CERN SIS / INSPIRE and HEPData

Alexander Kohls / Stella Christodoulaki

HEPData Advisory Board, Durham, 28 January 2020

CERN
Scientific
Information
Service

Agenda

1. Overview CERN Scientific Information Service
2. Current and future SIS Projects
3. HEPData – CERN/INSPIRE history
4. INSPIRE Overview
5. The new INSPIRE
6. Possible HEPData – INSPIRE collaboration

CERN's Scientific Information Service (est. 1954)

Alexander Kohls	• Open Access Policy
Jens Vigen	• CERN Publishing

Archive

Anita Hollier

- CERN Archive
- Pauli Archive
- Records Mgmt.

Inspire

Stella Christodoulaki

- INSPIRE Collaboration
- HEPdata
- SCOAP^3 Repository

Library

Tullio Basaglia

- Electronic Resources Mgmt.
- On-site Library
- CERN Bookshop
- Document mgmt.

Open Science

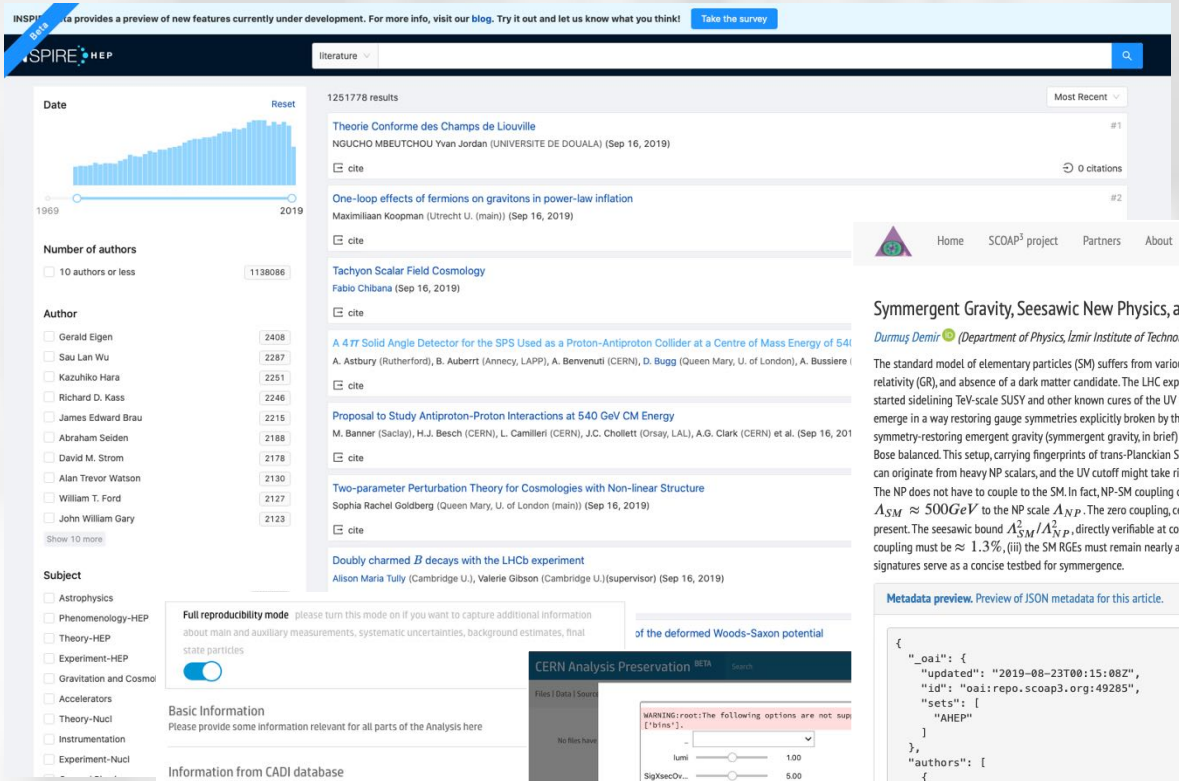
Kamran Naim

- CERN Analysis Preservation
- Persistent Identifiers
- SCOAP^3 Collaboration

Our Mission

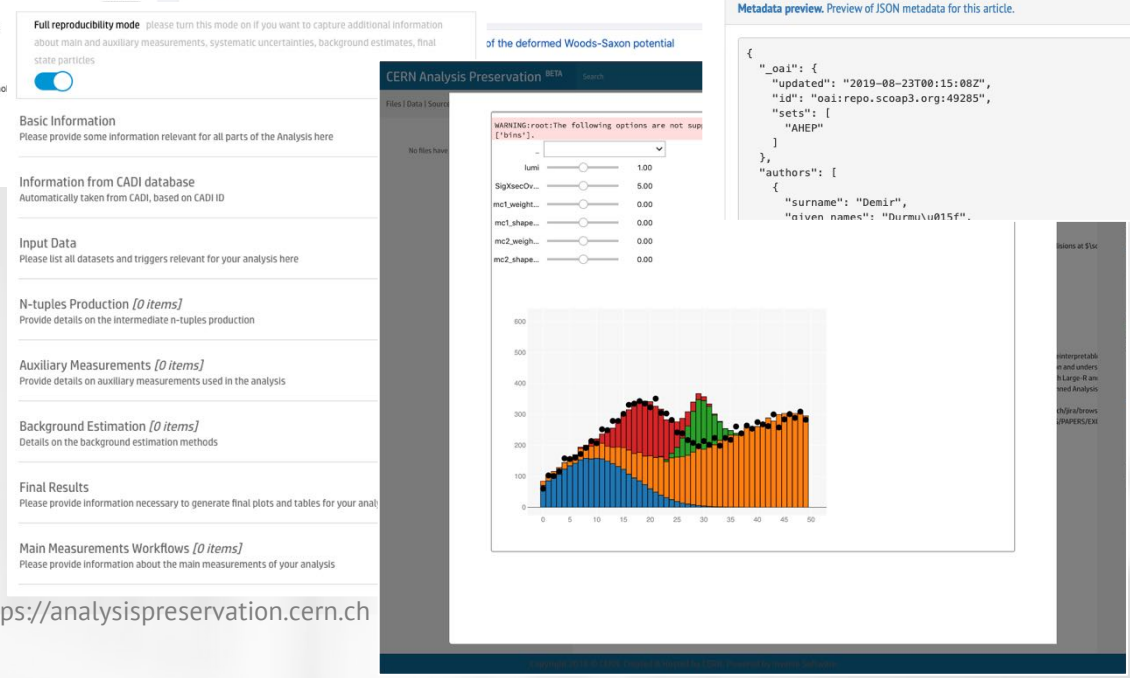
The CERN Scientific Information Service aims at efficiently managing, preserving and disseminating scientific information to make it openly accessible and reusable to CERN and the worldwide High-Energy Physics community.

