

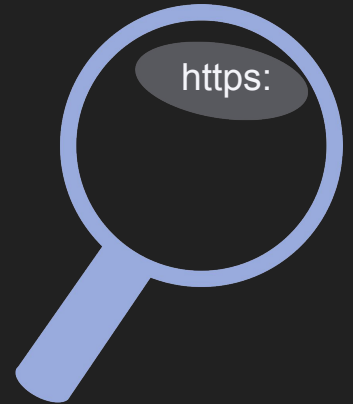
Supporting Sustainable Lookup Services

Steven Folsom, Cornell University Library

<https://orcid.org/0000-0003-3427-5769>

SWIB 2023

bit.ly/SustainableLookups



Background and Outline

- LD4P Authority Lookup Service (Questioning Authority tooling presented at previous SWIBs)
- Requirements for robust lookup services
- Opportunities and challenges of relying on caching
- Our transition to a more API-based approach (rather than caching)
- Work to support efforts to support caching in the future

What I'm talking about when I talk about lookups.

speech

Lookup with: LOC all genres (QA), MESH publication types (QA), GETTY_AAT Objects__Object_Genres (QA)

Add "speech" as literal or select from:

LOC all genres (QA) (7) MESH publication types (QA) (0) GETTY_AAT Objects__Object_Genres (QA) (8)

Speeches

Alternative Label: Televised speeches, Radio speeches, Academic addresses, Addresses (Speeches), Filmed speeches, Speechlets, Talks (Speeches), Public addresses, Legislative speeches, Orations

Citation note: (speech: A talk or formal address given to a public audience, extemporaneously or from a prepared text. In the 18th and 19th centuries, speeches were often separately published in print. Today, major public speeches are published in the New York Times and in the biweekly periodical Vital Speeches of the Day. Landmark speeches are usually anthologized.), (Legislative speeches), Commencement speech (A commencement speech or commencement address is a speech given to graduating students, generally at a university, generally in the US, although the term is also used for secondary education institutions) Baccalaureate service: (Baccalaureate addresses can range in length from under half an hour to as long as four hours) Inauguration (The "inaugural address" is a speech given during this ceremony which informs the people of his/her intentions as a leader) State of the Union (The State of the Union is the address presented by the President of the United States to a joint session of the United States Congress, typically delivered annually; State of the Union address) State of the State address (The State of the State Address is a speech customarily given once each year by the governors of each of the states of the United States (although the terminology for this speech differs for some states: in Iowa, the speech is called the Condition of the State Address; in Kentucky, Massachusetts, Pennsylvania, and Virginia it is called the State of the Commonwealth Address.) State of the City address (The State of the City Address - or State of the City Speech - is a speech customarily given once each year by the mayors of many cities in the United State, Canada, and the Philippines. Other international cities have also adopted the tradition), (speeches (documents). UF addresses (documents). Documents containing the text of any public address or talk. NT lectures; monologues; orations; sermons), (talk 1.2. An informal address or lecture), (Speeches Use Addresses; Addresses. UF Orations; Speeches. BT Discursive works. NT Academic addresses; Convicts' addresses; Lectures; Legislative addresses; Occasional addresses; Sermons), (talk: the act of talking formally about something before a group of people : a speech or lecture), p. 77 (speech: talk, conversation, discourse; statement, remark; formal address, oration; printed copy of this; speechlet: short speech)

Citation source: Reitz, J.M. ODLIS : online dictionary for library and information science, July31, 2014, Genre/form terms for law materials [report prepared by the Technical Services Special Interest Section of AALL], 2010.; Wikipedia, July 31, 2014.; Art & architecture thesaurus online, July 31, 2014, Oxford dictionaries website, Sept. 7, 2016, Genre terms : a thesaurus for use in rare book and special collections cataloging, via WWW, July 31, 2014, Merriam-Webster dictionary online, Sept. 7, 2016, The 10 best graduation speeches of all time, via CNBC website, viewed July 31, 2014., Görlach, M. An alphabetical list of English text types, in Text types and the history of English, c2004:

