



# The Distributed Library: OAI for Digital Library Aggregation

UIUC's Role:

Registry of OAI Data Providers

<http://gita.grainger.uiuc.edu/registry>

Thomas G. Habing

Grainger Engineering Library Information Center  
University of Illinois at Urbana-Champaign

[thabing@uiuc.edu](mailto:thabing@uiuc.edu)



# Catalyst for the Registry

- Various OAI Related Projects at UIUC
  - Mellon Funded Gateway to Cultural Heritage Material
  - IMLS Digital Collections and Content
  - NSDL Digital Mathematics Resources
  - CIC OAI Metadata Harvesting Service
  - OAI Portal to Engineering, Physics, and Computer Science

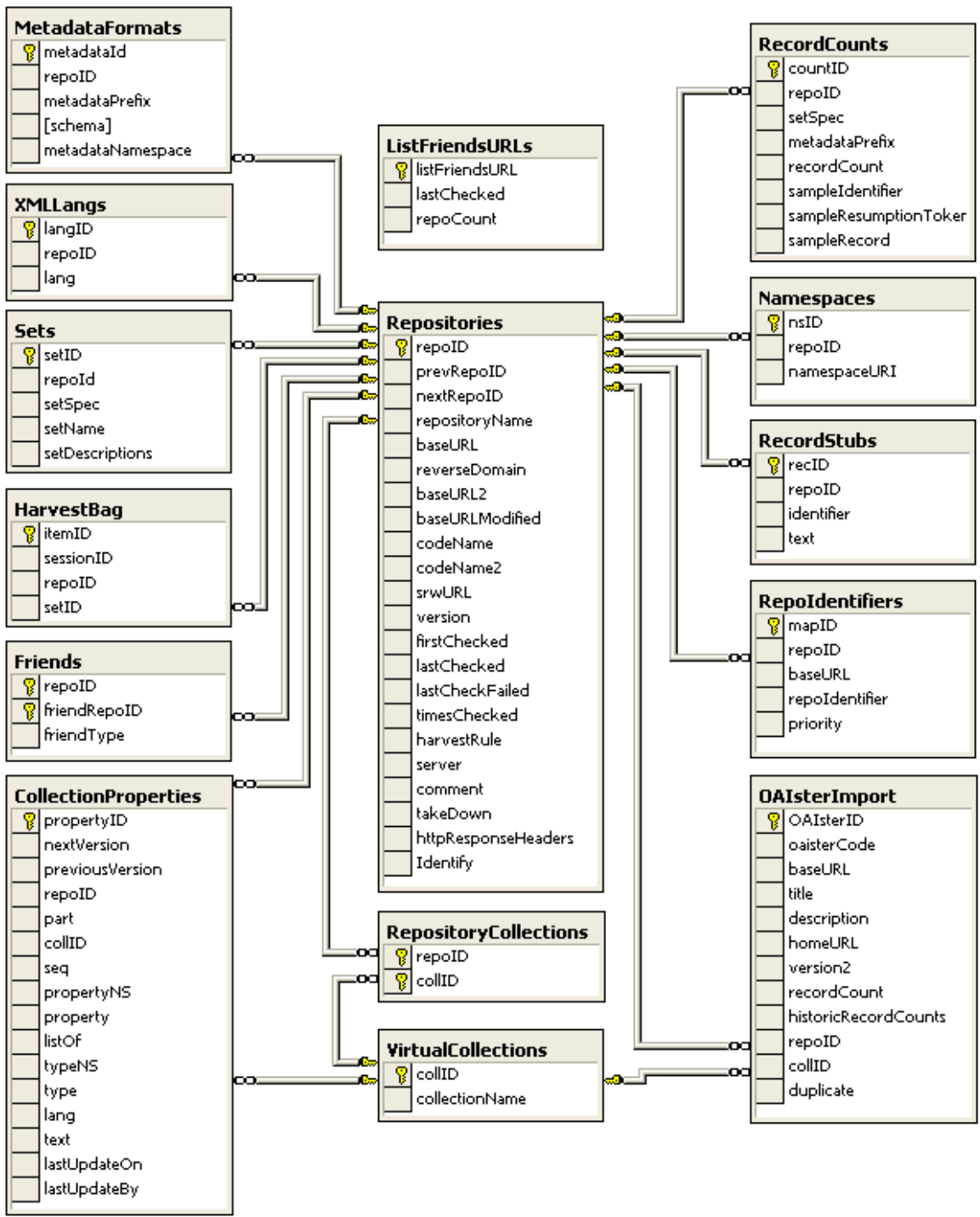


# Why Another OAI Registry

- Difficulty finding relevant OAI repositories to harvest, especially in selecting individual sets within a repository
- Other questions:
  - Which repositories are aggregating other repositories?
  - How big are individual repositories?
  - How to find new repositories as they come online?
  - What is the state of the OAI world?
- Enable other meta services



# Relational Database





# Sources

- Various other registries:
  - <http://celestial.eprints.org/cgi-bin/ListFriends.pl>
  - <http://oaisrv.nsd.cornell.edu/Register/ListFriends.pl>
  - [http://www.language-archives.org/register/archive\\_list.php4](http://www.language-archives.org/register/archive_list.php4)
  - <http://www.mds.rmit.edu.au/~ajk/oai/interop/targets.xml>
  - <http://www.openarchives.org/OAI/RejectedSites>
  - <http://www.openarchives.org/Register/ListFriends.pl>
- Google ([allinurl:verb=Identify](#))
- friends
- provenance



# Features

- Many ways to Search and Browse
- Web Interface
- Harvestable via OAI