The role of CERN's Scientific Information Service changed over time…



https://inspirehep.net

https://repo.scoap3.org

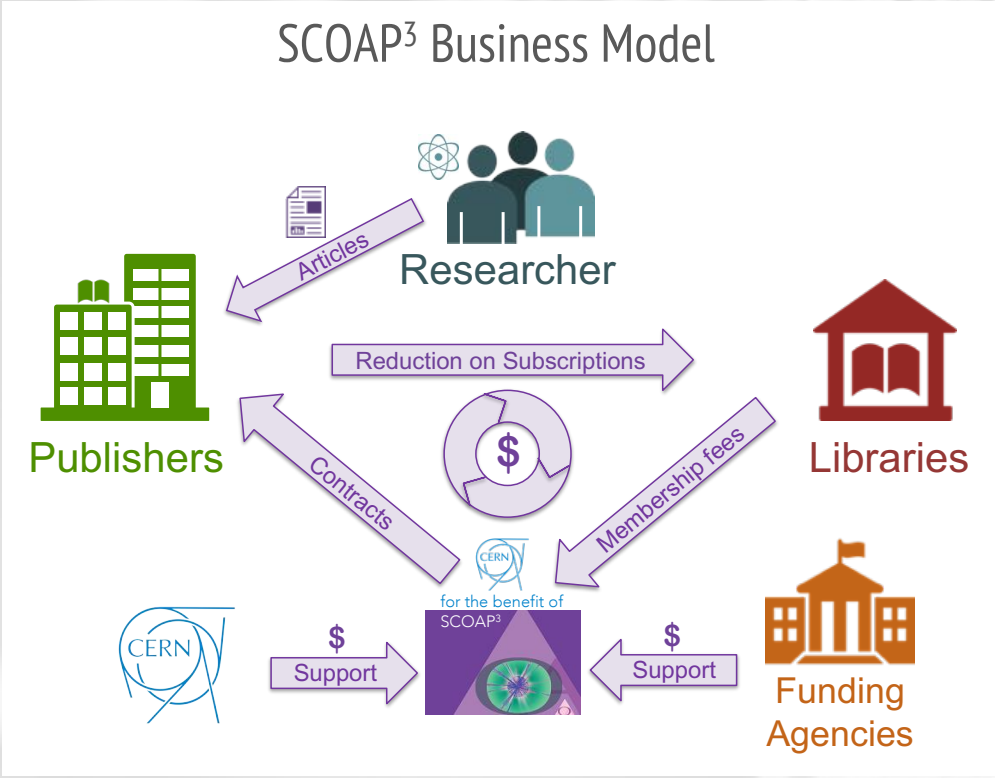


https://analysispreservation.cern.ch

Scientific Information Service

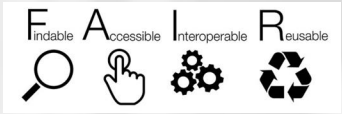
SCOAP³ – the largest OA initiative in the world

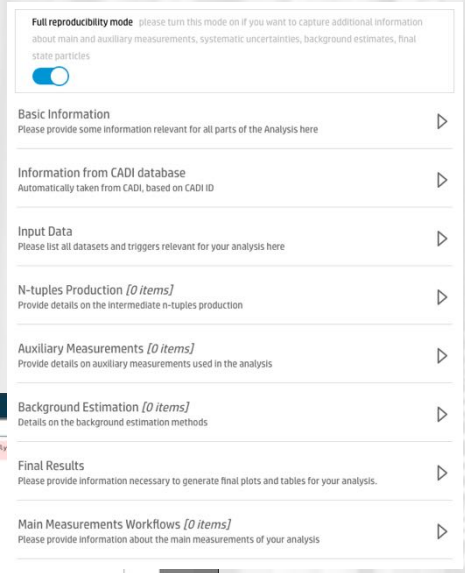
- SCOAP³ (Sponsoring Consortium for Open Access Publishing in Particle Physics) makes ~90% of HEP journal articles OA
- Supports now 7,000 articles/year in 11 journals
- > 32,000 OA articles since its start in 2014
- 3,000 partner libraries from 43 countries and 3 IGO's
- Partner libraries redirect funds previously used to pay subscriptions
- CERN is host organization and contractual counterpart of all stakeholders



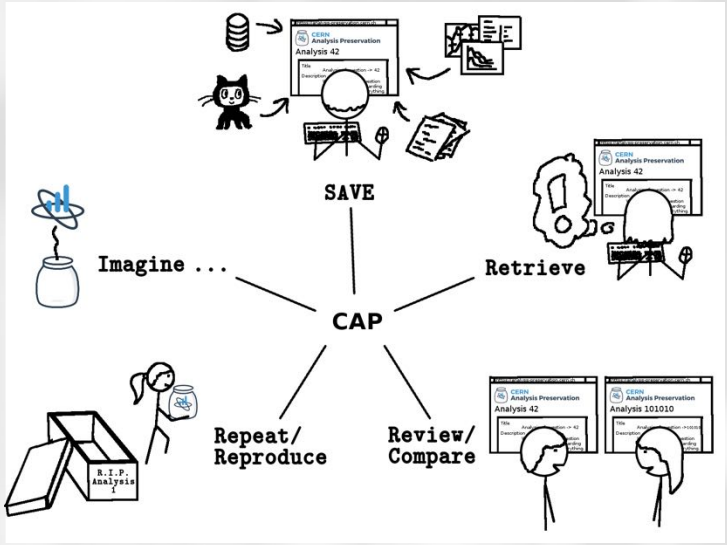
CERN Analysis Preservation – help researchers to preserve their work