Citation status: found

Broader: Discursive works

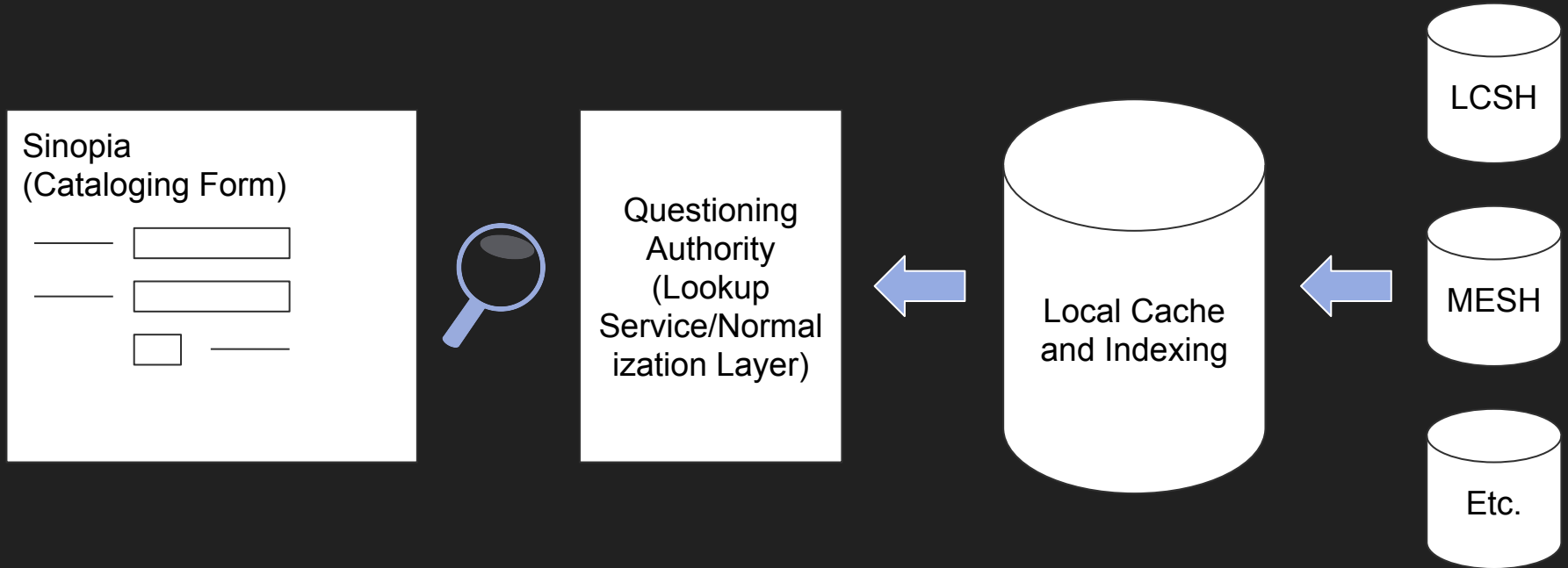
Narrower: Toasts (Speeches), Occasional speeches

Type: <http://www.loc.gov/mads/rdf/v1#Authority>, <http://www.loc.gov/mads/rdf/v1#GenreForm>, <http://www.w3.org/2004/02/skos/core#Concept>

Toasts (Speeches)

Citation note: (Toast (honor). A toast is a ritual in which a drink is taken as an expression of honor or goodwill. The term may be applied to the person or thing so honored, the drink taken, or the verbal expression accompanying the drink.), (toast: A call to a gathering of people to raise their glasses and drink together in honor of a person or thing, or an instance of drinking in this way), (toast: an occasion when people all drink together and say someone's name in order to express their admiration or their good wishes). (toast: a salutation or a few words of

LD4P Environment



Benefits of Caching RDF and Local Indexing

- Ability to determine what is searched and relevance
- Ability to determine what data is returned for display
- Uptime even if the vocabulary itself goes down