- RSS
- SRU
- Many ways to customize repository lists
- Many misc. reports and lists
- Harvest Bag (still in work)



# Features added for the DLF Distributed Library Project

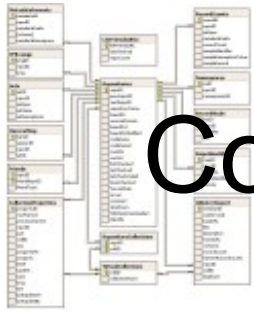
- Collection-level Description
  - Imported from OAIster
  - Detailed Manual Cataloging for DLF-Member Repositories
    - Sets can be Described as Sub-collections
    - Includes: GEM & LC Subject Heading, Item Types, Temporal & Spatial Coverage, Audience, Etc.
    - 33 Repositories, 155 Sets, 6853 Properties
  - Schema Conforms to the IMLS DCC Profile
    - <http://imlsdcc.grainger.uiuc.edu/about.htm#documentation>



# How to Access the Collection Descriptions

- Web Interface (Browse and Search)
- OAI <http://gita.grainger.uiuc.edu/registry/pxtest/oai.asp>
  - Each OAI repository is represented by one item
  - oai\_dc and imlstdcc metadata formats
- OAI gateway that exposes DLF-member repositories from one base URL  
<http://gita.grainger.uiuc.edu/registry/dlfgatewayplus/oai.aspx>
  - Each repository and set becomes a set in the gateway with the rich description contained in the <setDescription>





# Collection Description Next Steps

- Finish manual cataloging
- Explore automated enhancements
  - Harvest and analyze records and resources to embellish collection description
  - Also to embellish item-level search
  - Investigating Emory's MetaCombine Web Services to automatically cluster OAI metadata records
    - The clusters could be used to describe the collection?
    - Emory's command-line interface
    - We have also developed a simple web interface:
      - <http://gita.grainger.uiuc.edu/registry/cluster/doi.aspx>