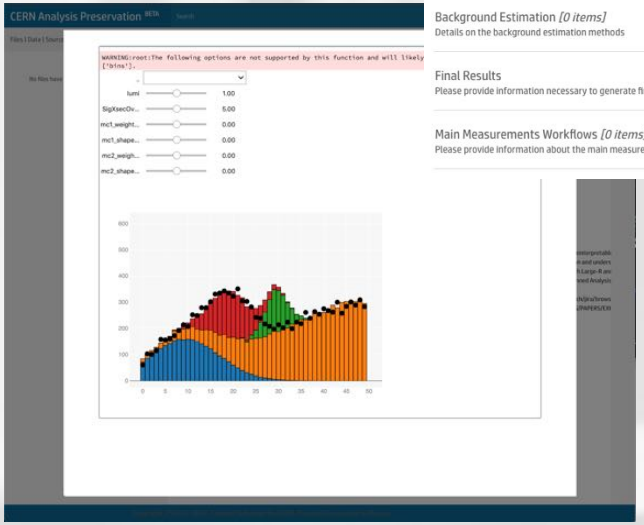
- Goal: help researchers to preserve and share their research objects and related metadata (i.e. scripts, workflows, notes, tables, plots, wikis, etc...)
- Capture directly all elements needed to understand and rerun an analysis





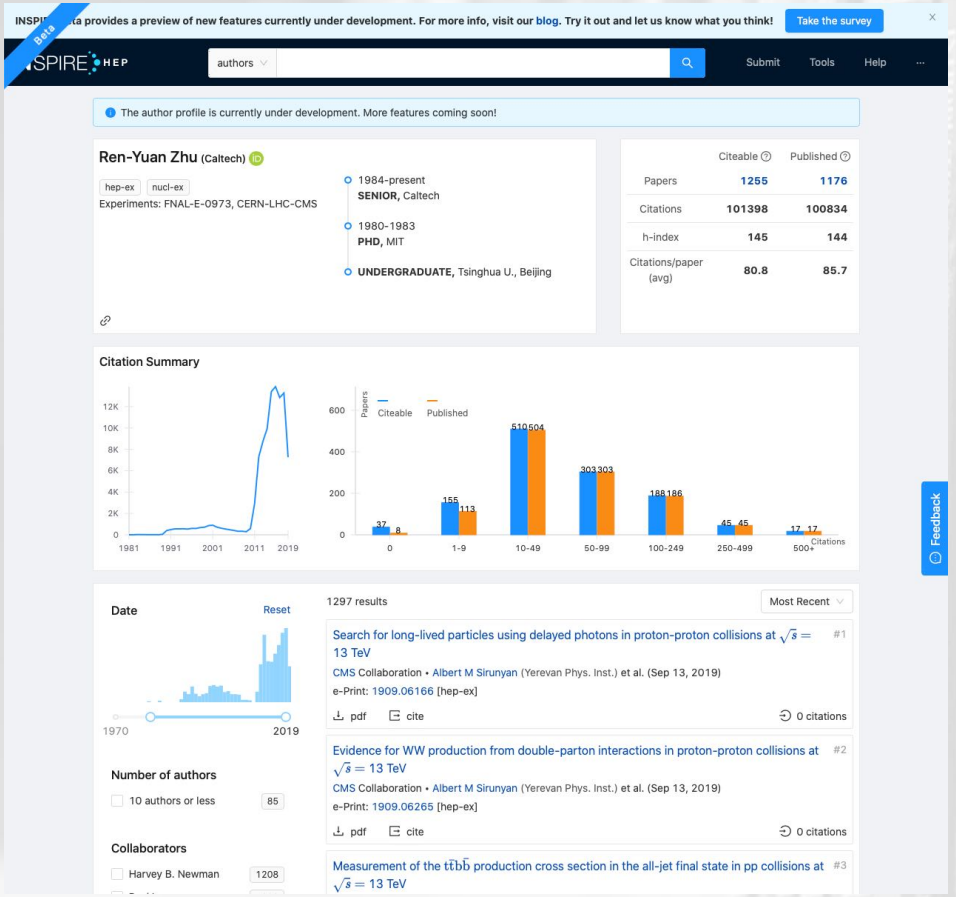
- Integration with related services (i.e. Github, Gitlab, Zenodo, etc.); ongoing integration with CERN IT services for remote execution and reuse (REANA)
- Piloted with 4 major LHC experiments





INSPIRE – the portal for particle physics

- Literature search (published & unpublished), author profiles, experiments and conference data base & HEP related jobs
- 50k active users (researchers)
- 1.4 M bibliographic records
- 23M citations
- 200k searches/day
- 6 collaborating institutions



Fermilab

CERN

SLAC

cnrs IN2P3 Les deux infinis

DESY.

Institute of High Energy Physics
Chinese Academy of Sciences

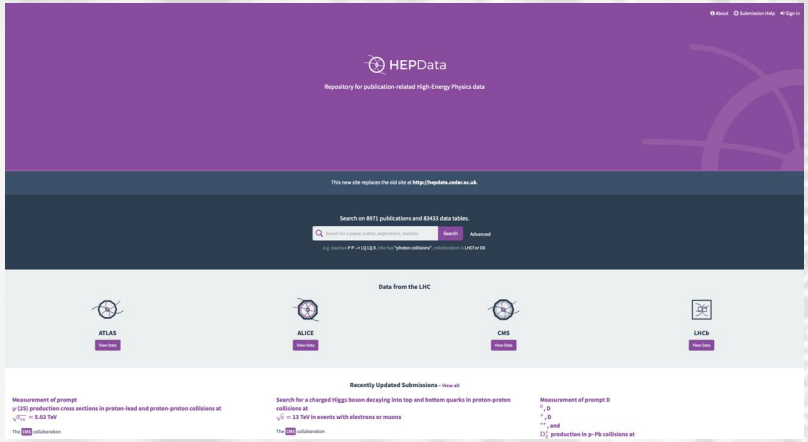
CERN Scientific Information Service

What's next?

- Expand SCOAP^3 to include books & monographs and more journal publications
- Create library innovation network to develop (cross-disciplinary) technology and education solutions to promote Open Science
- Launch and promote tools and processes to mitigate diminishing INSPIRE curation resources (machine learning, crowd-sourcing)
- Use INSPIRE to help the worldwide HEP research community to connect, exchange and apply Open Science practices

CERN SIS and HEPData

- Long collaboration with HEPData – from SPIRES to INSPIRE (facilitating discovery through article-data links)
- 2015-2016: CERN dedicates one developer to the creation of the new HEPData (Invenio based)
- Since 2016: CERN runs HEPData web service on its cloud and supports basic operation through INSPIRE ops expert
- 2017-2018: Discussions about a more formal collaboration with INSPIRE



INSPIRE: Who we are

INSPIRE is a collaboration of CERN, DESY, Fermilab, IHEP, IN2P3, and SLAC, and interacts closely with HEP publishers, arXiv.org, NASA-ADS, PDG, HEPData and other information resources.