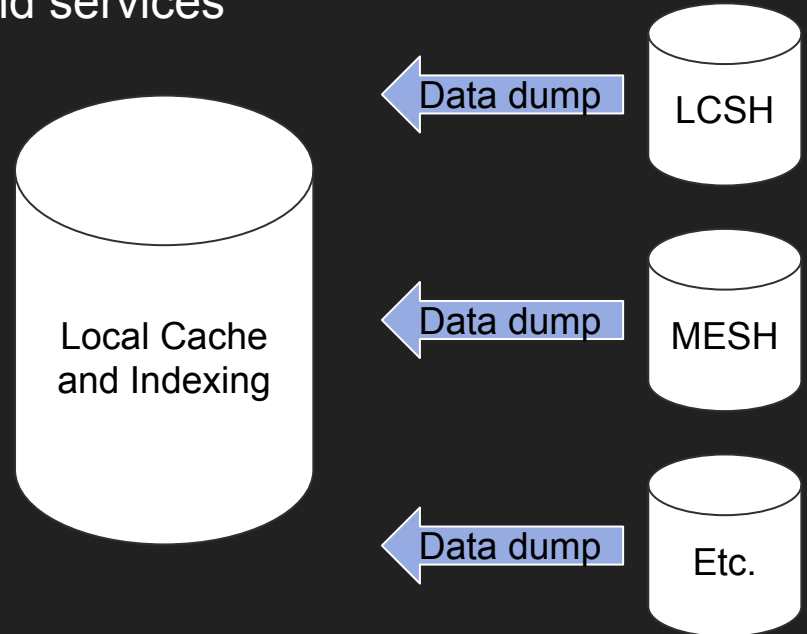


Meeting Cataloger Needs

A	B	C	D	E
EntityType	LDPath	Search	Display	RankingNotes
URI for the class of entity in the lookup. [Still trying to define best practices for noting when the the URI for the class (e.g. madsrdf:DeprecatedClass) is what we want to search on, but return a different URI. For now add an asterisk at the end of the URI and describe the situation in more detail in the Use Case column.]	LDPath to get to the information to be indexed in QA, see: https://marmotta.apache.org/ldpath/language.html . Confirm or add namespaces to the namespace tab.	Use an 'X' to mark if this data should be used to search against. N.B. some data is important to display to the cataloger, but perhaps would create messy results if searched against in a lookup environment, e.g. some notes are administrative in nature.	Use an X to mark if the value should be displayed. Include a label for the field. The label may simply be the property name in the property path column or you may decide another term is more appropriate.	If applicable provide notes on whether a particular property path should have a higher search ranking than others.
skos:Concept	Preferred label = skos:prefLabel	X	X - Preferred label	Rank above skos:altLabel
skos:Concept	Alternative label = skos:altLabel	X	X - Alternative label	Rank below skos:prefLabel
skos:Concept	Scope note = skos:note		X - Scope note	Shouldn't be used in searching (or if it is, should be weighted much lower than other notes)
skos:Concept	Citation note = madsrdf:hasSource / madsrdf:citation-note		X - Citation note	Shouldn't be used in searching (or if it is, should be weighted much lower than other notes)
skos:Concept	Citation source = madsrdf:hasSource / madsrdf:citation-source		X - Citation source	Shouldn't be used in searching (or if it is, should be weighted much lower than other notes)
skos:Concept	Citation status = madsrdf:hasSource / madsrdf:citation-status		X - Citation status	
skos:Concept	Broader = skos:broader / skos:prefLabel	?	X - Broader	Torn about whether to search these values to bring back the related term
skos:Concept	Narrower = skos:narrower / skos:prefLabel	?	X - Narrower	Torn about whether to search these values to bring back the related term
skos:Concept	Related = skos:related / skos:prefLabel		X - Related	
madsrdf:DeprecatedAuthority	rdf:type		X - Deprecated	
madsrdf:DeprecatedAuthority	Label = madsrdf:variantLabel	X	X - Label	Rank below skos:prefLabel of the Authority replacing this deprecated Authority
madsrdf:DeprecatedAuthority	Use instead = madsrdf:useInstead			
madsrdf:DeprecatedAuthority	Use instead = madsrdf:useInstead / skos:prefLabel		X - Use instead	
madsrdf:DeprecatedAuthority	Deletion note = madsrdf:deletionNote		X - Deletion note	Shouldn't be used in searching (or if it is, should be weighted much lower than other notes)

Costs and challenges with caching

- Indexing optimization for each vocabulary
- Keeping the data current, relying on data dumps
- Resources to maintain environment and services



Transition to API-based Lookups*

Sinopia
(Cataloging Form)

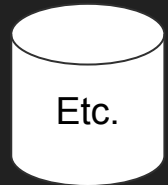
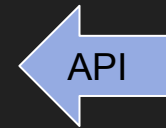
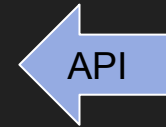
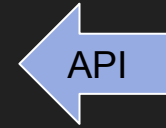
—

—

—



Questioning
Authority
(Lookup
Service/Normal
ization Layer)



Transition to API-based Lookups *and Local Lookups*

Sinopia
(Cataloging Form)

—

—

—



Questioning
Authority
(Lookup
Service/Normal
ization Layer)

Local
Lookup
Lists
(when API
isn't
available)

