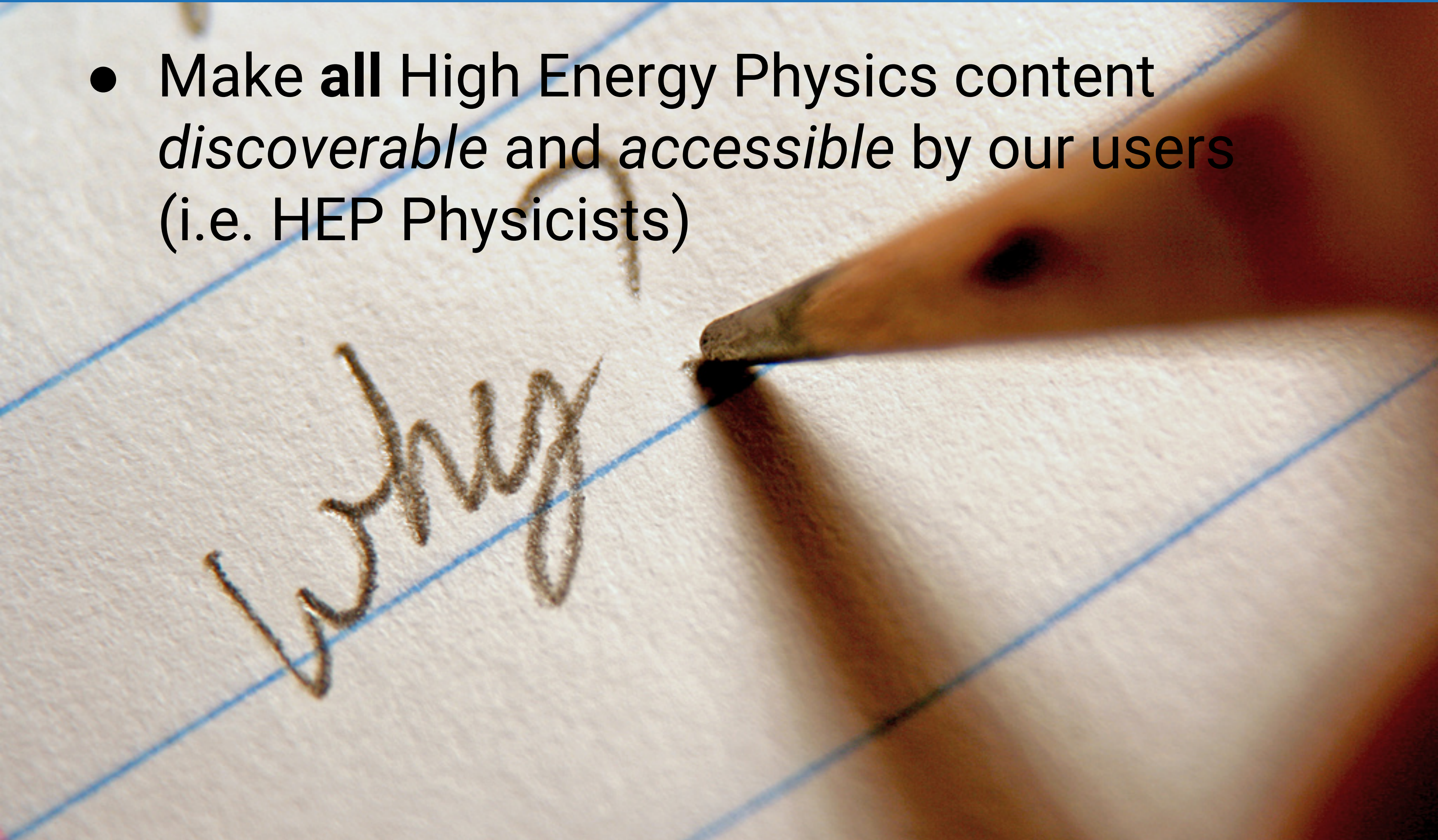
INSPIRE Service and Operations Manager

Invenio User Group Workshop

CERN, 12-15 October 2015

INSPIRE Mission

- Make **all** High Energy Physics content *discoverable* and *accessible* by our users (i.e. HEP Physicists)



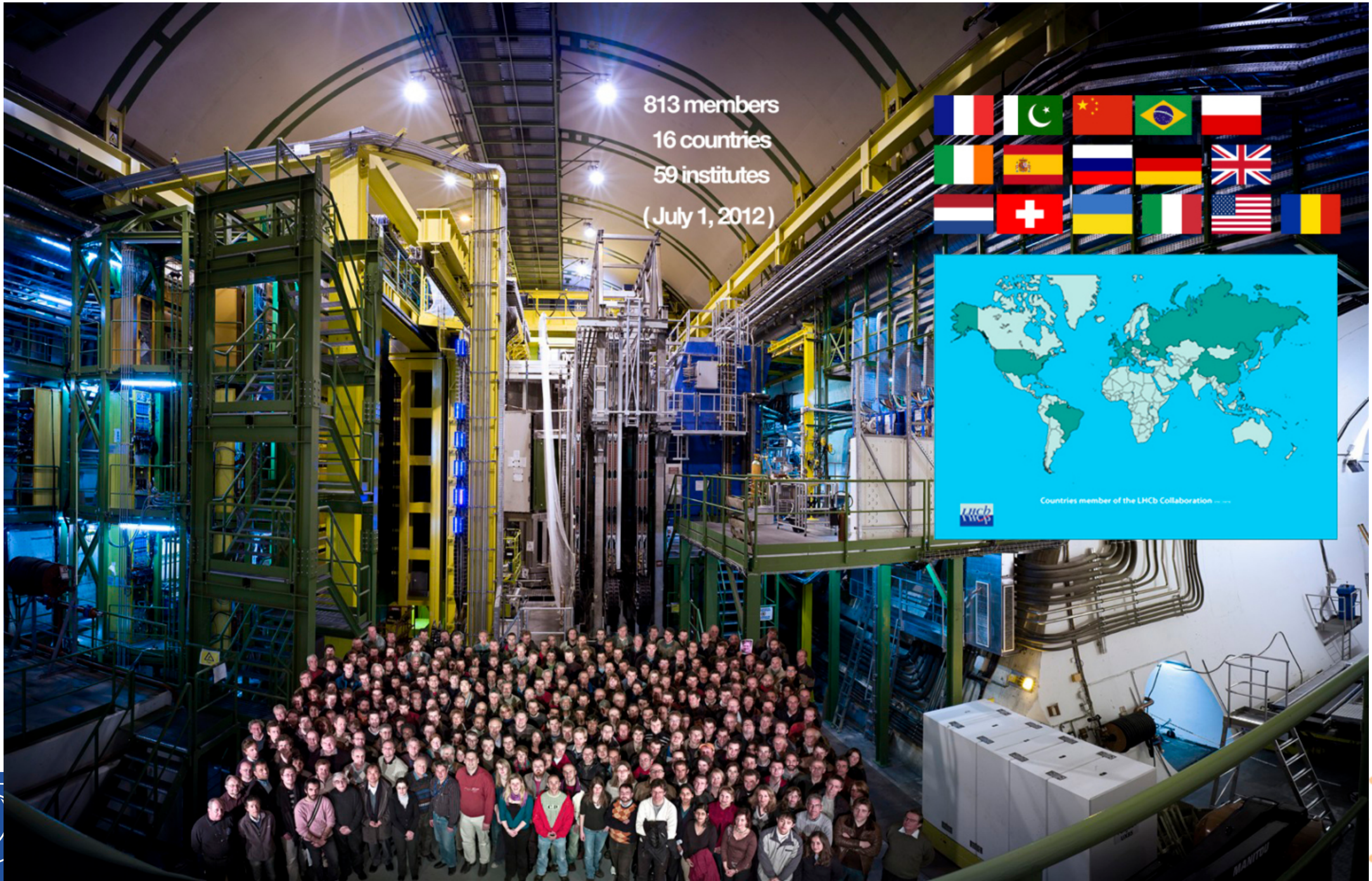
INSPIRE History

- 1969 **SPIRES** (SLAC)
- 1991 First accessible website in the US
- 2012 Ported to Invenio -> **INSPIRE**

