Open Access progress at CERN: technical implementation

Alberto Pepe

28 February 2005

Overview

CERN has pursued open dissemination of scientific results since its early beginnings:

The results of CERN experimental and theoretical work shall be published or otherwise made generally available

(CERN CONVENTION, ARTICLE II, 1954)

CERN Document Server (CDS) today offers access to over 750,000 bibliographic records, including 320,000 fulltext documents, of interest to people working in particle physics and related areas. Covers preprints, articles, books, journals, photographs, and much more.

(A presentation on the development of the OA policy at CERN is to follow in Session 3.)

Journals: Preprint to Postprint coverage

Ongoing goal: Improving the coverage of Journal publications

- about 1/3 of articles published by CERN authors are not available as OA fulltexts after publication

- enforcing submission of postprints by authors, when ©opyright allows it (common in physics)

- keeping constant track of OA journals

- collecting contact details of preprint submitters to follow the publication process

This will guarantee full coverage of postprints and up-to-date records

Conferences: Enforcing institutional self-archiving

Ongoing goal: ensuring Open Access to conference material presented by CERN authors

- ensure fast dissemination of conference contributions through OAI

- encourage paper submission within the CERN administrative procedures for travel request

- promote the use of OAI-compliant conference management software: InDiCo

Archives: Accessing CERN Document repository through OAI-PMH

- all CERN-produced documents exposed through the OAI-PMH.

- Metadata Formats `oai_dc, marcxml`

- MARC enriched over DC by
  (i) CERN Library (e.g. conference information)
  (ii) automatic processing (e.g. keywords and citations)

- Need to improve recognition of papers by locally available ranking methods:

  - ranking by number of citations
  - ranking by downloads and by downloaders
  - ... and even ranking by journal impact factors

Archives: Harvesting new metadata through OAI-PMH

- At present, CDS harvests metadata from more than 100 sources:

    - only about 2-3 sources are OAI-compliant(!)
    - ... but the most important source is
      (arXiv.org, about 70% of import volume traffic)

- Current harvesting mechanism relies on arXiv.org email submission system

    - richer metadata content
    - ... but non-OAI

- Need richer metadata format, current `oai_dc` is very spartan

- Need more OAI-compliant data sources