The INSPIRE team is composed by a diverse team of professionals: physicists and librarians as content managers, software developers, system engineer and product manager.

INSPIRE: What we do

HEP AGGREGATOR

COMMUNITY DRIVEN

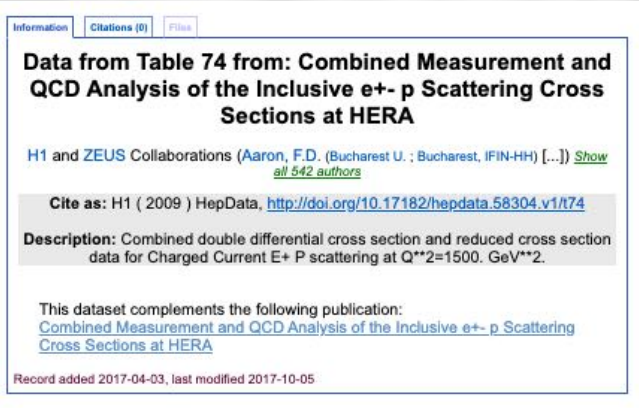
CONTENT QUALITY

DISCOVERABILITY

- Discover HEP research
- Extract accurate citation metrics
- Find impact of one's research
- Find new collaborators
- Hire researchers
- Find HEP jobs and conferences
- ...

INSPIRE and HEPData

- INSPIRE harvests HEPData weekly and connects data to the approprate paper and author


Information | Citations (0) | Files

Data from Table 74 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA

H1 and ZEUS Collaborations (Aaron, F.D. (Bucharest U. ; Bucharest, IFIN-HH) [...]) Show all 542 authors

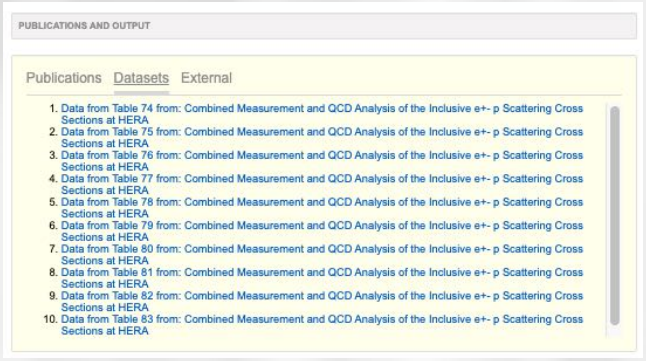
Cite as: H1 (2009) HepData, http://doi.org/10.17182/hepdata.58304.v1/t74

Description: Combined double differential cross section and reduced cross section data for Charged Current E+ P scattering at Q**2=1500. GeV**2.

This dataset complements the following publication:
Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA

Record added 2017-04-03, last modified 2017-10-05

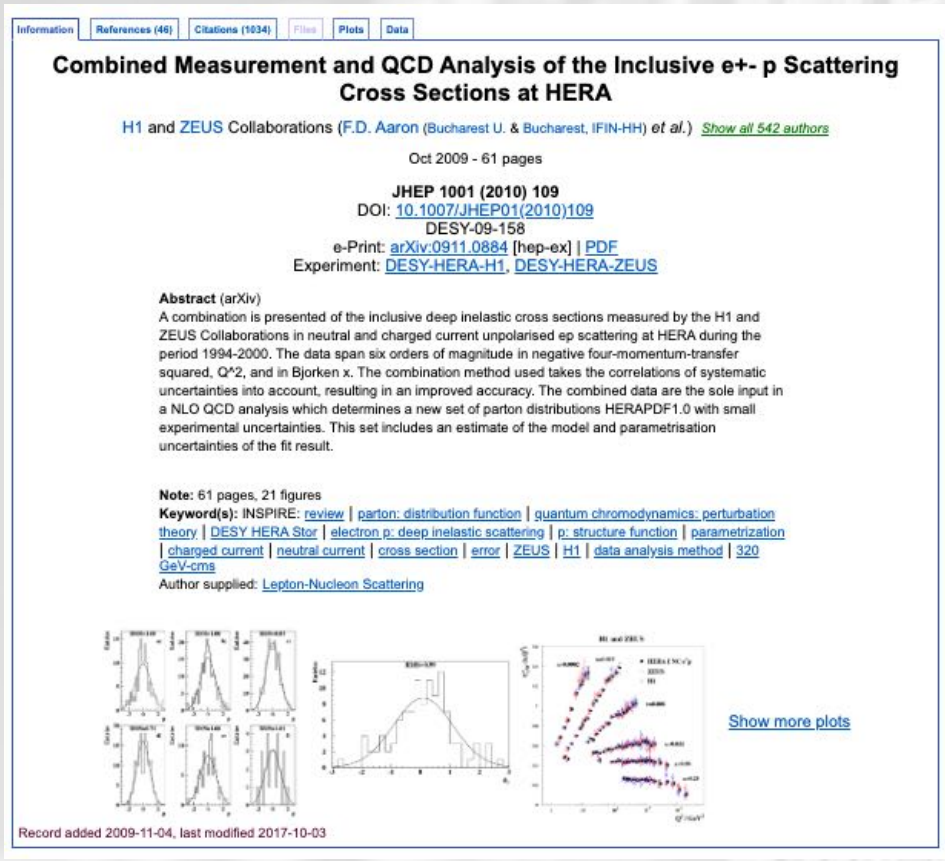
(Data record)


PUBLICATIONS AND OUTPUT

Publications | Datasets | External

1. Data from Table 74 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
2. Data from Table 75 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
3. Data from Table 76 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
4. Data from Table 77 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
5. Data from Table 78 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
6. Data from Table 79 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
7. Data from Table 80 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
8. Data from Table 81 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
9. Data from Table 82 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA
10. Data from Table 83 from: Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA

(Author record)


Information | References (46) | Citations (1034) | Files | Plots | Data

Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA

H1 and ZEUS Collaborations (F.D. Aaron (Bucharest U. & Bucharest, IFIN-HH) *et al.*) Show all 542 authors

Oct 2009 - 61 pages

JHEP 1001 (2010) 109
DOI: 10.1007/JHEP01(2010)109
DESY-09-158
e-Print: arXiv:0911.0884 [hep-ex] | PDF
Experiment: DESY-HERA-H1, DESY-HERA-ZEUS

Abstract (arXiv)
A combination is presented of the inclusive deep inelastic cross sections measured by the H1 and ZEUS Collaborations in neutral and charged current unpolarised ep scattering at HERA during the period 1994-2000. The data span six orders of magnitude in negative four-momentum-transfer squared, Q^2, and in Bjorken x. The combination method used takes the correlations of systematic uncertainties into account, resulting in an improved accuracy. The combined data are the sole input in a NLO QCD analysis which determines a new set of parton distributions HERAPDF1.0 with small experimental uncertainties. This set includes an estimate of the model and parametrisation uncertainties of the fit result.

