Linked Open Dedup Vectors: An experiment with RDFa in Primo

Corey Harper
2011-09-12 – IGeLU
Outline

• Problem Statement
• How it works
  ▫ The code
  ▫ The output
• Challenges & Next Steps
• Leaving off the SEO
• Proof of concept – simplicity
• Value Proposition:
  ▫ FRBR & Dedup Ids
  ▫ Identifier Management & Equivalence
  ▫ Larger network of data sources
  ▫ Integrate selected metadata with other systems
- Understand vendor question of ROI
- Not all investment is “heavy”
- Low barrier to doing linked data
- Problems / fixes tend to be general web design issues:
  - RESTful architecture
  - Persistent IDs
  - “Thoughtful” redirects
Primo “FRBR-ization”
Mostly Manifestations

• Manifestations:
  ▫ Singleton Aleph Records
  ▫ DedupMrg Cluster
  ▫ DedupMrg groups have alt manifestations
  ▫ Unfortunately, IDs don’t resolve

• FRBR too rigid
• Bibo too flat
• Need Something *in between*
• Custom tileset; edited details.jsp (JSTL)
• Text strings from “display” elements
• Standalone div, property, content tags
• Convert to URIs for rel links?
• Jquery (or more JSP extensions?)
  ▪ Enrich with descriptions of related works / manifestations (TODO)
  ▪ Enrich with external data (TODO)
<div
xmlns:dc="purl.org/dc/elements/1.1/"
xmlns: [many more here]"
about="http://primodev1.bobst.nyu.edu/primo_library/libweb/action/dlDisplay.do?onCampus=false&docId=${result.id}&group=GUEST&vid=NYURDF&institution=NYU&fromLogin=true#book"
typeof="rdfrbr:Manifestation">
<div property="bibo:isbn"
    content="${result.values['isbn'][0]}">
</div>

<div property="dc:subject"
    content="${result.values['subject'][0]}">
</div>
<c:set var=dup val=${results.dedup}>
<c:choose>
    <c:when test="fn:startsWith(dup,dedup)">
        <div rel="equivelantManifestation">
            <c:forEach items="${result.values['lds02']}" var = "recid">
                <div typeof="rdfrbr:Manifestation">
                    <div property="dc:dentifier"
                         content="${recid}'"></div>
                </div>
            </c:forEach>
        </div>
    </c:when>
</c:choose>
<#book> a rdfsbr:Manifestation ;
    rdrel:equivalenceRelationshipsManifestation
        [ a rdfsbr:Manifestation ; dc:dentifier
            "nyu_aleph000594679"@en_US  ],
        [ a rdfsbr:Manifestation ; dc:dentifier
            "nyu_aleph002013659"@en_US  ] ;
    rdrel:workManifested
        [ a rdfsbr:Work ;
            dc:dentifier "21827719"@en_US
        ] .
Challenges and Next Steps

- Request Enhanced DL functionality
- Possible Vocamp at DC2011?
  - Vocabulary for ID Management
  - (More) Lateral FRBR Relationships
- Linking to external data
  - lccn, lcsh, lcnaf, viaf
- More data modeling
- Preparation of raw dumps
- Re-work blank-nodes
Why do this?

• Not for SEO, else I’d use Microdata. ☺
• Primo processing as part of ID Mgt Strategy
  ▫ Related Manifestations / Works =
  ▫ Related Identifiers!!
• Other Relationships:
  ▫ EAD to Bib to Collection to Digital Object
  ▫ For certain types of dedups, use LD fields for other identifiers.
Questions?

corey.harper@nyu.edu

@chrpr