Collaboration among: CERN, SLAC, Fermilab, DESY and IHEP

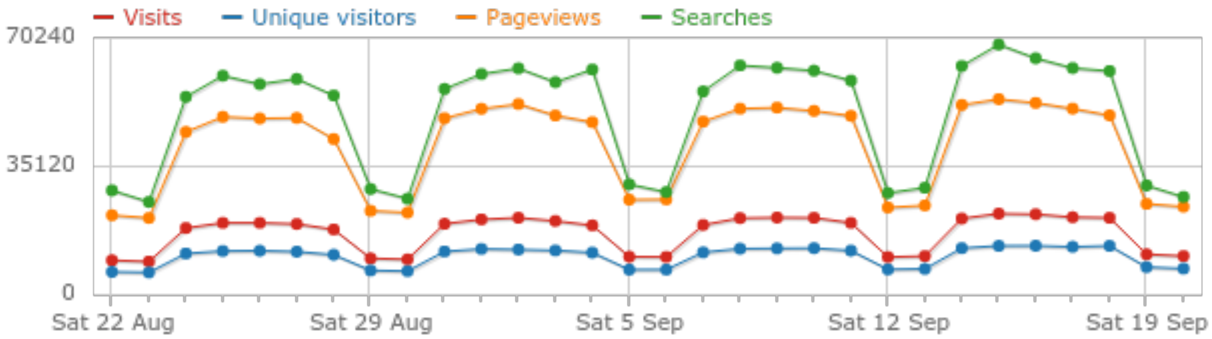
INSPIRE Users: Theoreticians

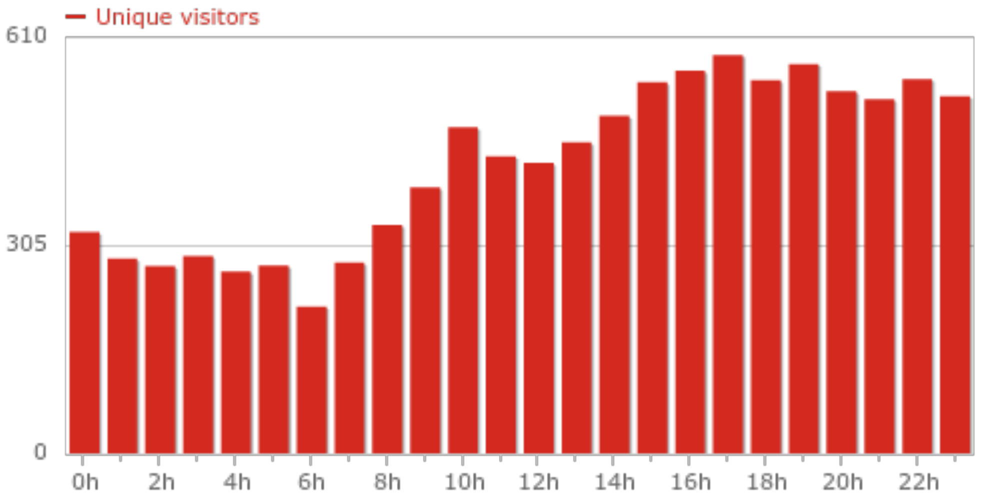
INSPIRE Users: Experimentalist



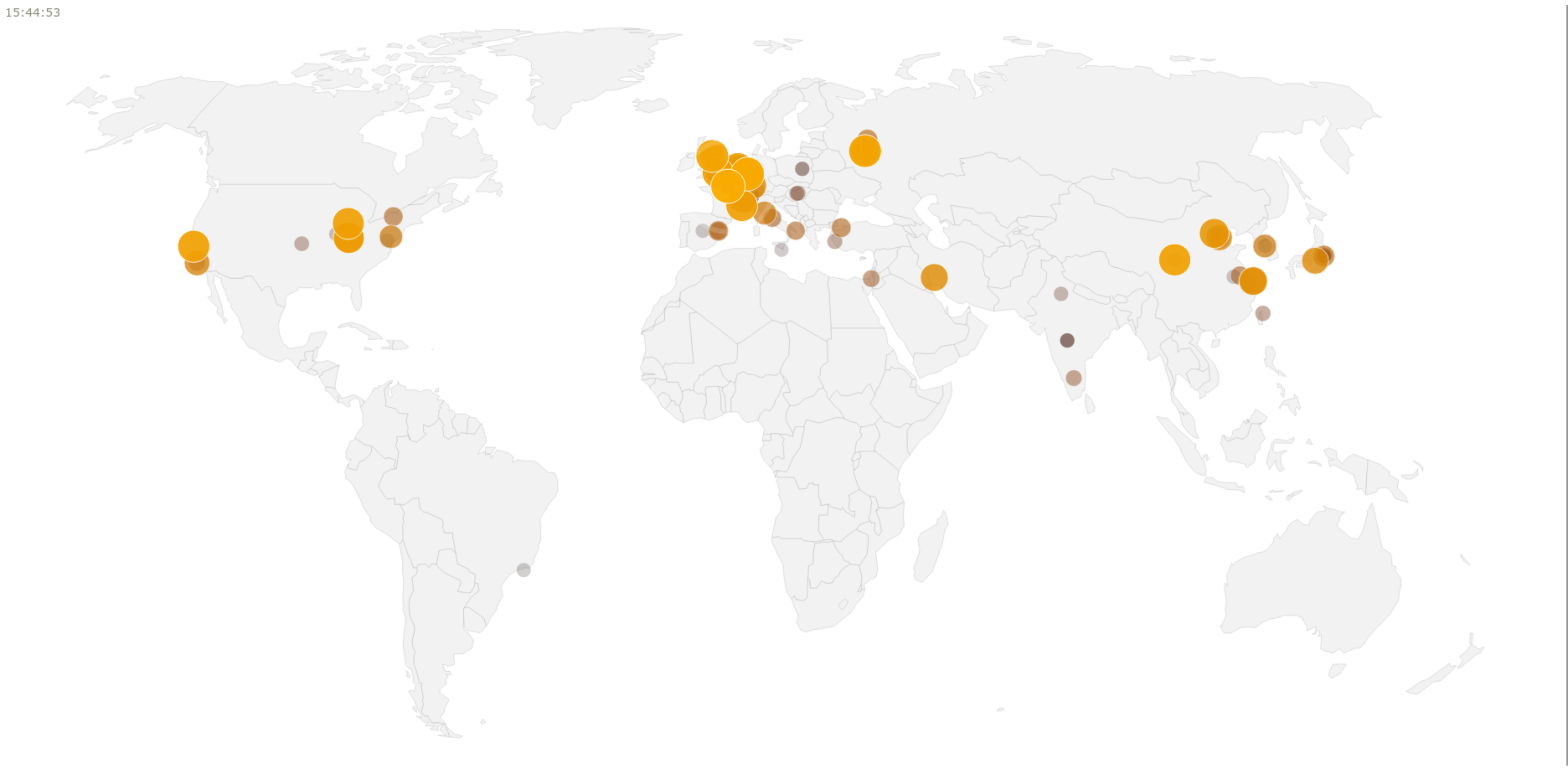
813 members
16 countries
59 institutes
(July 1, 2012)