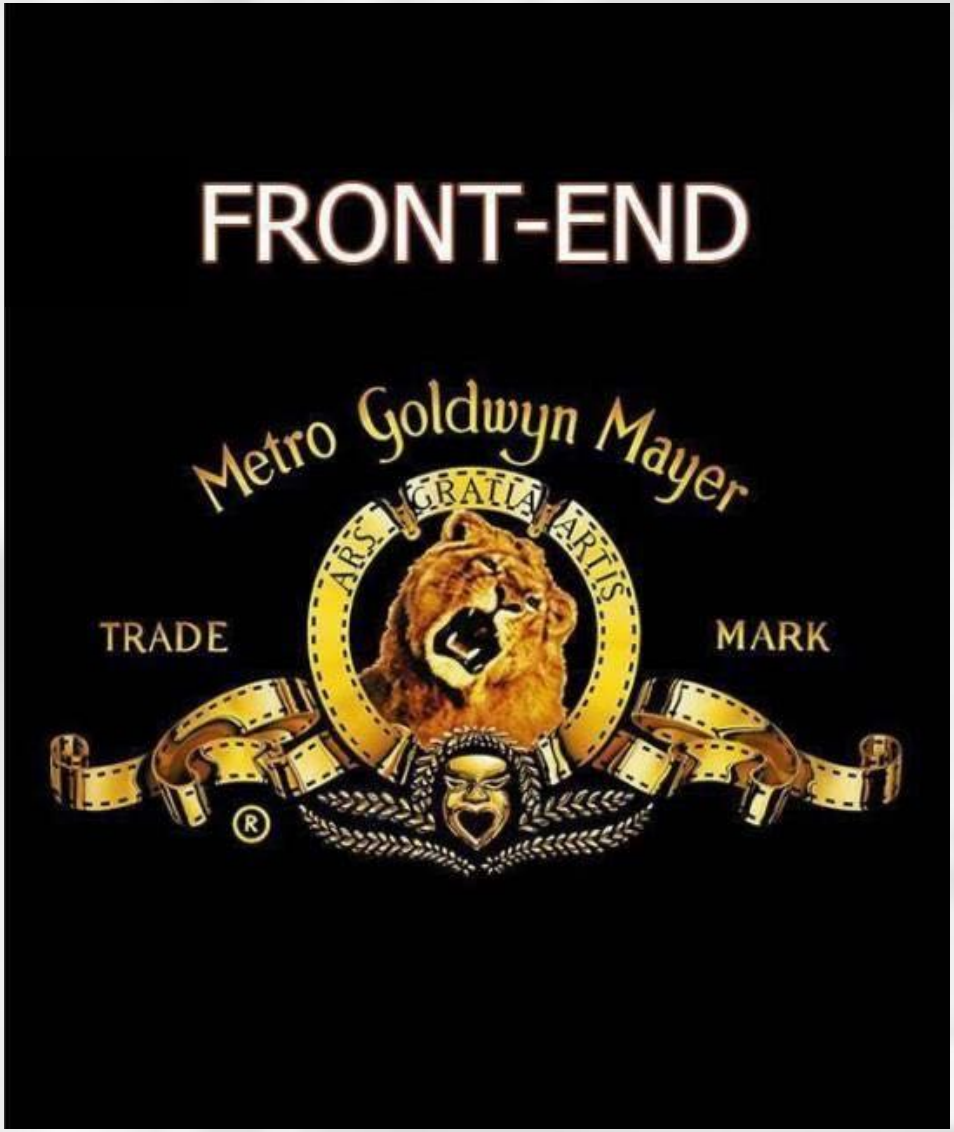
Note: 61 pages, 21 figures
Keyword(s): INSPIRE: review | parton: distribution function | quantum chromodynamics: perturbation theory | DESY HERA Stor | electron p: deep inelastic scattering | p: structure function | parametrization | charged current | neutral current | cross section | error | ZEUS | H1 | data analysis method | 320 GeV-cms
Author supplied: Lepton-Nucleon Scattering

Show more plots

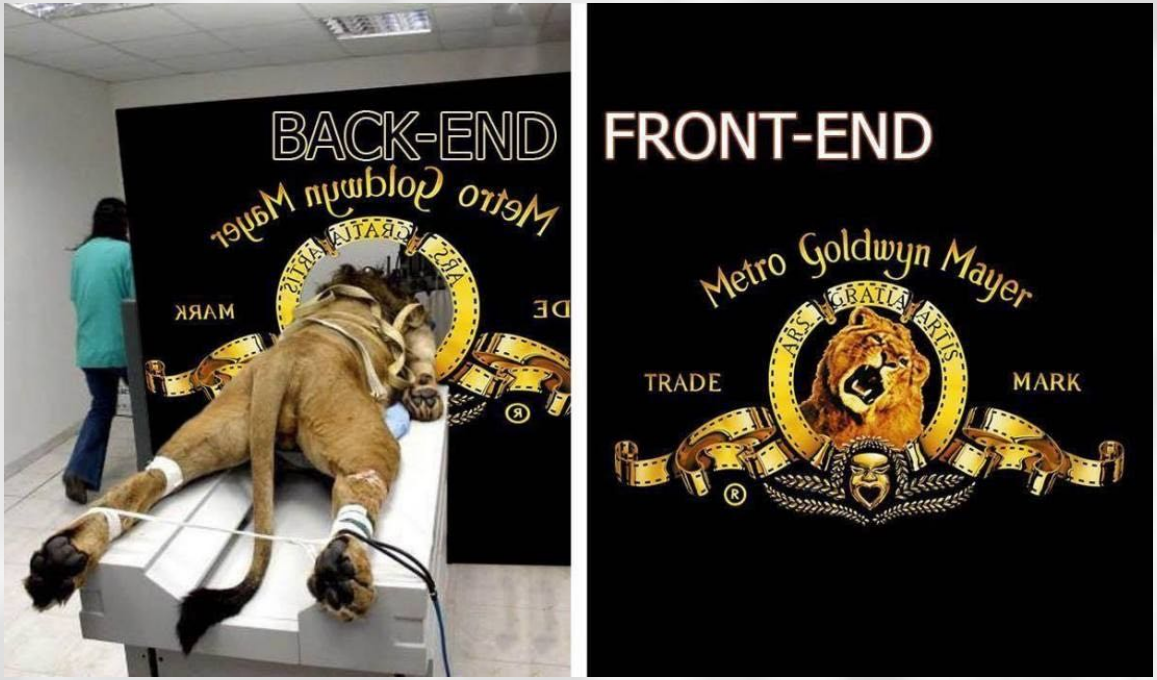
Record added 2009-11-04, last modified 2017-10-03

