Adding Structured Data to the NLM Digital Repository to Enhance Search Results

Kelly Quinn

Introduction

This project allowed the National Library of Medicine (NLM) to experiment with adding structured data to its Digital Repository resource pages to enhance search results displayed by Google and other search engines. Possible workflows were evaluated and a recommended process was determined. The project used Google Analytics’ Search Console to create a baseline for resource usage and to measure the project’s effectiveness.

Methodology

- Selected Schema.org ontology and JSON-LD linked data encoding based on Google’s recommendations.
- Created initial mapping of Schema.org to NLM’s DMDINDEX metadata schema.
- Selected two resources from the digital repository to receive the JSON-LD (the video Private SNAFU vs Malaria Mike and the book Applied Psychology and Scientific Living).
- Selected two other resources as a control group, which did not receive the JSON-LD.
- Configured Google’s Search Console to monitor statistics for the resources.
- Created JSON-LD manually and added the code to the resource HTML pages.

Next Steps

- Finalize DMDINDEX mapping to Schema.org.
- Create XSLT to transform MARCXML into JSON-LD for each resource format (e.g., book, image, and video).
- Add JSON-LD to all resources in the Digital Repository and monitor results.

Recommended Workflow

1. Retrieve bib records from ILS in MARC using z39.50
2. Obtain URIs using MarcEdit’s MARCNext tool
3. Transform MARC to MARCXML using MarcEdit
4. Transform MARCXML to JSON-LD using XSLT
5. Ingest JSON-LD datastream to Fedora via Fedora Object XML (FOXML)
6. Add JSON-LD to HTML resource pages

Sample Records

Preliminary Results

- Total Impressions:
  - Video: ▲1063%, ▲258%
  - Book: ▲4%

- Links to Our Site:
  - Video: ▲9%
  - Book: ▲20%

- Average Page Position:
  - Video: ▼5%
  - Book: ▼5%