The NDLTD Union Catalog: Issues at a Global Scale

Hussein Suleman
University of Cape Town
hussein@cs.uct.ac.za

Format: paper
Length of presentation: 20 minutes

The NDLTD Union Catalog is an international collection of ETD metadata that is harvested from various institutional, regional and cross-institutional collections. The Union Catalog has grown substantially in the 10 years since its launch and now contains almost 2 million records. However, various issues have surfaced during the maintenance of the Union Catalog and its downstream service providers. For example, at this scale, the well-known best practice of the OAI-PMH to restrict the size of a response to 100 records or 1MB has a severe impact on harvesting time. This paper describes this problem and other issues that are relevant to the Union Catalog and similar projects. For each such issue, solutions are discussed. Together these present a set of guidelines not only for large union catalogues but also for the design of large digital library collections in general.