

INSPIRE: A new information system for High-Energy Physics. Lessons learnt.

e-Science brings opportunities and challenges for the world of scholarly communication: it amplifies the needs of scientists for fast, effective, unrestricted communication of ideas and scientific results, through Open Access; it enables automation of librarianship intelligence, providing new services to the scientific community for the discovery of information; it calls on libraries and information professionals to fill new roles, as evolving actors in the scholarly communication chain.

The field of High-Energy Physics (HEP) has pioneered infrastructures for scholarly communication, with half a century of tradition in Open Access and pre-print dissemination and two decades of experiences in repositories. Scholarly communication in HEP is now moving fast in the e-Science era.

This contribution will analyze the status of scholarly communication in the field and the potential offered by the inception of INSPIRE, the next-generation repository for the field.

To provide context, it should be noted that contemporary scholarly discourse follows many alternative routes in addition to the three-century old tradition of publication in peer-reviewed journals. HEP has explored alternative communication strategies for decades, initially via the mass mailing of paper copies of preliminary manuscripts, then via the inception of the first online repositories and digital libraries. HEP is therefore uniquely placed to answer recurrent questions raised by the current trends in scholarly communication: is there an advantage for scientists to make their work available through repositories, often in preliminary form? Is there an advantage to publishing in Open Access journals? Do scientists still read journals or do they use digital repositories?

The analysis of citation data demonstrates that free and immediate online dissemination of preprints creates an immense citation advantage in HEP, whereas publication in Open Access journals presents no discernible advantage. In addition, the analysis of clickstreams in the leading digital library of the field shows that HEP scientists seldom read journals, preferring preprints instead. Quantitative evidence in support of these theses will be provided.

This unique situation is the background for another study aimed to understanding the evolving needs of the user community in light of a transforming information environment. A survey of about 10% of practitioners in the field reveals usage trends and information needs. Community-based services, such as the pioneering arXiv and SPIRES systems, largely answer the need of these scientists, with a limited but increasing fraction of younger users relying on Google. Commercial services offered by publishers or database vendors are essentially unused in the field. The survey offers an insight into the most important features that users require to optimize their research workflow. These results are discussed in detail and inform the future evolution of information management in HEP in the INSPIRE platform.

The results of these studies have a wider interest, as HEP researchers are traditionally 'early adopters' of innovation in scholarly communication. Other disciplines and service providers can be inspired by these findings.

This contribution will then focus on a new, groundbreaking, initiative in the field: the INSPIRE project.

The INSPIRE project, jointly executed by the four leading laboratories in the field, in the US and Europe, has a unique role in the daily workflow of HEP scientists. It is the successor of the SPIRES platform, and in synergy with arXiv.org and leading publishers in the field is building a next-generation platform for scientific information in HEP. Based on the Invenio Open Source CERN digital library software, it provides continuous access to almost 1 Million records among which 500000 Open Access full-text documents.

This contribution presents the rationale for the evolution of INSPIRE, the decisions which have guided the implementation of the beta version, the strategy to build new repository services.

The present services offered by INSPIRE, and its adoption by the users community are described: citation analysis, author disambiguation, harvesting of Open Access content, strategic partnerships with other information providers and leading publishers in the field.

INSPIRE is now designing and implementing new services, which will be discussed in the final part of this contribution: user-pulled and not technology pushed, to match the expectations of its community. These will be shortly described: tagging, automatic key-wording, crowdsourcing of curation, automatic author disambiguation, widening of the scope of the collections, semantic analysis of the content, provision of innovative services on images, partnership with other initiatives such as OpenAIRE and D4ScienceII.

Finally, the contribution will list the strategic, design and operational lessons learnt by this amazing adventure in provision of scientific information to a dynamic, lively and engaged user community.