INSPIRE Users: facts



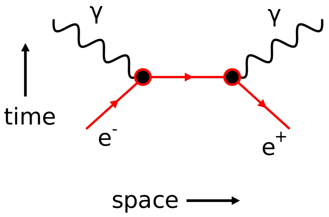


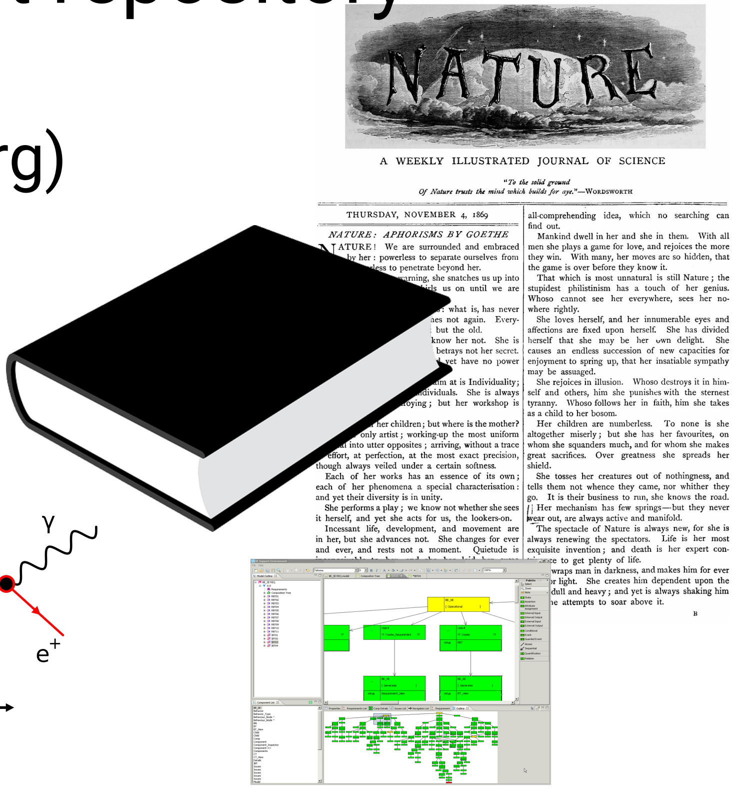
INSPIRE Users: facts



What is INSPIRE

- *High Energy Physics* subject repository
- Aggregator of
  - Preprints (mainly from arXiv.org)
  - Journal Articles
  - Notes
  - Conference Proceedings
  - Theses
  - Books
  - Scientific Data
  - Scientific Software
- 1M+ records




NATURE

A WEEKLY ILLUSTRATED JOURNAL OF SCIENCE

"To the solid ground
Of Nature trusts the mind which builds for aye."—Wordsworth

THURSDAY, NOVEMBER 4, 1869

NATURE: APHORISMS BY GOETHE

NATURE! We are surrounded and embraced
by her : powerless to separate ourselves from
less to penetrate beyond her.
warning, she snatches us up into
hirls us on until we are

: what is, has never
nes not again. Every-
but the old.
know her not. She is
betrays not her secret.
d yet have no power

im at is Individuality;
ndividuals. She is always
troying; but her workshop is

her children; but where is the mother?
only artist; working-up the most uniform
al into utter opposites; arriving, without a trace
effort, at perfection, at the most exact precision,
though always veiled under a certain softness.

Each of her works has an essence of its own; each of her phenomena a special characterisation: and yet their diversity is in unity.

She performs a play; we know not whether she sees it herself, and yet she acts for us, the lookers-on.

Incessant life, development, and movement are in her, but she advances not. She changes for ever and ever, and rests not a moment. Quietude is

all-comprehending idea, which no searching can find out.

Mankind dwell in her and she in them. With all men she plays a game for love, and rejoices the more they win. With many, her moves are so hidden, that the game is over before they know it.

That which is most unnatural is still Nature; the stupidest philistinism has a touch of her genius. Whoso cannot see her everywhere, sees her nowhere rightly.

She loves herself, and her innumerable eyes and affections are fixed upon herself. She has divided herself that she may be her own delight. She causes an endless succession of new capacities for enjoyment to spring up, that her insatiable sympathy may be assuaged.

She rejoices in illusion. Whoso destroys it in himself and others, him she punishes with the sternest tyranny. Whoso follows her in faith, him she takes as a child to her bosom.

Her children are numberless. To none is she altogether miserly; but she has her favourites, on whom she squanders much, and for whom she makes great sacrifices. Over greatness she spreads her shield.

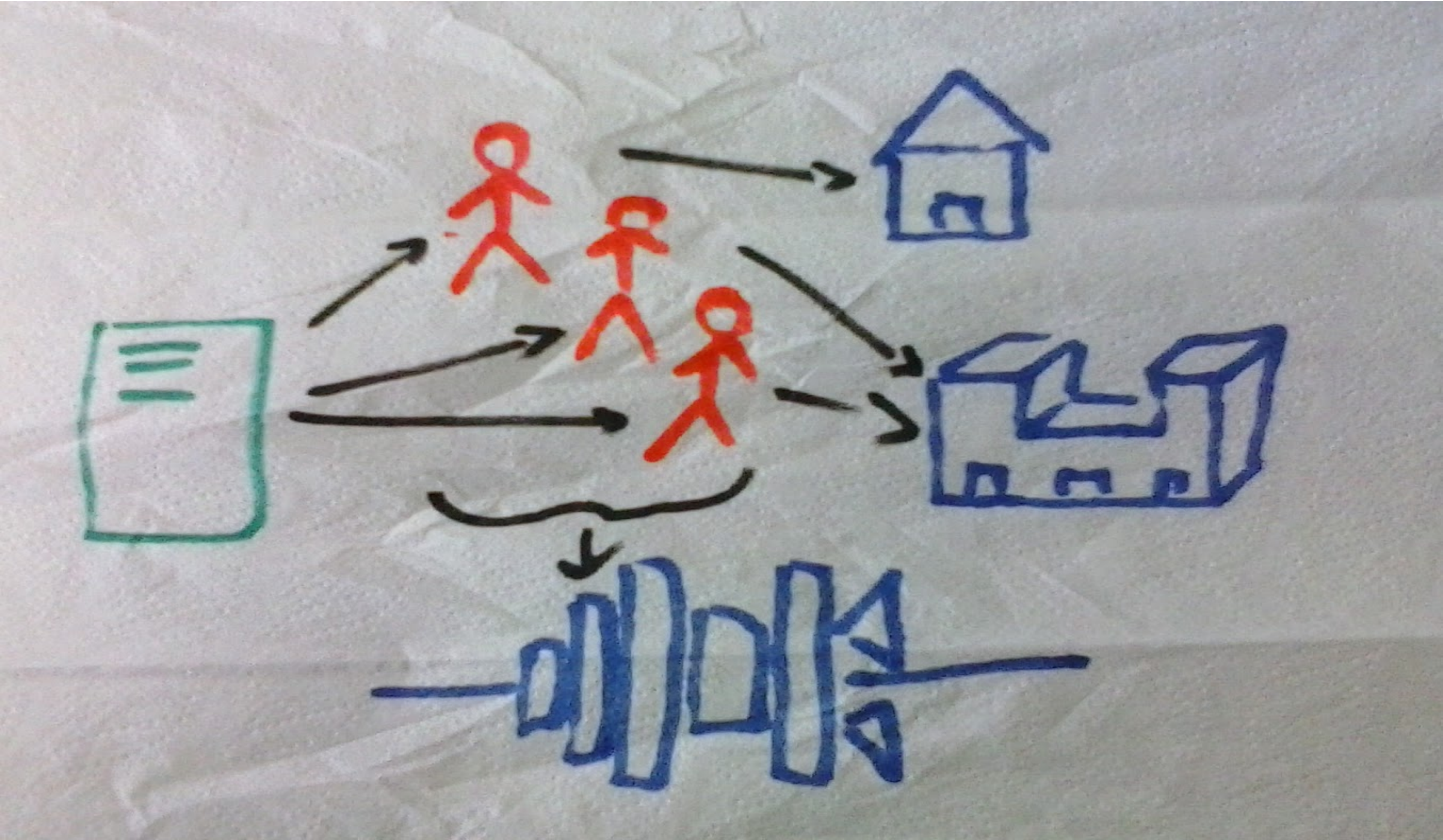
She tosses her creatures out of nothingness, and tells them not whence they came, nor whither they go. It is their business to run, she knows the road.

Her mechanism has few springs—but they never wear out, are always active and manifold.

