

Harvard Scholarship Indexing Project (HSIP) – Progress Report

August 30, 2013

Introduction

The purpose of this project is to create a metadata repository for Harvard faculty scholarship. We will accomplish this by adding “push” APIs to the Harvard Faculty Finder (HFF) website to enable the owners of publication databases across Harvard to send metadata about those publications to HFF in a standardized way.

Progress to Date

We created two database scripts. The first parses publication data in Scopus XML format in order to extract individual elements, such as the author names, article title, and publication date. It saves the result in a staging area within the HFF database. The second script takes data in the staging area, removes any publications that already exist in HFF, and loads the remaining publications into the main HFF database tables.

Next Steps

The first script currently only parses Scopus XML data. We will build additional parsers for EndNote XML, RIS, VIVO RDF, and Web of Science XML. We will also wrap the scripts in a web service so that data can be sent to them remotely.

Budget Justification

Implementing the Library Lab project required creating a new development environment within HMS IT to host a copy Harvard Faculty Finder (HFF). The project manager, Steve Wimberg, coordinated efforts within HMS IT to load the database, setup file directories for the website and APIs, and configure firewall rules. He also assisted in the testing of the new database scripts to load publication data into HFF and parse Scopus XML data.