

Developing a Collection Registry for IMLS National Leadership Grant Digital Collections

Goals of the IMLS Digital Collections & Content Project (Funded 2002-2005)

- **Build a registry and search and discovery tools** for all *National Leadership Grant* (NLG) collections with digital content.
- **Assist and guide** NLG projects in making item-level metadata sharable via the Open Archives Initiative Protocol for Metadata Harvesting.
- **Build a repository and search and discovery tools** for integrated access to the item-level metadata in NLG collections.
- **Research** how resource developers can best represent collections and items to meet the requirements of divergent service providers and user communities.

Stage 1:

What to describe?
Why?

Stage 1: What to describe? Why?

- What is the purpose of the collection registry?
- Do we describe projects or collections? Institutions? Agents? Sub-collections?
- Just what is a 'collection' anyway?

Some definitions of 'collection'

An aggregation of individual items based on any number of criteria (Johnston and Robinson)

Information environments or contexts of selected and ordered resources which facilitate information seeking by users (Lee)

An aggregate of items, which is maintained by an actor following a plan of cultural relevance over time. A collection is designed for a certain public, and the conservation of the collected items is normally catered for. (CIDOC CRM)

Primary purpose: to enable **discovery** of and **access** to digital collections

Primary focus: description of digital collections

- Brief entities for NLG projects, institutions, agents
- Allow multiple collections for projects and multiple institutions for collections and projects
- Allow sub-collections and super-collections
- Criteria for designating a collection (all required)
- **Cohesive** (topic area, type of material, etc)
- Searchable as a **distinct** collection or sub-collection
- A **unique** point of entry (URL)

Criteria will be revisited as project progresses.

References:

- Johnston, P. & Robinson B. (2002) Collections and Collection Description, *Collection Description Focus Briefing Paper*, No. 1.
- Lee, H. (2000). What is a collection? [Electronic version]. *Journal of the American Society for Information Science*, 51 (12): 1106-1113.
- RSLP Collection Description Schema, 2003 <http://www.ukain.ac.uk/metadata/rslp/schema/>
- Dublin Core Collection Description, <http://www.ukoln.ac.uk/metadata/dcmi/collection-op-summary/2003-08-25/>
- Encoded Archival Description, <http://www.loc.gov/ead/ead.html>

Stage 2:

Review of Similar Projects

Stage 2: Review of Similar Projects

Schemas Examined:

RSLP Collection Description (CD) (UKOLN)
→ Describes collections, locations, agents, external relationships

Dublin Core CD (DCMI)
→ Based upon the RSLP CD.
→ Still in development

Encoded Archival Description (SAA/JOC)
→ Used for encoding archival finding aids
→ Descriptive summary of collection and contents

Sample of Collection Registries Examined:

NSDL: <http://www.nsl.org>
→ Science/technology/engineering/math
→ Qualified Dublin Core

Cornucopia: <http://www.cornucopia.org.uk/>
→ Registry of museum collections in UK
→ Variation of RSLP CD Schema

EnrichUK: <http://www.enrichuk.net/>
→ Registry of NOF funded digital collections
→ Variation of RSLP CD Schema

Stage 2 Decisions:

Develop schema based on RSLP CD.
→ Mature and in widespread use

Adapt as needed

- Define properties for digital coll.
- Drop properties specific to physical coll.
- Develop entity for project
- Adapt entities for institution/agents

Maintain mappings to Dublin Core CD

Next Steps:

- Database development (completed)
- Initial population of registry through survey
- Incorporate feedback into registry design
- Allow projects to verify/ enhance entries
- Browse / search and discovery interface designed
- Usability testing and refinement (beta release)
- Public release of IMLS Collection Registry

Stage 3:

Schema Development

Display Name (Property Name)	Definition / Correlation to DC Collection Description App. Profile (as of Sept. 2003)	O	R
Title of collection (title)	The name of the digital collection created by the NLG project / dc:title	N	N
Collection ID (identifier)	The unique identifier for the collection / dc:identifier	N	N
URL (hasServiceURL)	The URL(s) where an end user can access the collection / sp:cd:hasLocation	N	Y
Object represented (type)	The name of the object(s) or resource(s) represented by the digital items (e.g. lithograph) / dc:type	N	Y
Format (format)	The format(s) of the digital items in the collection / dc:format	N	Y
Size (format_extnt)	The number of digital items within the collection / dc:terms:extent	N	N
Frequency of additions (dc:cruid:periodicity)	Frequency of additions to the digital collection (e.g. closed, irregular, daily) / r:sp:dc:cruid:status	N	N
Language of items (language)	If textual, the language(s) of the items in the digital collection / dc:language	Y	Y
Audience (audience)	The primary audience(s) for the digital collection / dc:audience	Y	Y
Access (lights_accessRights)	A statement of any access restrictions placed on the digital collection / dc:terms:accessRights	Y	Y
Rights (rights)	Information about rights (copyrights, etc) held in and over the digital collection / dc:rights	Y	Y
Collection development policy (collection:policy)	A collection development policy or URL pointing to such a policy for the digital collection / r:sp:dc:cruid:status	Y	Y
Alternative access (hasService)	A web service(s) (e.g. OAI or Z39.50) available to access the digital collection (beyond the URL) / r:sp:dc:hasLocation	Y	Y
Metadata schema used (metadata:schema)	The metadata standard(s) used to describe items in the digital collection / NONE	Y	Y
Topic (subject)	Topic(s) (from GEM) that describe the overall topical content of the items in the digital collection / dc:subject	N	Y
Description (description)	A description of the digital collection / dc:description	Y	Y
Strength (strength)	An indication of the strength(s) of the digital collection / sp:cd:strength	Y	Y
Topic - person/org (subject_name)	A personal or corporate name(s) associated with the items in the digital collection / r:sp:cd:agentName	Y	Y
Coverage - geographic (coverage_spatial)	A place(s) or area(s) associated with most or all of the items in the digital collection / dc:terms:spatial	Y	Y
Time period (coverage_temporal)	A time period(s) associated with most or all of the items in the digital collection / dc:terms:temporal	Y	Y
Hosting institution (institution_hosting)	The name of the institution who is hosting the digital collection / sp:cd:owner	N	N
Contributing institution (institution_contributing)	The name of the institution(s) contributing to the digital collection / NONE	Y	Y
Parent collection (relation_isPartof)	The name of any second collection that contains the current digital collection / dc:terms:isPartof	Y	Y
Sub-collection (relation_hasPart)	The name of any other collection(s) that is contained within the current digital collection / dc:terms:hasPart	Y	Y
Physical collection (relation_hasPhysical)	The name of the physical collection(s) from which the current digital collection is derived / dc:relation	Y	Y
Other associated collection (relation)	The name of any other collection(s) that is associated with or complements the current digital collection / dc:relation	Y	Y