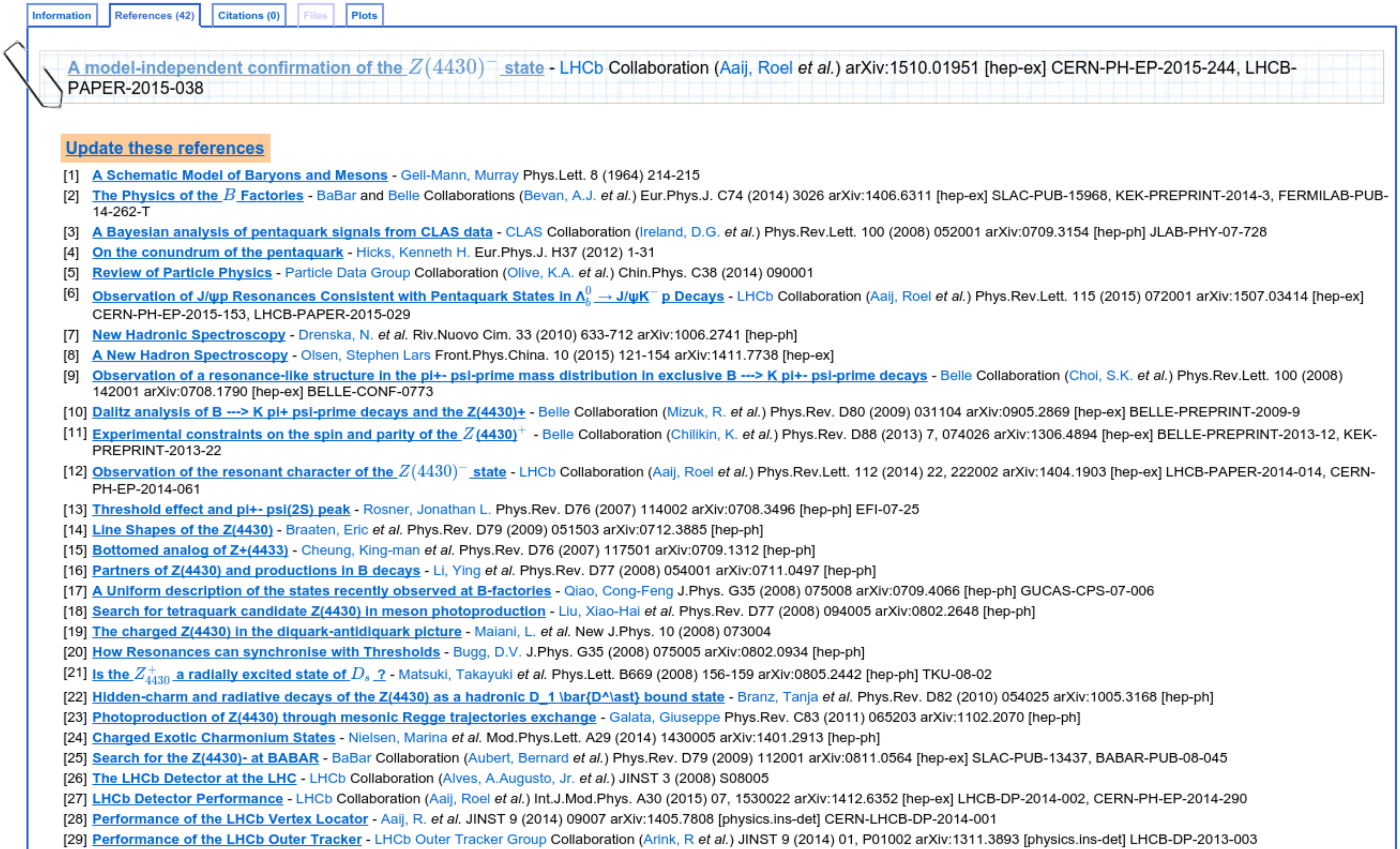
The spectacle of Nature is always new, for she is always renewing the spectators. Life is her most exquisite invention; and death is her expert con-
ce to get plenty of life.
wraps man in darkness, and makes him for ever
or light. She creates him dependent upon the
dull and heavy; and yet is always shaking him
he attempts to soar above it.

B

INSPIRE data model



Advanced functionalities: Citations & References

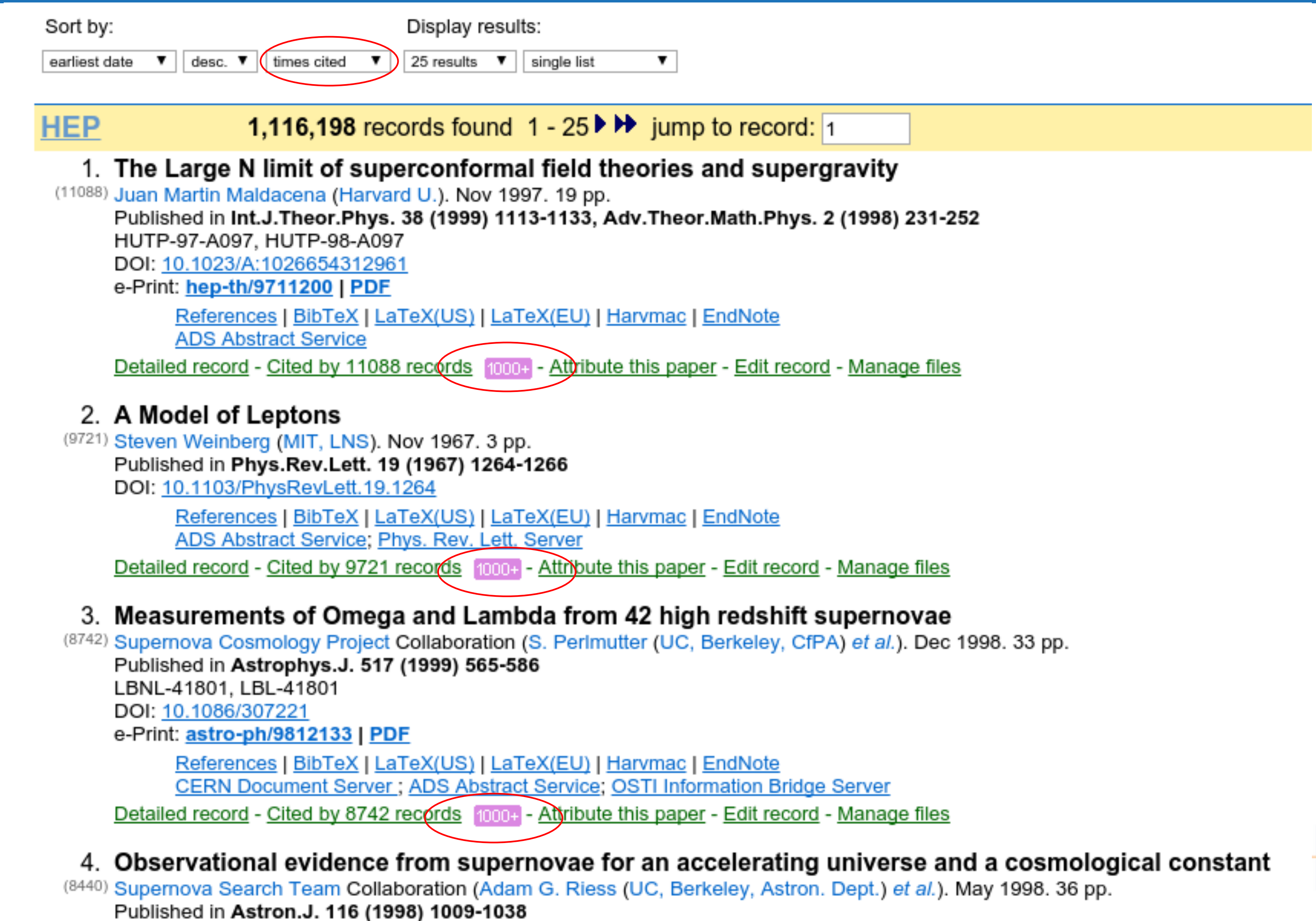


A model-independent confirmation of the $Z(4430)^-$ state - LHCb Collaboration (Aaij, Roel *et al.*) arXiv:1510.01951 [hep-ex] CERN-PH-EP-2015-244, LHCB-PAPER-2015-038

