Discovery Futures: Linked Data & AI

Tom Cramer
Associate University Librarian
Stanford University
@tcramer

CNI “San Diego”
May 21, 2020
Linked Data & Libraries

Linked Data is the gateway between library data and the Web.

- Get library data onto the web
- Get web data into the library
Discovery is where...

...we will see tangible benefits of LOD
Linked Data & Library Discovery

1. Schema.org (et al.) mark up
2. Knowledge Panels
3. Semantic Browse
4. Semantic Search
5. Type Ahead / Auto-Suggest
Knowledge Panel

**BibCard implementation from University of Wisconsin**

- **https://search-ld.library.wisc.edu/**
- **“Gertrude Stein on Picasso”**
Semantic Browse

Author

Subject

Place
Semantic Search

Semantic Search is a feature of the U.S. National Library of Medicine (NIH) that allows users to search for medical terms using MeSH (Medical Subject Headings) descriptors. This search engine provides detailed information about various medical terms, including their definitions, synonyms, and related concepts.

The example shown in the image is a search for "heart attack". The search results display the MeSH heading for Myocardial Infarction, which includes a unique ID, tree number(s), and additional information such as scope notes and entry terms. The scope note explains that myocardial infarction is NECROSIS of the myocardium caused by an obstruction of the blood supply to the heart, and it is not synonymous with acute disease like "acute infarct".

Other terms related to myocardial infarction include Cardiovascular Stroke, Heart Attack, and Myocardial Infarct, which are also listed under the entry terms section. The NLM Classification # for myocardial infarction is WG 310.
Type Ahead / Auto-Suggest
LD4 Knowledge Panel Recipe

DRAFT (see original outline in separate doc)

Table of Contents (proposed new structure)

1. Introduction
2. What is a Knowledge Panel, Illustrated
3. Benefits of using Knowledge Panels
4. Data
   a. Sources
   b. Sufficiency (Minimum Viable Displayability)
   c. Example Queries
   d. Data Integrity and Ethical Concerns
5. Implementation
   a. Technical Strategies - Client Side vs. Server Side querying
   b. UX/Interaction considerations
6. Future Work

LD4 Discovery Affinity Group Charge

Rationale

There is a pressing need to provide concrete demonstrations of the benefits of linked data in order to help motivate the transition of library catalog from MARC. Within LD4P2, work to enhance the discovery of library resources is the focus of WP4: Discovery. The partners have committed to working with the Blacklight framework to demonstrate enhanced discovery features such as:

1. knowledge panel in search results to present contextual information powered by linked data;
2. browsing based on authority files and links to related entities in external data;
3. semantic search, which could include laying out alternative terms in a "no results" page, suggesting semantically related terms in a type-ahead, or providing richer, geo-based browse by leveraging URIs for places; and
4. microdata on item pages to enable machine crawling.

In all of these areas there is opportunity for community input on and discussion of the best approaches and the desired outcomes. The LD4 Discovery Affinity Group will provide a forum for these discussions, with the following goals:

- identify exemplar systems/features that demonstrate the benefits of linked data and could be applied to library resources
- identify opportunities for enhanced discovery of library resources using linked data that can be implemented in the short to medium term
- document connections between cataloging practices and discovery outcomes
- document effective assessment approaches
- provide advice to LD4P2 partners and cohort members for WP4: Discovery

• Biweekly open calls
• Extensive knowledge base

https://wiki.lyrasis.org/display/LD4P2/LD4+Discovery+Affinity+Group+Charge
AI & Discovery
AI: Two Big Opportunities

1. More (meta)data
2. New interfaces
AI
More Metadata
Metadata at Scale

Text
- Named Entity Recognition
- Text Classification

Images
- Labelling
- Object Recognition

Time-based Media
- Speech-to-Text
- Structural Analysis
New Interfaces
New Interfaces

Yewno Concept Explorer
Pathfinder Projects at Stanford

Enhanced Cataloging for Theses & Dissertations

**Question:** Can we use AI to automatically add keywords & subjects to catalog records?

**Models / Engines:** Yewno, ANNIF, BERT

**Key Points:** Integration with traditional cataloging workflow & ILS

Computer Vision for Archaeological Photos

**Question:** Can we usefully classify & mine under-described photos from a 20 year dig?

**Models & Engines:** Claif.ai, Google CV, AutoML

**Key Points:** Comparison of commercial vs. custom models, integration with dig records

[http://catalhoyuk.com/research/imageai](http://catalhoyuk.com/research/imageai)
AI4LAM is an international, participatory community focused on advancing the use of artificial intelligence in, for and by libraries, archives and museums.

https://ai4lam.org