

# Managing Institutional Repository with CERN Document Server Digital Library Software

Diane Berkovits<sup>1</sup>, Martin Vesely<sup>1</sup>

<sup>1</sup> European Organization for Nuclear Research,  
CH-1211 Geneva 23

**Abstract.** Institutional repositories are becoming a standard self-archiving instrument for sharing scientific documents among researchers across scientific institutions. Evolution of institutional repositories was significantly accelerated by specification of the OAi-PMH protocol for metadata harvesting. We discuss the practical aspects of the OAi-PMH and its impact on scientific communication and we demonstrate the key features of CERN Document Server Software - an institutional repository management system developed and used at CERN.

**Keywords:** institutional repositories; metadata harvesting; self-archiving; Open Archives initiative

In recent years scientific institutional repositories have gained more attention due to the increased availability of self-archived documents and institutional open access. Among the main issues raised in the scientific community, one of the most important in the perspective of the present contribution are the technological aspects of scientific communication and document sharing. In particular, we focus on metadata harvesting as the main technological framework for metadata sharing using standardized solutions such as the Open Archives initiative protocol for metadata harvesting (OAi-PMH).

OAi-PMH aims at developing scalable and robust technology for dissemination and sharing of scientific documents. One of the central ideas is to construct upon the originally centralized self-archiving practice put in place by Ginsparg's initiative in early 1990's and to create a distributed and globally open document sharing model on top of it. In this model, metadata about self-archived documents can be obtained by partner research institutes by harvesting.

At CERN, the electronic institutional repository powered by CERN Document Server digital library software has been developed since 1996 and became OAi compatible in 2002 with initial volume of 21366 OAi harvestable documents organized in several ad-hoc OAi collections covering CERN preprints in theoretical and experimental High Energy Physics. Later, the OAi collections were continuously being extended in number and volume, including CERN theses and CERN conference contributions reaching today's 76435 records with a monthly increase of ca 1000 metadata records.

In CERN Document Server, OAi collections are organized according to the arXiv.org standard. In practice each record that belongs to an OAi collection is attached an arXiv.org OAi set identifier. Document collections are decomposed, or more precisely structured, in hierarchical structures according to a domain specific and standardized classification scheme. We believe that harmonization of metadata about document collections is central to today's practice of OAi-PMH-based scientific document sharing. Indeed the organization of domain specific documents for targeted segments of readers has been a subject of several recent works and initiatives aiming at providing OAi sharable metadata. Note that for well described and interoperable document collections we sometimes use the term 'document server'. In this paper we will use the term 'institutional repository' as adopted in the scope of the Open Access movement. Worldwide, the CERN Document Server Software has become the biggest data provider across the different platforms installed.