Update these references

[1] A Schematic Model of Baryons and Mesons - Gell-Mann, Murray Phys.Lett. 8 (1964) 214-215
[2] The Physics of the B Factories - BaBar and Belle Collaborations (Bevan, A.J. *et al.*) Eur.Phys.J. C74 (2014) 3026 arXiv:1406.6311 [hep-ex] SLAC-PUB-15968, KEK-PREPRINT-2014-3, FERMILAB-PUB-14-262-T
[3] A Bayesian analysis of pentaquark signals from CLAS data - CLAS Collaboration (Ireland, D.G. *et al.*) Phys.Rev.Lett. 100 (2008) 052001 arXiv:0709.3154 [hep-ph] JLAB-PHY-07-728
[4] On the conundrum of the pentaquark - Hicks, Kenneth H. Eur.Phys.J. H37 (2012) 1-31
[5] Review of Particle Physics - Particle Data Group Collaboration (Olive, K.A. *et al.*) Chin.Phys. C38 (2014) 090001
[6] Observation of J/ψp Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays - LHCb Collaboration (Aaij, Roel *et al.*) Phys.Rev.Lett. 115 (2015) 072001 arXiv:1507.03414 [hep-ex] CERN-PH-EP-2015-153, LHCB-PAPER-2015-029
[7] New Hadronic Spectroscopy - Drenska, N. *et al.* Riv.Nuovo Cim. 33 (2010) 633-712 arXiv:1006.2741 [hep-ph]
[8] A New Hadron Spectroscopy - Olsen, Stephen Lars Front.Phys.China. 10 (2015) 121-154 arXiv:1411.7738 [hep-ex]
[9] Observation of a resonance-like structure in the pi+- psi-prime mass distribution in exclusive B ---> K pi+- psi-prime decays - Belle Collaboration (Choi, S.K. *et al.*) Phys.Rev.Lett. 100 (2008) 142001 arXiv:0708.1790 [hep-ex] BELLE-CONF-0773
[10] Dalitz analysis of B ---> K pi+ psi-prime decays and the Z(4430)+ - Belle Collaboration (Mizuk, R. *et al.*) Phys.Rev. D80 (2009) 031104 arXiv:0905.2869 [hep-ex] BELLE-PREPRINT-2009-9
[11] Experimental constraints on the spin and parity of the $Z(4430)^+$ - Belle Collaboration (Chilikin, K. *et al.*) Phys.Rev. D88 (2013) 7, 074026 arXiv:1306.4894 [hep-ex] BELLE-PREPRINT-2013-12, KEK-PREPRINT-2013-22
[12] Observation of the resonant character of the $Z(4430)^-$ state - LHCb Collaboration (Aaij, Roel *et al.*) Phys.Rev.Lett. 112 (2014) 22, 222002 arXiv:1404.1903 [hep-ex] LHCB-PAPER-2014-014, CERN-PH-EP-2014-061
[13] Threshold effect and pi+- psi(2S) peak - Rosner, Jonathan L. Phys.Rev. D76 (2007) 114002 arXiv:0708.3496 [hep-ph] EFI-07-25
[14] Line Shapes of the Z(4430) - Braaten, Eric *et al.* Phys.Rev. D79 (2009) 051503 arXiv:0712.3885 [hep-ph]
[15] Bottomed analog of Z+(4433) - Cheung, King-man *et al.* Phys.Rev. D76 (2007) 117501 arXiv:0709.1312 [hep-ph]
[16] Partners of Z(4430) and productions in B decays - Li, Ying *et al.* Phys.Rev. D77 (2008) 054001 arXiv:0711.0497 [hep-ph]
[17] A Uniform description of the states recently observed at B-factories - Qiao, Cong-Feng J.Phys. G35 (2008) 075008 arXiv:0709.4066 [hep-ph] GUCAS-CPS-07-006
[18] Search for tetraquark candidate Z(4430) in meson photoproduction - Liu, Xiao-Hai *et al.* Phys.Rev. D77 (2008) 094005 arXiv:0802.2648 [hep-ph]
[19] The charged Z(4430) in the diquark-antidiquark picture - Maiani, L. *et al.* New J.Phys. 10 (2008) 073004
[20] How Resonances can synchronise with Thresholds - Bugg, D.V. J.Phys. G35 (2008) 075005 arXiv:0802.0934 [hep-ph]
[21] Is the Z^+_{4430} a radially excited state of D_s ? - Matsuki, Takayuki *et al.* Phys.Lett. B669 (2008) 156-159 arXiv:0805.2442 [hep-ph] TKU-08-02
[22] Hidden-charm and radiative decays of the Z(4430) as a hadronic D_1 \bar{D^\ast} bound state - Branz, Tanja *et al.* Phys.Rev. D82 (2010) 054025 arXiv:1005.3168 [hep-ph]
[23] Photoproduction of Z(4430) through mesonic Regge trajectories exchange - Galata, Giuseppe Phys.Rev. C83 (2011) 065203 arXiv:1102.2070 [hep-ph]
[24] Charged Exotic Charmonium States - Nielsen, Marina *et al.* Mod.Phys.Lett. A29 (2014) 1430005 arXiv:1401.2913 [hep-ph]
[25] Search for the Z(4430)- at BABAR - BaBar Collaboration (Aubert, Bernard *et al.*) Phys.Rev. D79 (2009) 112001 arXiv:0811.0564 [hep-ex] SLAC-PUB-13437, BABAR-PUB-08-045
[26] The LHCb Detector at the LHC - LHCb Collaboration (Alves, A.Augusto, Jr. *et al.*) JINST 3 (2008) S08005
[27] LHCb Detector Performance - LHCb Collaboration (Aaij, Roel *et al.*) Int.J.Mod.Phys. A30 (2015) 07, 1530022 arXiv:1412.6352 [hep-ex] LHCB-DP-2014-002, CERN-PH-EP-2014-290
[28] Performance of the LHCb Vertex Locator - Aaij, R. *et al.* JINST 9 (2014) 09007 arXiv:1405.7808 [physics.ins-det] CERN-LHCB-DP-2014-001
[29] Performance of the LHCb Outer Tracker - LHCb Outer Tracker Group Collaboration (Arink, R *et al.*) JINST 9 (2014) 01, P01002 arXiv:1311.3893 [physics.ins-det] LHCB-DP-2013-003

Advanced functionalities: Citations & References



Sort by: earliest date ▼ desc. ▼ times cited ▼

Display results: 25 results ▼ single list ▼

HEP **1,116,198** records found 1 - 25 ▶ ⏩ jump to record: 1

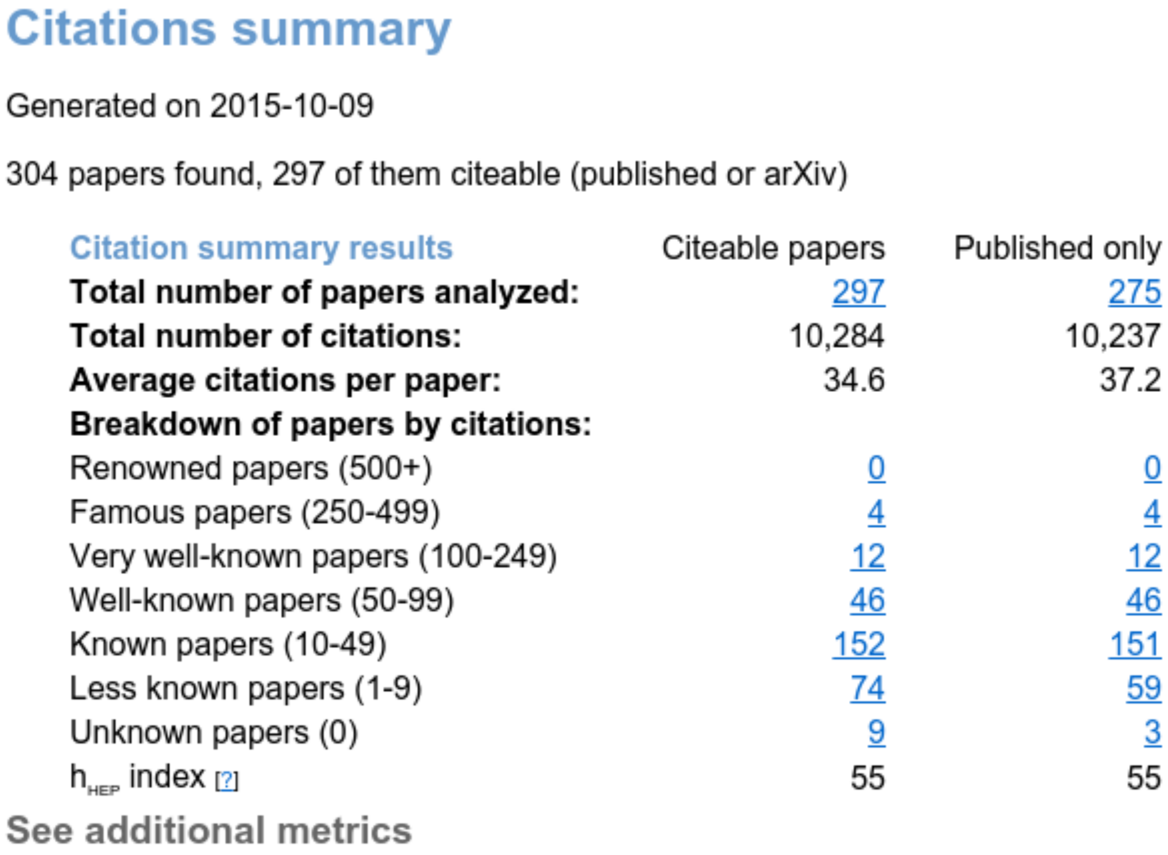
1. **The Large N limit of superconformal field theories and supergravity**
(11088) Juan Martin Maldacena (Harvard U.). Nov 1997. 19 pp.
Published in **Int.J.Theor.Phys. 38 (1999) 1113-1133, Adv.Theor.Math.Phys. 2 (1998) 231-252**
HUTP-97-A097, HUTP-98-A097
DOI: 10.1023/A:1026654312961
e-Print: **hep-th/9711200** | **PDF**
References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
ADS Abstract Service
Detailed record - Cited by 11088 records 1000+ - Attribute this paper - Edit record - Manage files

