

How do High-Energy Physics scholars discover scientific information?

Anne Gentil-Beccot

CERN, European Organization for Nuclear Research, Switzerland

Abstract

Grey literature has always been the main conduit of scholarly communication for High-Energy Physics (HEP) researchers. An efficient way of searching and accessing this information is a central part of their research workflow. In 2007, a survey was conducted to understand which information resources HEP scholars use to find the information they need. The results of this survey are presented. Over 2000 answers, representing about one-tenth of the active HEP community, were collected and show that community-driven resources largely dominate the landscape, with commercial services serving only a small proportion of the users. In addition, HEP scholars appear to use different tools for different information needs, which are clearly prioritized. Finally, the results of the survey shed light on the future information needs of HEP scientists over the next five years.

Introduction

The High-Energy Physics community is a small and cohesive community, counting about 20,000 scientists. Grey literature has always been the main vehicle of scholarly communication for this community, which has built specific information systems to access and use these resources. As well as information services specifically developed by the community, some other information services exist: commercial online databases, providing metadata about all scientific literature; journal platforms, serving metadata and full-text of articles in the field; the pervasive Google and its scholarly instance, Google scholar...

These three categories of information systems constitute the current landscape of information provision and search platforms used by the HEP scientists. In addition, online services are changing more and more, new tools are developed and new ways of interacting with users evolve. In this environment it is important to understand what is the current usage of information servicesⁱ in HEP. Do scientists prefer publishers' own portals, or Google, or the field-specific access provided by HEP information resources? Do they expect to see some changes? Which of these platforms is better placed to answer any need for change? What kind of evolution should be envisaged for an information system to meet the needs of HEP scholars?

After describing the specific relation there is between HEP community and grey literature, this article describes the survey jointly conducted in 2007 by several HEP research institutes about the usage of HEP information systems. An overview of the methods used in the survey is first presented, followed by the main results: the large domination of community-based services and an insight into current and future information needs of HEP scholars.

Grey literature in High-Energy Physics

The High-Energy Physics (HEP) community has always used grey literature as its main vehicle of communicationⁱⁱ. Half a century ago, before the time of electronic journals, the delay between the submission of a scientific paper and the time when it reached the reader appeared unacceptable to HEP scientists, who embraced preprints as their main means of communication. One of the origins of this phenomenon is the peculiarity of a two-soul community: experimental physicists working at accelerators of subatomic particles of ever-increasing energy, witnessing a discovery a week during the early stages of the discipline; and theoretical physicists interpreting these results, refining their theories and suggesting new studies, as well as providing insight into the inner working of our universe. Months of delay in communication were out of the question: enter grey literature.

With the onset of use of the internet, the process continued electronically. And the community launched its own tools to manage this literature: in 1991, even before the web was invented, Paul Ginsparg, at the Los Alamos laboratories, launched arXiv, the first physics preprint repositoryⁱⁱⁱ. SPIRES^{iv}, the SLAC e-catalogue of preprints, which had existed since 1974, was able to benefit from the invention of the web, and become the first web server in the U.S. - it rapidly became the first HEP information system in the world. In 1993, CDS^v, the CERN Document Server, was launched.

Even if publication in journals is essential for the career evolution of HEP scholars, as for scholars in any other field, this is almost decoupled from the effective information transfer between scientists. Preprints remain today the main communication channel; "only" the discovery and submission processes have changed. Authors still submit a paper to arXiv before submitting to a journal: the date of submission of the paper to arXiv carries more importance to them than the "received" stamp on the subsequent journal article. Even in the era of electronic journals, therefore, grey literature fully retains its importance in the discipline. Instead of being phased out together with the paper era, grey literature is actually growing in importance, reflecting the evolution of modern scholarly communication beyond the printed word. Fifty years ago, scientists communicated by writing articles, of which the preprint versions made up the bulk of grey literature in the field. Nowadays, scientists go to conferences and their slides, generated in digital form, constitute a new form of grey literature that other scientists want to access immediately, and often to quote in their subsequent publications, without having to wait for a conference paper to be written and submitted as a preprint to a repository somewhere.