# Collection Properties

repoID	set	propertyNS	property	listOf	typeNS	type	lang	text
848	NULL	http://purl.org/dc/elements/1.1/	identifier	0	NULL	NULL	NULL	RepoID=848
848	NULL	http://purl.org/dc/elements/1.1/	title	0	NULL	NULL	NULL	California Digital Library (CDL)...
848	NULL	http://purl.org/dc/elements/1.1/	title	0	NULL	NULL	NULL	California Digital Library (CDL)
848	NULL	http://purl.org/dc/terms/	abstract	0	NULL	NULL	NULL	The CDL Collection contains Coun...
848	NULL	http://purl.org/dc/elements/1.1/	type	0	http://purl.org/dc/terms/	DCMIType	NULL	Collection
848	NULL	http://purl.org/dc/elements/1.1/	type	0	http://example.org/cld/...	CLDType	NULL	CollectionCollection
848	NULL	http://purl.org/dc/elements/1.1/	language	0	http://purl.org/dc/terms/	ISO639-2	NULL	eng
848	NULL	http://example.org/cld/terms#	logo	0	http://purl.org/dc/terms/	URI	NULL	http://alcme.oclc.org/oaicat/oi...
848	NULL	http://example.org/gen/terms#	isLocatedAt	0	http://purl.org/dc/terms/	URI	NULL	http://www.cdlib.org/
848	NULL	http://example.org/gen/terms#	isAccessedVia	0	http://purl.org/dc/terms/	URI	NULL	http://oai.cdlib.org/
848	NULL	http://purl.org/dc/terms/	hasPart	0	NULL	NULL	NULL	coca:numeric
848	NULL	http://purl.org/dc/terms/	hasPart	0	NULL	NULL	NULL	escholeditions:ucpress:public
848	NULL	http://purl.org/dc/terms/	hasPart	0	NULL	NULL	NULL	escholeditions:ucpress:uonly
848	NULL	http://purl.org/dc/terms/	hasPart	0	NULL	NULL	NULL	images
848	NULL	http://purl.org/dc/terms/	hasPart	0	NULL	NULL	NULL	oac:ead
848	NULL	http://purl.org/dc/terms/	hasPart	0	NULL	NULL	NULL	oac:text
848	NULL	http://purl.org/dc/terms/	isReferencedBy	0	http://purl.org/dc/terms/	URI	NULL	http://www.cdlib.org/news/publis...
848	NULL	http://purl.org/dc/elements/1.1/	publisher	0	http://purl.org/dc/terms/	LCSH	NULL	California Digital Library
848	NULL	http://imlsdcc.grainger.uiuc...	managedBy	0	http://purl.org/dc/terms/	URI	NULL	dloy@ucop.edu
848	NULL	http://imlsdcc.grainger.uiuc...	metadataSchema	0	http://imlsdcc.grainger...	MetadataSchema	NULL	oai_dc
848	coc...	http://purl.org/dc/elements/1.1/	identifier	0	NULL	NULL	NULL	coca:numeric
848	coc...	http://purl.org/dc/elements/1.1/	title	0	NULL	NULL	NULL	Counting California : numeric da...
848	coc...	http://purl.org/dc/terms/	abstract	0	NULL	NULL	NULL	This collection provides access ...
848	coc...	http://purl.org/dc/elements/1.1/	format	0	http://purl.org/dc/terms/	IMT	NULL	text/html
848	coc...	http://purl.org/dc/elements/1.1/	format	0	http://purl.org/dc/terms/	IMT	NULL	text/csv
848	coc...	http://purl.org/dc/terms/	extent	0	NULL	NULL	NULL	8977 records
848	coc...	http://purl.org/dc/elements/1.1/	type	0	http://purl.org/dc/terms/	DCMIType	NULL	Collection
848	coc...	http://purl.org/dc/elements/1.1/	type	0	NULL	NULL	NULL	Numeric tables
848	coc...	http://purl.org/dc/elements/1.1/	type	0	http://example.org/cld/...	CLDType	NULL	CollectionDataset
848	coc...	http://purl.org/dc/elements/1.1/	type	0	http://purl.org/dc/terms/	DCMIType	NULL	Dataset



# Other Recent Enhancements

- Browse By Subject Headings
- Lists of Related Repositories
  - Including a couple summary reports



# More enhancements to come

- Advanced search
- Adding OAI Service Providers to the Registry
  - Which service providers are utilizing which repositories
- Self-service



# Questions