2. **A Model of Leptons**
(9721) Steven Weinberg (MIT, LNS). Nov 1967. 3 pp.
Published in **Phys.Rev.Lett. 19 (1967) 1264-1266**
DOI: 10.1103/PhysRevLett.19.1264
References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
ADS Abstract Service; Phys. Rev. Lett. Server
Detailed record - Cited by 9721 records 1000+ - Attribute this paper - Edit record - Manage files

3. **Measurements of Omega and Lambda from 42 high redshift supernovae**
(8742) Supernova Cosmology Project Collaboration (S. Perlmutter (UC, Berkeley, CfPA) *et al.*). Dec 1998. 33 pp.
Published in **Astrophys.J. 517 (1999) 565-586**
LBNL-41801, LBL-41801
DOI: 10.1086/307221
e-Print: **astro-ph/9812133** | **PDF**
References | BibTeX | LaTeX(US) | LaTeX(EU) | Harvmac | EndNote
CERN Document Server ; ADS Abstract Service; OSTI Information Bridge Server
Detailed record - Cited by 8742 records 1000+ - Attribute this paper - Edit record - Manage files

4. **Observational evidence from supernovae for an accelerating universe and a cosmological constant**
(8440) Supernova Search Team Collaboration (Adam G. Riess (UC, Berkeley, Astron. Dept.) *et al.*). May 1998. 36 pp.
Published in **Astron.J. 116 (1998) 1009-1038**

Advanced functionalities: Citations & References



Citations summary

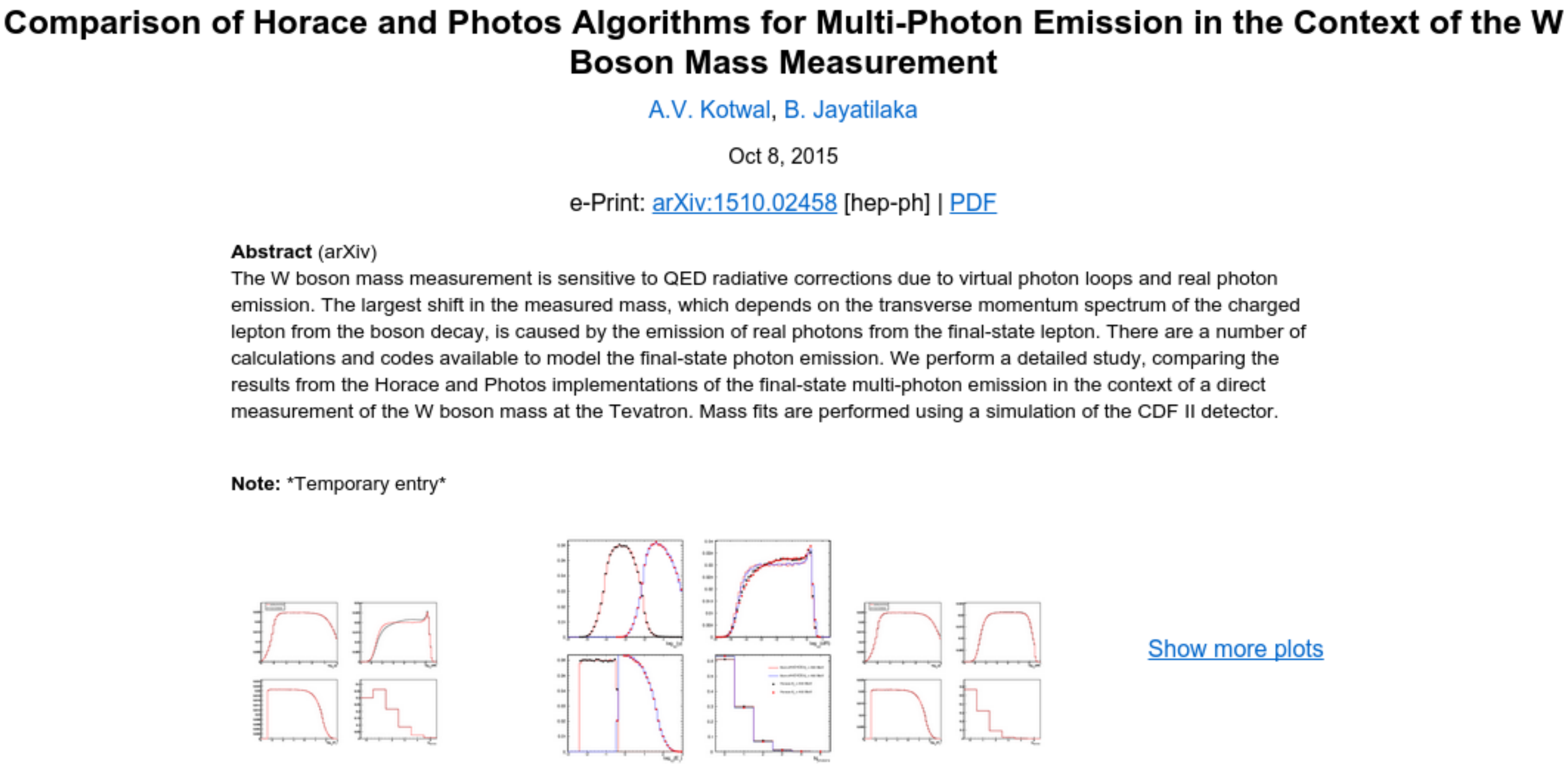
Generated on 2015-10-09

304 papers found, 297 of them citeable (published or arXiv)

Citation summary results	Citeable papers	Published only
Total number of papers analyzed:	297	275
Total number of citations:	10,284	10,237
Average citations per paper:	34.6	37.2
Breakdown of papers by citations:		
Renowned papers (500+)	0	0
Famous papers (250-499)	4	4
Very well-known papers (100-249)	12	12
Well-known papers (50-99)	46	46
Known papers (10-49)	152	151
Less known papers (1-9)	74	59
Unknown papers (0)	9	3
h_{HEP} index [?]	55	55

See additional metrics

Advanced functionalities: Plots



Comparison of Horace and Photos Algorithms for Multi-Photon Emission in the Context of the W Boson Mass Measurement