Plan for swapping out caches for API-based lookups

- Identify critical data sources
- Background analysis
 - Are there already existing configurations in the community developed QA code base that we can turn on?
 - If not, identify search APIs for dataset to create new code/configurations
 - If no APIs, are the datasets small enough to create local pick list configurations (e.g RDA value lists)
- 2 sprints to create new lookups (happening later this month)
- Create test templates to make sure they work in Sinopia
- Communicate new lookups exist
- Give Sinopia template admins time to swap out, before removing cached lookups

Pros and Cons

What we gain:

- Lookups will be in sync with official vocabulary
- Selecting a lookup for template to create a Sinopia form will be simplified (don't have to choose between cached version, old cache versions, and direct)
- Once set up, low maintenance

What we give up:

- We are limited to data providers' API choices about search and display when we translate their API.
- If their API goes down, QA will not be able to provide search results to Sinopia.
- We still need a plan for data sources that lack a suitable search API for QA to translate and are too large/dynamic for a config file for each entity in the data source

Entity Metadata Management API

- LD4 Working Group is defining an API specification for communicating new, updated, deprecated and/or deleted entities within a dataset.

The image shows a screenshot of a document page for the Entity Metadata Management API. At the top left is the logo for EMM (Entity Metadata Management). The main title is "Entity Metadata Management API 0.1". Below the title, there are two columns. The left column contains a table of contents with sections: "Status of this Document", "1. Introduction" (with sub-sections 1.1 Objectives and Scope, 1.2 Use Cases, 1.3 Terminology), and "2. Architecture". The right column also has a "Status of this Document" section, which includes a warning box stating "This is a draft API specification. Expect changes to this document." and lists the following information: "This Version: 0.0.1", "Latest Stable Version: None", and "Previous Version: None".

EMM ENTITY METADATA
MANAGEMENT

Entity Metadata Management API 0.1

Status of this Document

1. Introduction

- 1.1. Objectives and Scope
- 1.2. Use Cases
- 1.3. Terminology

2. Architecture

Entity Metadata Management

Status of this Document

This is a draft API specification. Expect changes to this document.

This Version: 0.0.1

Latest Stable Version: None

Previous Version: None

LD4 Working Group

- Representation
 - Academic Libraries
 - National Libraries
 - Data Scientists
 - Domain Libraries
- Process
 - Regular meetings
- Working Spaces
 - GitHub
 - Approval process for pull requests
 - Slack
 - Communication
 - Asynchronous discussions between meetings

Editors

- [David Eichmann](#), School of Library & Information Science
- [Christine Fernsebner Eslao](#), Harvard Library
- [Nancy J. Fallgren](#), National Library of Medicine
- [Steven Folsom](#), Cornell University Library
- [Kevin M. Ford](#), Library of Congress
- [Jim Hahn](#), University of Pennsylvania Libraries
- [Kirk Hess](#), OCLC R&D
- [Anna Lionetti](#), Casalini Libri
- [Tiziana Possemato](#), Casalini Libri
- [Erik Radio](#), University of Colorado Boulder
- [E. Lynette Rayle](#), Cornell University Library
- [Gregory F. Reeve](#), Brigham Young University
- [Vitus Tang](#), Stanford University Libraries
- [Simeon Warner](#), Cornell University Library

Important moments in the lifecycle of an entity

Creation

New entities to incorporate into workflows. Important in the caching context to understand what entities are available for use.

Updates

Updates to existing entities are important in order to reflect the current state of an entity description, whether that be labels, definitions, other attributes. Some Updates require more attention than others.

