INSPIRE - The Next-Generation HEP Information System

Annette Holtkamp[1,2]

[1]CERN, 1211 Geneve 23, Switzerland
[2]on leave of absence from DESY, Notkestraße 85, 22607 Hamburg, Germany

CERN, DESY, Fermilab and SLAC have joined forces to build INSPIRE, the next-generation HEP information platform offering innovative tools for information discovery and communication. Representing a natural community-based evolution of SPIRES, INSPIRE provides fast access to the entire body of HEP literature. As a subject repository it will host fulltexts of preprints, Open Access journal articles and supplementary material like conference slides and multimedia, enabling novel text- and data mining applications. In the spirit of Web2.0 INSPIRE will also supply tools for collaboration and user-enriched content.

1 Introduction

Since decades, the SPIRES [1] database of High-Energy Physics (HEP) literature and related information, maintained by DESY, Fermilab and SLAC, is serving as the dominant information resource for particle physicists [2]. To adapt current technological advances to the workflow of HEP scientists SPIRES is moving to a new platform, Invenio [3], a modern open-source multimedia digital library platform developed at CERN. The three SPIRES partners DESY, Fermilab and SLAC have joined forces with CERN to build INSPIRE, the next-generation gateway to all HEP relevant information. Based on hosting the entire body of HEP literature metadata and the fulltext of all open access publications, it will empower scientists with state-of-the-art discovery tools. This paper highlights some of the innovations INSPIRE will offer in the near future.

2 Extended content

A survey launched in 2007 [2] revealed the need of the HEP community for access to material beyond the current scope of SPIRES. Following these wishes INSPIRE will over time significantly broaden the spectrum of covered material.

Since highest priority was given in the survey to the access to fulltext, INSPIRE will host the fulltext of all freely accessible preprints, journal articles, conference contributions, theses as well as research notes of experimental collaborations. The indexing of conference talks will be accompanied by long-term archiving of the corresponding slides. In addition, INSPIRE will act as repository for all kinds of supplementary material, like data, software or multimedia, which authors may want to supply to increase the usability of their scientific results. Thus it

will serve as a long-term archive for material that is in danger of getting irrevocably lost over time.

Another explicit wish of the HEP community is seamless access to older material. Complementing the role of arXiv [4] as the place where particle physicists deposit their new material, INSPIRE will ingest pre-arXiv articles wherever possible, aiming to cover the complete historical corpus of HEP literature.

3 Extended functionality

In addition to reproducing SPIRES' successful functionality on a modern platform, INSPIRE profits from Invenio's advanced technology to offer new features.

Searches are not only much faster and more efficient, but also, besides supporting the SPIRES specific syntax, INSPIRE enables Google-like free-text searches across metadata and fulltext.

For each article, a detailed page shows abstract, keywords and links to different fulltext versions, supplementary material and a growing wealth of additional information. A new feature to be expected on the detailed article page in the near future is the display of figures extracted from the fulltext. In parallel, figure captions will become independently searchable.

Citation analysis is supplemented by a citation history, visualizing the citation counts of an article over time. The "co-cited with" feature opens up new paths to find related articles.

Author pages provide a citation summary and comprehensive information on affiliation history, research subjects, frequent coauthors, research output. Work is under way to disambiguate authors and correctly attribute articles and in the future other scientific output to their creators.

Personal accounts will enable options like private bookshelves, alerts or RSS feeds, personalized display formats or tools for sharing information within a collaboration. Web2.0 functionality will allow scientists in the future to submit material, to attach comments, to tag content or to aggregate related objects.

Semantic techniques for information retrieval are under development that, based on a taxonomy of HEP concepts, will allow e.g. to translate a search for "SQCD" into a query for "supersymmetry" and "quantum chromodynamics", to search for material of similar content or to facilitate broadening and narrowing searches. Another application currently being refined is the automatic categorization of material on ingestion, so that a paper is automatically recognized e.g. as a conference talk on Higgs search or as an internal note on detector calibration. Automatic assignment of keywords from the HEP taxonomy will allow unified searches across different materials.

4 Outlook

At the forefront of information technology, INSPIRE will continuously develop innovative tools to put at the disposal of particle physicists.

Moving beyond the paper-centric paradigm, INSPIRE provides a platform not only to store supplementary material as attachments to an article but to treat all kinds of scientific output as independent objects, be it a dataset, a video or a software program. A user friendly search interface and a standardized set of metadata will allow comprehensive searches over all database content. Through the assignment of unique identifiers (similar to DOI's [5]) to non-article objects they will become independently citable. Measuring their usage will contribute to the

development of new metrics creating more comprehensive profiles for researchers and allowing to track the impact of ideas and retrieve the most relevant material.

INSPIRE's capacity to handle non-article objects blends in perfectly with initiatives towards data preservation whose importance is increasingly recognized by the HEP community [6, 7]. Thanks to its role as central HEP information system, INSPIRE is ideally placed to become an essential partner in digital preservation e.g. by hosting high-level data files and making them discoverable and citable, by ingesting and preserving documentation in the form of internal notes or wikis under the protection of a flexible access control system, and by participating in the development of standards.

To exploit synergies with other information providers and neighboring fields, collaboration will be intensified among others with the Astrophysics Data System (ADS)[8] and the Particle Data Group (PDG) [9] with the aim of providing seamless access to related information resources. Thus, finding an astroparticle physics paper on INSPIRE will smoothly lead to the underlying data sets indexed at ADS. An experimental or phenomenological paper on leptonic Z decay will be linked to corresponding information on the PDGlive website [10], and searches using PDG codes are planned to be enabled in INSPIRE. Projects undertaken in cooperation with publishers aim e.g. to improve ingestion of metadata or to enable fulltext searches of articles predating arXiv.

INSPIRE will thus continue and develop further the role that SPIRES has played as "the" reference HEP scientific information platform, providing inspiration for information management in other fields of science.

INSPIRE will go into production in Spring 2010. A public beta version is already available for testing.

References

[1] http://spires.slac.stanford.edu

[2] A. Gentil-Beccot *et al.*, J. Am. Soc. Inf. Sci. Technol. **60** (2009) 150, arXiv:0804.2701

[3] http://cdsware.cern.ch/invenio/index.html

[4] http://arXiv.org

[5] http://www.doi.org

[6] A. Holzner *et al.*, arXiv:0906.0485

[7] R. Mount *et al.*, arXiv:0912.0255;
https://www.dphep.org

[8] http://adswww.harvard.edu

[9] http://pdg.lbl.gov

[10] http://pdglive.lbl.gov