A.V. Kotwal, B. Jayatilaka

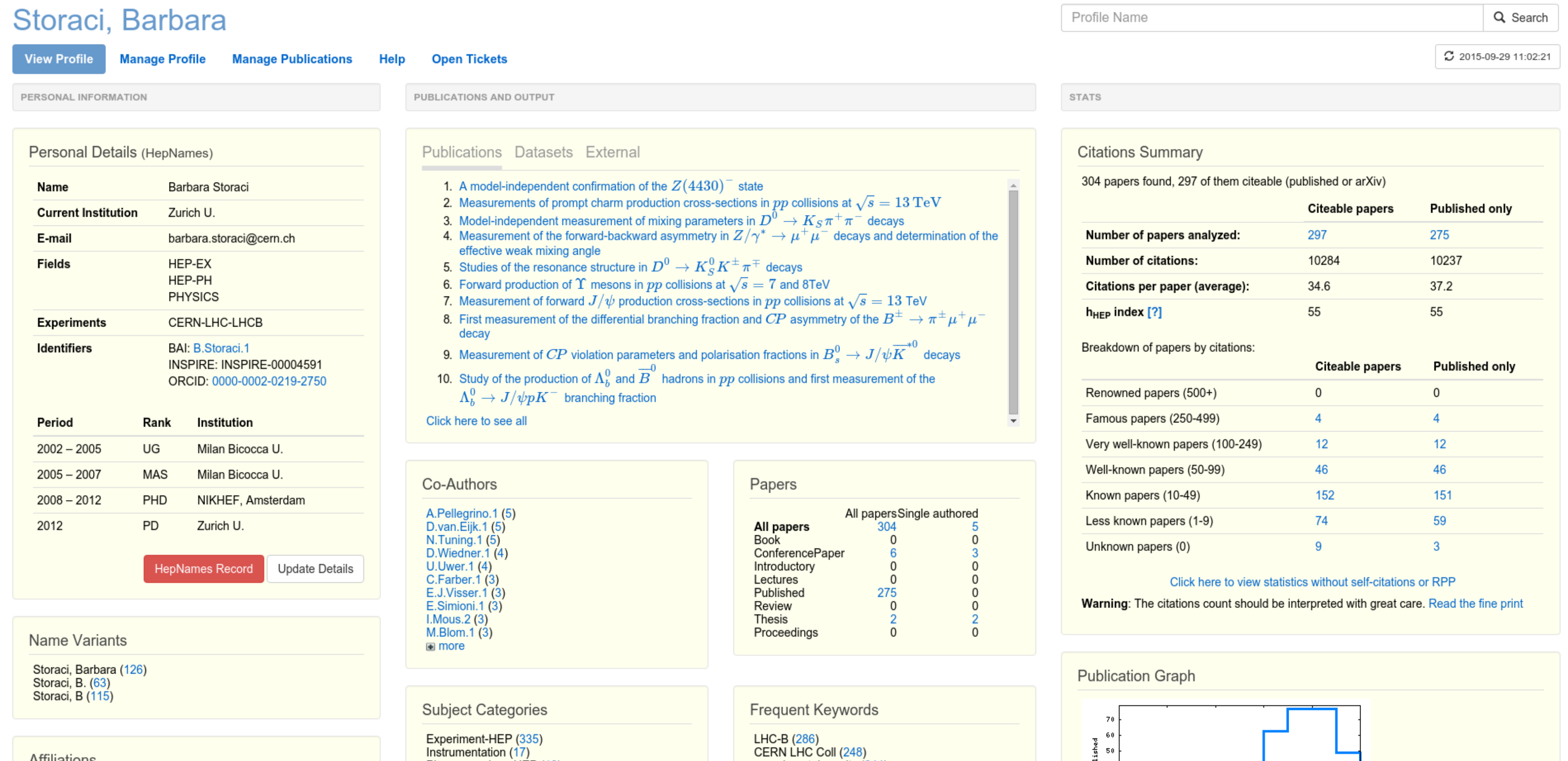
Oct 8, 2015

e-Print: arXiv:1510.02458 [hep-ph] | PDF

Abstract (arXiv)
The W boson mass measurement is sensitive to QED radiative corrections due to virtual photon loops and real photon emission. The largest shift in the measured mass, which depends on the transverse momentum spectrum of the charged lepton from the boson decay, is caused by the emission of real photons from the final-state lepton. There are a number of calculations and codes available to model the final-state photon emission. We perform a detailed study, comparing the results from the Horace and Photos implementations of the final-state multi-photon emission in the context of a direct measurement of the W boson mass at the Tevatron. Mass fits are performed using a simulation of the CDF II detector.

Note: *Temporary entry*

Advanced functionalities: Author profiles



Storaci, Barbara

Profile Name | Search

View Profile | Manage Profile | Manage Publications | Help | Open Tickets

2015-09-29 11:02:21

PERSONAL INFORMATION

Personal Details (HepNames)

Name	Barbara Storaci
Current Institution	Zurich U.
E-mail	barbara.storaci@cern.ch
Fields	HEP-EX HEP-PH PHYSICS
Experiments	CERN-LHC-LHCB
Identifiers	BAI: B.Storaci.1 INSPIRE: INSPIRE-00004591 ORCID: 0000-0002-0219-2750

Period	Rank	Institution
2002 – 2005	UG	Milan Bicocca U.
2005 – 2007	MAS	Milan Bicocca U.
2008 – 2012	PHD	NIKHEF, Amsterdam
2012	PD	Zurich U.

HepNames Record | Update Details

Name Variants

Storaci, Barbara (126)
Storaci, B. (63)
Storaci, B (115)

Affiliations

PUBLICATIONS AND OUTPUT

Publications | Datasets | External

1. A model-independent confirmation of the $Z(4430)^-$ state
2. Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$
3. Model-independent measurement of mixing parameters in $D^0 \to K_S\pi^+\pi^-$ decays
4. Measurement of the forward-backward asymmetry in $Z/\gamma^* \to \mu^+\mu^-$ decays and determination of the effective weak mixing angle
5. Studies of the resonance structure in $D^0 \to K^0_S K^\pm\pi^\mp$ decays
6. Forward production of Υ mesons in pp collisions at $\sqrt{s} = 7$ and 8TeV
7. Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV
8. First measurement of the differential branching fraction and CP asymmetry of the $B^\pm \to \pi^\pm\mu^+\mu^-$ decay
9. Measurement of CP violation parameters and polarisation fractions in $B^0_s \to J/\psi\overline{K}^{*0}$ decays
10. Study of the production of Λ^0_b and $\overline{B}^0$ hadrons in pp collisions and first measurement of the $\Lambda^0_b \to J/\psi pK^-$ branching fraction

Click here to see all

Co-Authors

A.Pellegrino.1 (5)
D.van.Eijk.1 (5)
N.Tuning.1 (5)
D.Wiedner.1 (4)
U.Uwer.1 (4)
C.Farber.1 (3)
E.J.Visser.1 (3)
E.Simioni.1 (3)
I.Mous.2 (3)
M.Blom.1 (3)
more

Papers

	All papers	Single authored
All papers	304	5
Book	0	0
ConferencePaper	6	3
Introductory	0	0
Lectures	0	0
Published	275	0
Review	0	0
Thesis	2	2
Proceedings	0	0

Subject Categories

Experiment-HEP (335)
Instrumentation (17)

Frequent Keywords

LHC-B (286)
CERN LHC Coll (248)

STATS

Citations Summary

304 papers found, 297 of them citeable (published or arXiv)

	Citeable papers	Published only
Number of papers analyzed:	297	275
Number of citations:	10284	10237
Citations per paper (average):	34.6	37.2
h$_{\mathrm{HEP}}$ index [?]	55	55

Breakdown of papers by citations:

	Citeable papers	Published only
Renowned papers (500+)	0	0
Famous papers (250-499)	4	4
Very well-known papers (100-249)	12	12
Well-known papers (50-99)	46	46
Known papers (10-49)	152	151
Less known papers (1-9)	74	59
Unknown papers (0)	9	3

Click here to view statistics without self-citations or RPP

Warning: The citations count should be interpreted with great care. Read the fine print

