DLF 2011 Forum, Baltimore, October 31 – November 2, 2011

Presentation proposal

UDFR: A Semantic Registry for Format Representation Information

Lisa Colvin, Abhishek Salve, and Stephen Abrams

Abstract

A deep understanding of the diverse digital formats used to represent content is fundamental to effective long-term preservation analysis, planning, and intervention. The Unified Digital Format Registry project (UDFR, https://bitbucket.org/udfr/main/wiki/Home) will provide an open source platform for an online, semantically-enabled registry of significant format representation information, building on the experience of the existing PRONOM and GDFR registries. The UDFR project, now at its midpoint, uses a semantic wiki to manage and expose its holdings. This presentation will review the progress to date and describe plans for the remainder of the UDFR project activities.

Description

A deep understanding of the diverse digital formats used to represent content is fundamental to effective long-term preservation analysis, planning, and intervention. The Unified Digital Format Registry project (UDFR, https://bitbucket.org/udfr/main/wiki/Home) will provide an open source platform for an online, semantically-enabled registry of significant format representation information, building on the data modeling experience and collection activities of the existing PRONOM and GDFR registries and supporting a superset of their combined features. The UDFR project, now at its midpoint, uses a semantic wiki to manage and expose its holdings as linked data for interoperability with the semantic web. UDFR project staff have evaluated a number of potential semantic platforms for use as the core of the UDFR technology stack, finally choosing OntoWiki and the Virtuoso triple store on the basis of their comprehensive feature sets. This presentation will review the progress to date – including stakeholder community building, ontology development, and technology evaluation – and describe plans for the remainder of the UDFR project activities leading towards an operational registry initially populated with an export of PRONOM and GDFR data early in 2012.