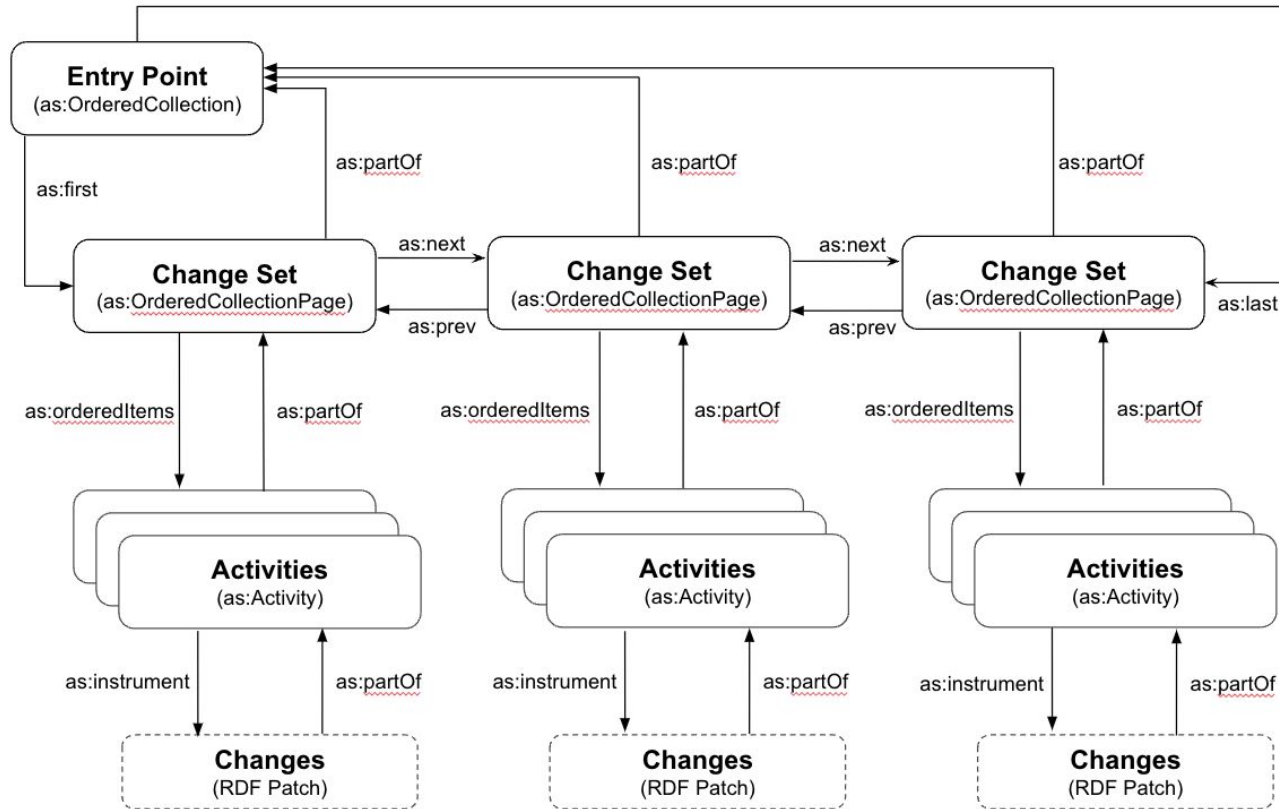
Deletes

Deprecation and deleting of entities are critical in understanding that an entity previously available for use is now discouraged from use.

Why Activity Streams?

- W3C Spec, widely used and supports linked data
- Prior Art
 - Getty Vocabularies Implementation
 - Library of Congress Implementation
- Sufficient Semantics for Minimal Requirements
 - Provides a model for Activities/Actions taken on a given entity
 - Dated
 - Typed
 - Not formal “Forces and Functions” evaluation Serven mentioned this morning, but we see a match with one way messaging (Awareness).
- Potential for extensions
 - Types
 - Can have additional information to share precise changes
-

EMM API Architecture using Activity Streams



Note: Changes using RDF Patch are optional.

Current status: Types

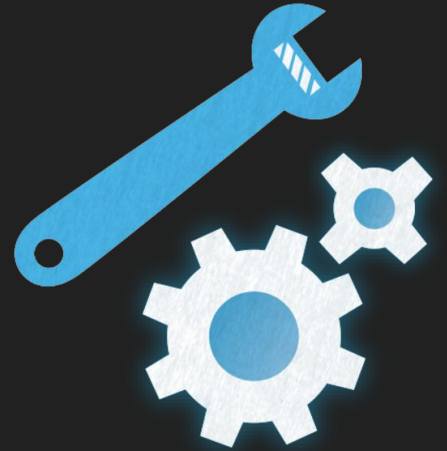
as:Creation/as:Add (For new entities to pay attention to)

as:Update (For changes to entities)

ex:Deprecation (For entities that still exist but are discouraged from use)

as>Delete (For entities that no longer deference... *a paper trail*)

Concluding this round of API spec work



Conclusion: We maintain hope for cache-based lookups

- We've learned a lot about what data sources cataloger's want to link to and lookup expectations for these specific vocabularies.
- Our decision to turn to API-based lookups was largely based on resources when the LD4P Grant ends.
 - Maintaining caches should be possible with pooled resources, perhaps by consortia and/or vendors.
- There are emerging practices (Activity Streams) that could really improve the ability to maintain the currentness of caches.

Links

QA:

- https://lookup.ld4l.org/authority_list
- https://github.com/LD4P/qa_server

Entity Metadata Management API:

- https://github.com/LD4/entity_metadata_management
- LD4 Slack Channel: [authority_api_wg](#)