Publication Graph

Challenges

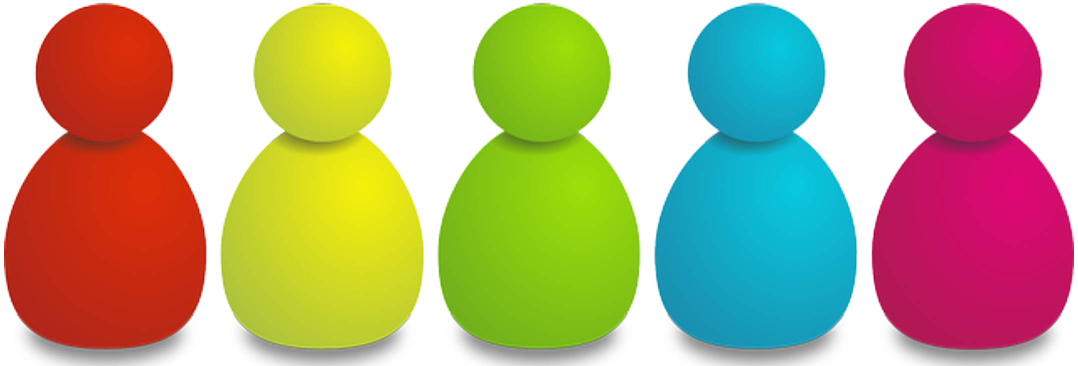
- Collaboration papers: ~3000 authors, i.e. **1MB** of metadata per record!
- **Heterogeneous metadata** from various sources to be normalized and merged (e.g. preprint Vs. published version)
- **Thorough users** spotting missing citations :-)

Objectives of our development: Maximizing curators efficiency

- Cataloguing tools
  - **Automatic duplicate records identification** (*invenio-matcher*)
  - **Advanced record editor** (schema-based, autocompletion everywhere, mouse-free, supporting record merging, integrated with history and ticketing system) (*invenio-editor to come*)
  - **Batch record editor** (*invenio-checker*)
  - Advanced workflow to preserve cataloguing work in case of external updates (*dictdiffer, holdingpen, workflow...*)

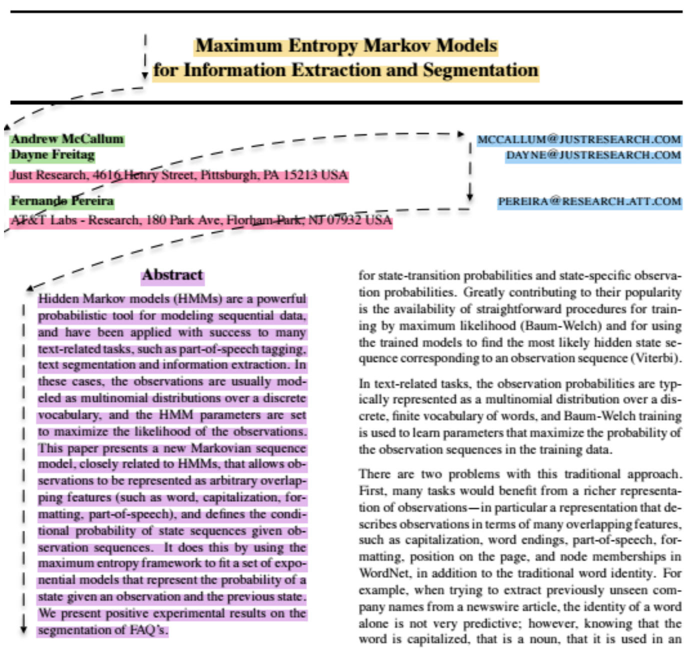
Objectives of our development: Crowdsourcing

- Users to have an active part in the quality of data:
  - suggesting new content (through easy forms)
  - proposing corrections of any record
  - claiming/rejecting proposed papers association to their user profile
  - helping correcting wrong/missing citation or references



Objectives of our development: Machine learning

- Automatic learning from cataloger/user input to:
  - suggest potential user profiles (*beard, beard-server, invenio-beard*)
  - tag records as core/non core records upon ingestion
  - recognize metadata from PDFs (e.g. to guess references/affiliations) (*invenio-grobid*)



Maximum Entropy Markov Models
for Information Extraction and Segmentation

Andrew McCallum MCCALLUM@JUSTRESEARCH.COM
Dayne Freitag DAYNE@JUSTRESEARCH.COM
Just Research, 4616 Henry Street, Pittsburgh, PA 15213 USA

Fernando Pereira PEREIRA@RESEARCH.ATT.COM
AT&T Labs - Research, 180 Park Ave, Florham Park, NJ 07932 USA

Abstract

Hidden Markov models (HMMs) are a powerful probabilistic tool for modeling sequential data, and have been applied with success to many text-related tasks, such as part-of-speech tagging, text segmentation and information extraction. In these cases, the observations are usually modeled as multinomial distributions over a discrete vocabulary, and the HMM parameters are set to maximize the likelihood of the observations. This paper presents a new Markovian sequence model, closely related to HMMs, that allows observations to be represented as arbitrary overlapping features (such as word, capitalization, formatting, part-of-speech), and defines the conditional probability of state sequences given observation sequences. It does this by using the maximum entropy framework to fit a set of exponential models that represent the probability of a state given an observation and the previous state. We present positive experimental results on the segmentation of FAQ's.

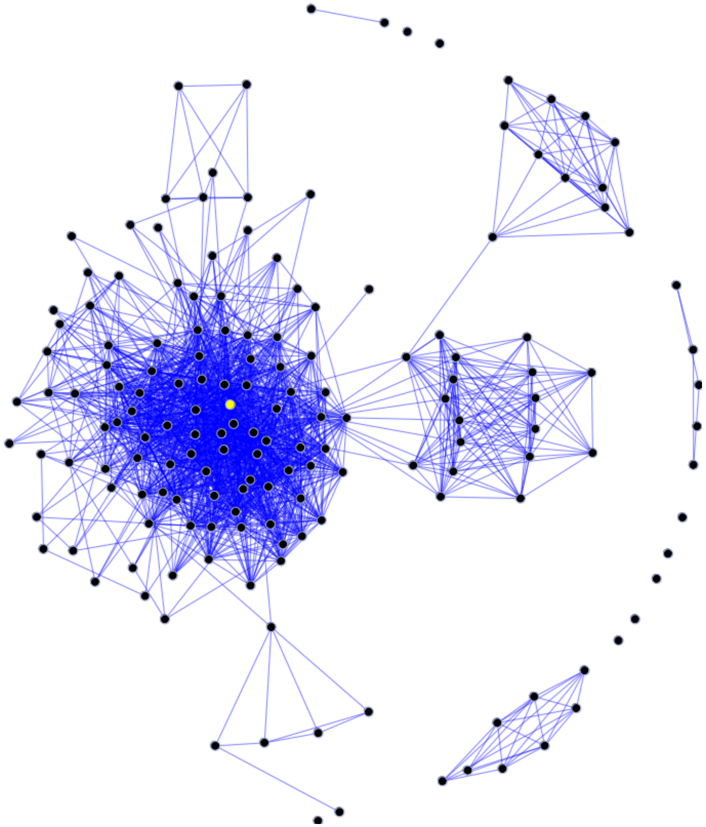
for state-transition probabilities and state-specific observation probabilities. Greatly contributing to their popularity is the availability of straightforward procedures for training by maximum likelihood (Baum-Welch) and for using the trained models to find the most likely hidden state sequence corresponding to an observation sequence (Viterbi).

In text-related tasks, the observation probabilities are typically represented as a multinomial distribution over a discrete, finite vocabulary of words, and Baum-Welch training is used to learn parameters that maximize the probability of the observation sequences in the training data.

There are two problems with this traditional approach. First, many tasks would benefit from a richer representation of observations—in particular a representation that describes observations in terms of many overlapping features, such as capitalization, word endings, part-of-speech, formatting, position on the page, and node memberships in WordNet, in addition to the traditional word identity. For example, when trying to extract previously unseen company names from a newswire article, the identity of a word alone is not very predictive; however, knowing that the word is capitalized, that is a noun, that it is used in an

Objectives of our development: Enriching metadata

- Capturing and exposing the **citation graph**
- Reliably connecting paper signatures to corresponding **author profiles**
- Reliably connecting paper signatures to corresponding **institutions**



Conclusion

- Serving the users is the first priority
- Covering the whole HEP subject
- High quality metadata
  - dedicated curation
  - crowdsourcing
  - machine learning
- Rewriting everything on top of the new Invenio