(Paper record)

Towards INSPIRE's next generation



Towards INSPIRE's next generation



HARD TO KEEP INSPIRE STABLE - Very old technology, hard to maintain and scale

DECREASING RESOURCES - Funding for (curation) resources is decreasing

REQUESTS FOR MORE FEATURES - Opportunity to benefit from Machine Learning and crowdsourcing tools

INSPIRE beta

MINIMUM VIABLE PRODUCT

Focus on the basic and most popular functionality at this initial stage: search papers, authors, jobs, conferences

DATA

Better integration with data is our next step

USER-CENTERED DEVELOPMENT

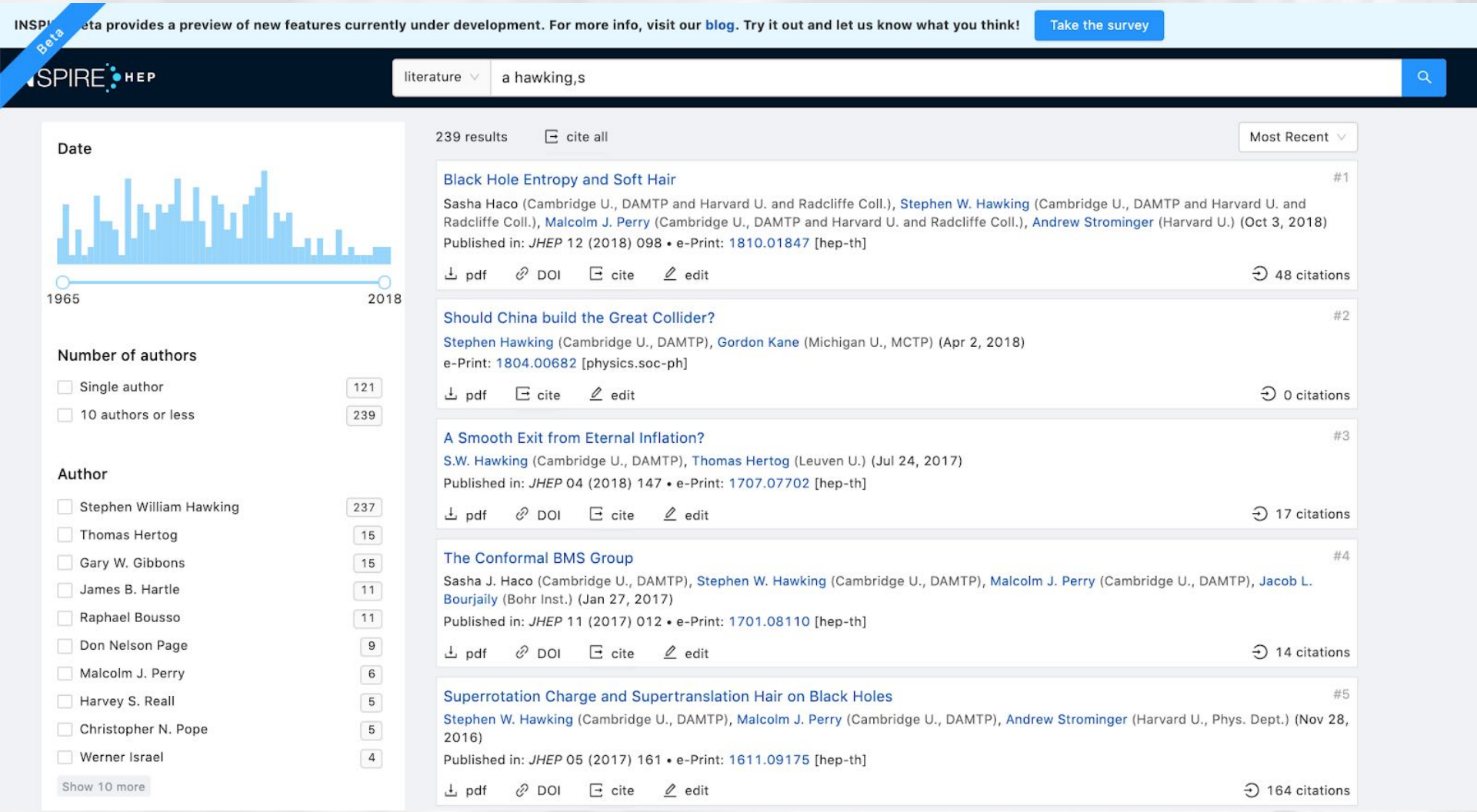
Thorough user testing (user interviews, focus groups, surveys, stats) to understand community needs

