

50 Years of Experience in Making Grey Literature Available: Matching the Expectations of the Particle Physics Community

Carmen O'Dell, David Dallman, Martin Vesely, and Jens Vigen.
CERN

Abstract:

The CERN Scientific Information Service has been active in the field of digital library research and in providing scientific information services to the high-energy physics community for almost five decades now. Most recently the research focus has been on interoperability issues in document storage and retrieval systems, metadata added-value services, digital library automation and networked information services. The achievements of this research and the implications for treating grey literature are presented, including practical implementation examples.

Introduction

CERN, the European Organization for Nuclear Research, was founded in 1954 and is the world's largest particle physics laboratory. The Member States financing CERN are all European but the laboratory has researchers from all over the world. In fact, there are some 6500 CERN users, representing 500 universities and over 80 nationalities. CERN itself employs around 2500 staff to provide the necessary technical, computing and administrative infrastructure. The CERN Scientific Information Service was created in 1955 and one of its key tasks is to: "acquire and manage information resources in all fields of relevance to the Organization, and make these accessible in the most convenient way to the worldwide particle physics community" [1]. To do this effectively the library has continually adapted and evolved its methods for collecting, treating and disseminating information resources.

1. The Preprint Culture

In all fields of scientific research, there is a need to be the 'first' but research into the tiniest of particles ironically requires some of the biggest machines in the world, with price tags to match. When research is this expensive, there is simply no room to do the same science twice. Over 40 years ago this drove the particle physics community into a culture based on preprints, to accelerate scientific communication beyond that of the lengthy journal publication process. Universities and laboratories produced their own series of preprints describing their research and experimental results, and distributed them free of charge to hundreds of other institutes in the field.

In 1958 the CERN library started issuing a weekly list of the preprints received and by the early 1960s, semi-automatic means were being used to produce it. Eventually each document was catalogued and the bibliographic data typed into a database from which the list could be produced. In 1983 this database was made available to users for searching.

By 1991, around 10,000 paper preprints were being catalogued per year with publication references to the corresponding journal articles also being added to the database by hand. The system was absolutely at the limit of what could be achieved with the two staff members available. In addition, there were problems of space for storing all these paper documents.

Fortuitously, in August 1991, Paul Ginsparg set up the Los Alamos electronic preprint server, now known as arXiv. The server provided an enormous improvement in the speed and ease of preprint communication. By 1992, it was starting to become so popular among particle theorists