CURRENTLY UNDER TESTING

INSPIRE beta runs in parallel with the current platform

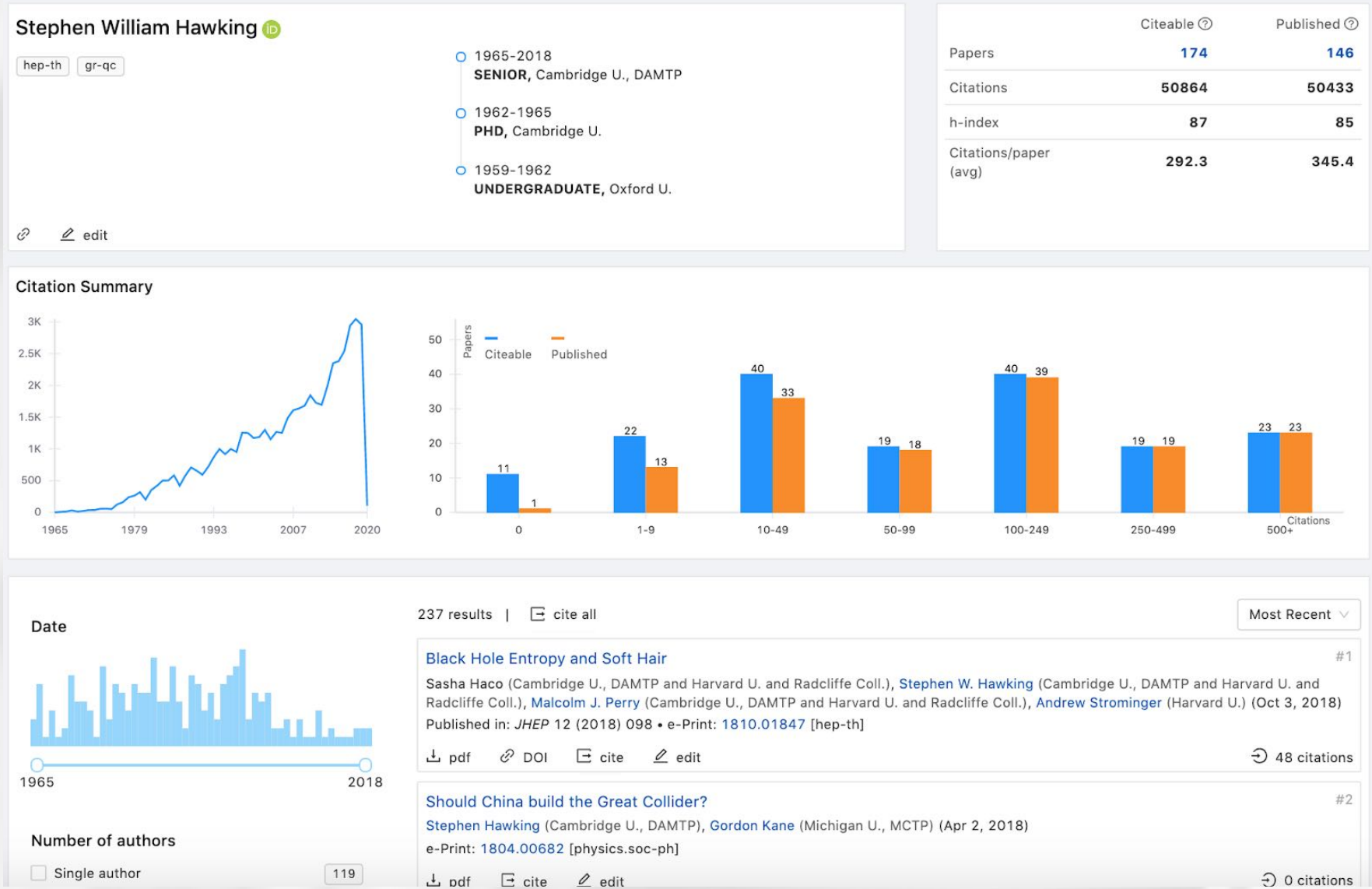
INSPIRE beta search

https://beta.inspirehep.net



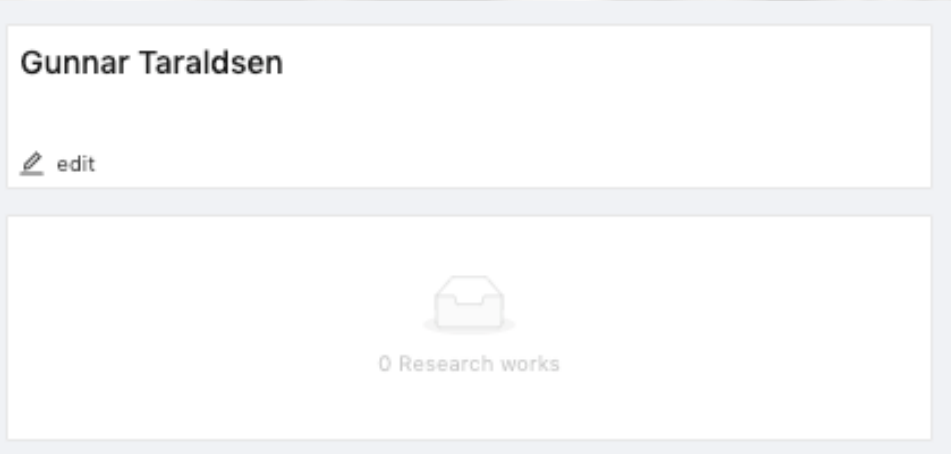
INSPIRE beta author profile

https://beta.inspirehep.net



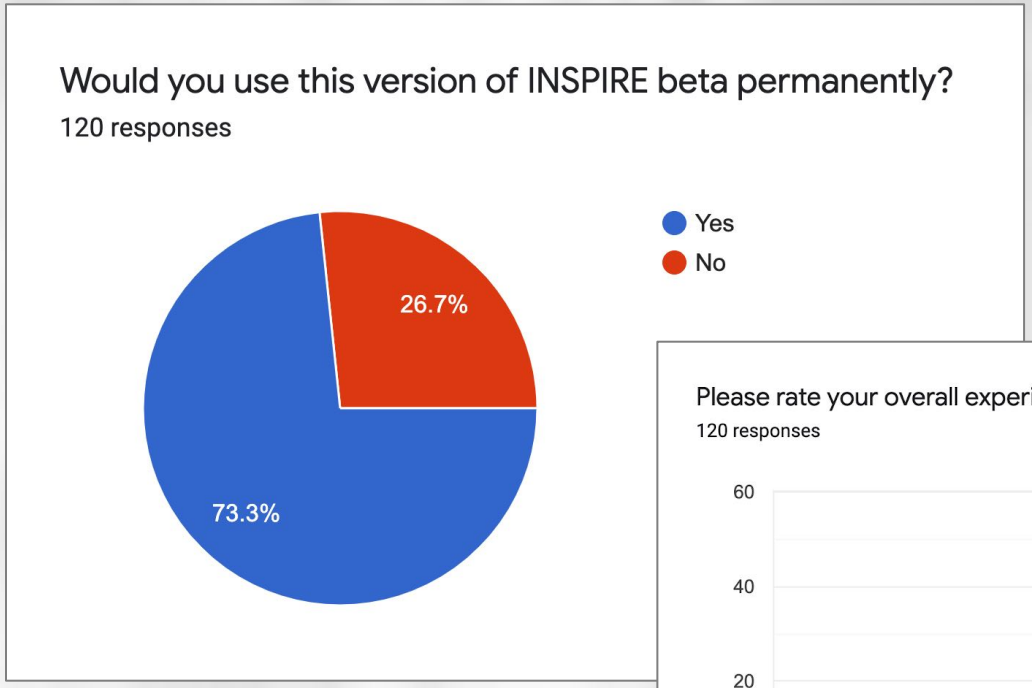
Focus on Open Science

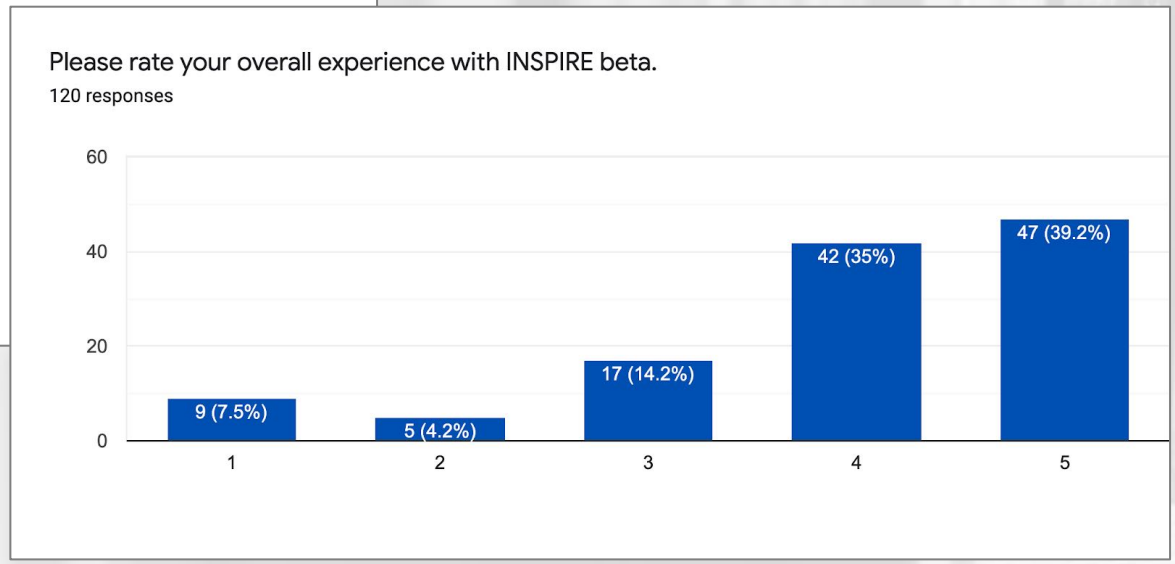
In the new INSPIRE, we use the term “Research works” to highlight that a researcher’s contribution is not just papers, but also data, software, analysis, etc.



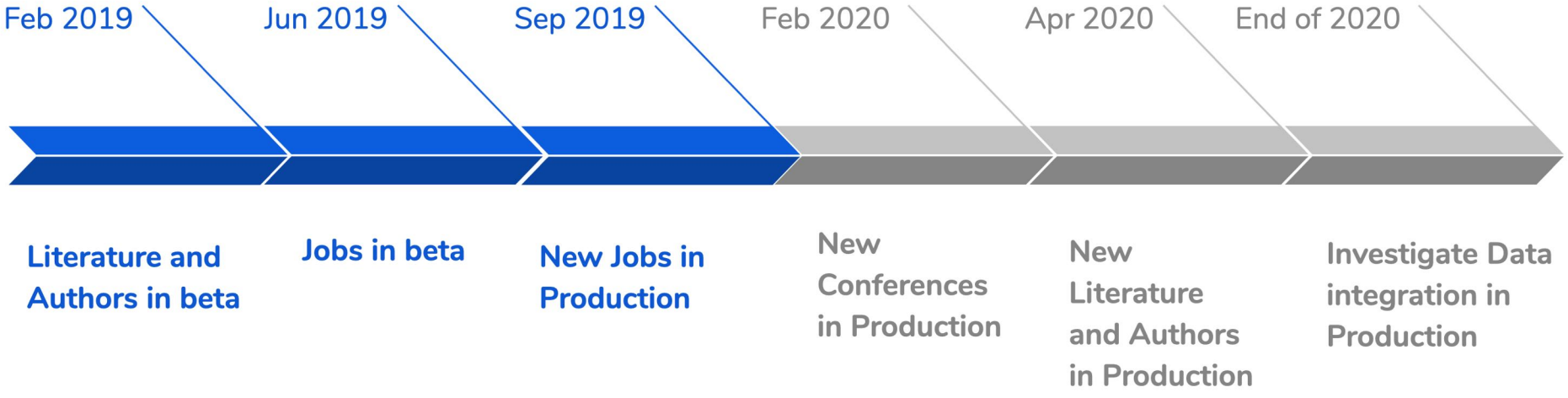
Very positive user feedback from beta tests

73.3% would use the new INSPIRE version permanently and **74.2%** rate their experience positive or excellent





New INSPIRE platform timeline



INSPIRE and HEPData

INSPIRE aims to be a community driver in Open Science and that goes hand in hand with HEPData

USER RESEARCH

- How to represent Data in INSPIRE so that it benefits more the HEP community?
- How to better expose Data in INSPIRE to create incentives to researchers to preserve and share their Data?
- How to leverage the collaboration HEPData – INSPIRE?

COLLABORATION

We envision a closer collaboration following community needs

INSPIRE governance model

- INSPIRE is a collaboration of 6 laboratories:

- Roles, contributions and governance defined in the INSPIRE Collaboration Agreement, signed by all partners
- All labs contribute to the ingestion workflows and curation activities (usually ~2 FTE)
- CERN hosts data bases and services and provides INSPIRE system engineer, development team and product manager
- New interested INSPIRE partners can start participation with individual bilateral collaboration agreement with CERN

What can we do together?

Idea

CERN and the INSPIRE collaboration invite IPPP Durham to enter into bilateral collaboration agreement with CERN to formalise the existing cooperation and to expand it further

- CERN to host database and provide operational support
- HEPData to support ingestion of records into INSPIRE and to curate HEPData records as needed
- HEPData and INSPIRE teams to develop ideas for closer integration of HEPData records