

MARC to Schema.org:

Providing Better Access to UIUC Library Holdings Data

Tim Cole • Michael A. Norman • Patricia Lampron • William Weathers • Ayla Stein • M. Janina Sarol • Myung-Ja Han

INTRODUCTION

Taking advantage of the Web as a means for disseminating large datasets, libraries have begun publishing their bibliographic catalogs on the Web. Initially, most libraries focused on releasing their catalogs as MARCXML suitable for bulk downloading, however, MARC consists primarily of string data with few, if any, URIs linking to ontologies or related resources. MARC was not designed for use with RDF. Libraries (and cooperatives like OCLC) are now experimenting with disseminating catalogs as linked open data. Semantics compatible with RDF are being used, but specific schemes vary.

The University of Illinois at Urbana-Champaign (UIUC) Library has released 5.5 million bibliographic catalog records for sharing. These include detailed information about physical holdings at Illinois, allowing consumers to know what manifestations and items of the creative work described are available at UIUC. Catalog records are being released as MARCXML, as MODS enriched with links to name (VIAF) and subject authorities (LCSH), and as RDF using schema.org semantics.

LINKING WITH LCSH AND VIAF

VIAF SEARCH STEPS

1. Remove all punctuation (., - () ?)
2. Remove b., d., or fl. in front of years
3. Search for the exact name string in the VIAF database*
4. If no matches are found, search the VIAF database for a record that begins with the name string*

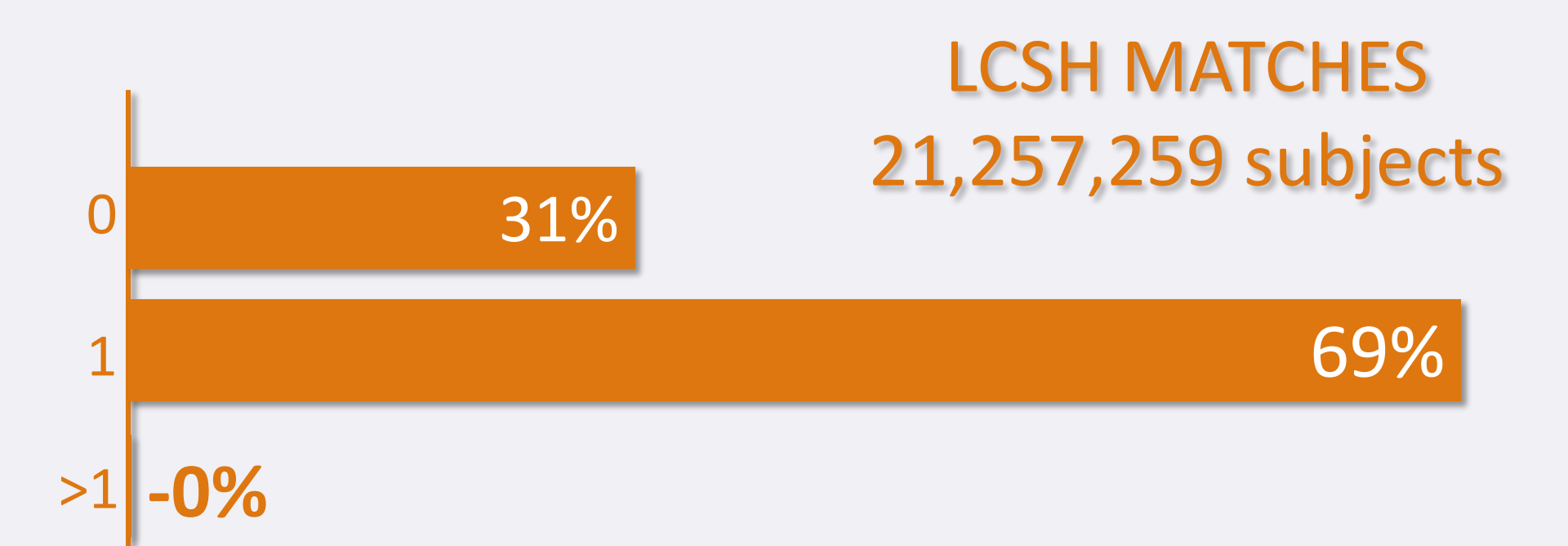
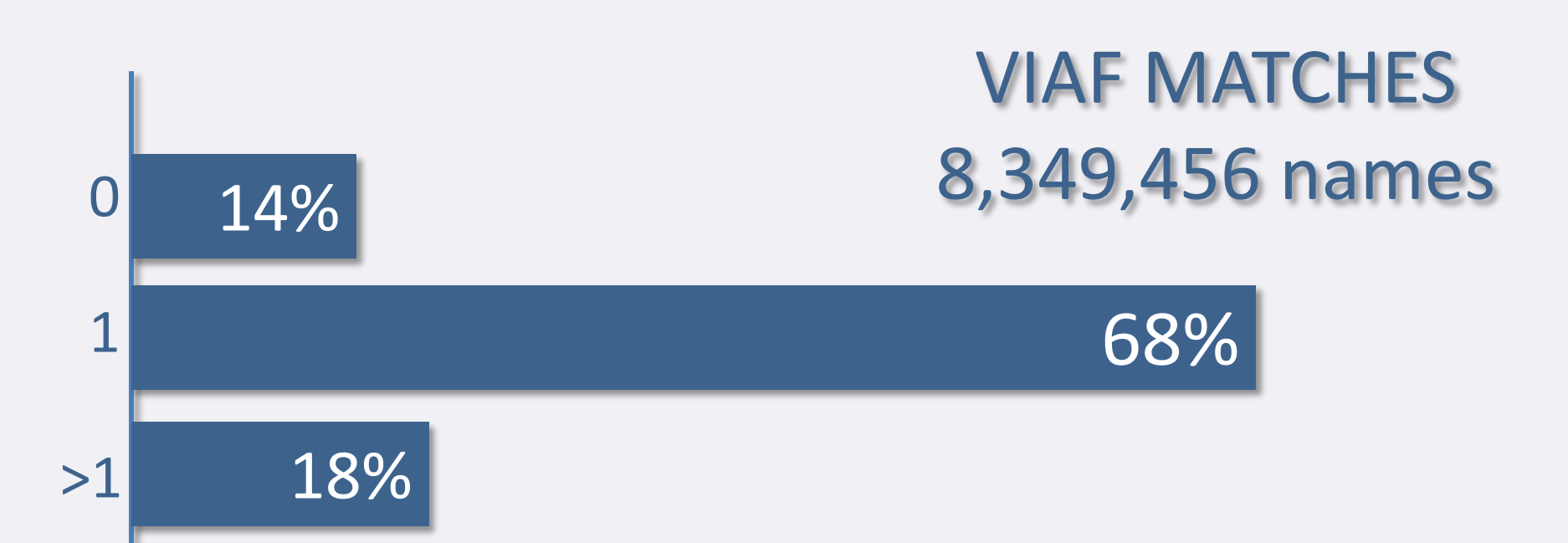
LCSH SEARCH STEPS

1. Search for the exact subject string, subject type, and heading type in the database*
2. If no matches are found, search for the exact subject string and subject type in the database*
3. If no matches are found, search for the exact subject string and heading type in the database*
4. If no matches are found, search for the exact subject string in the database*

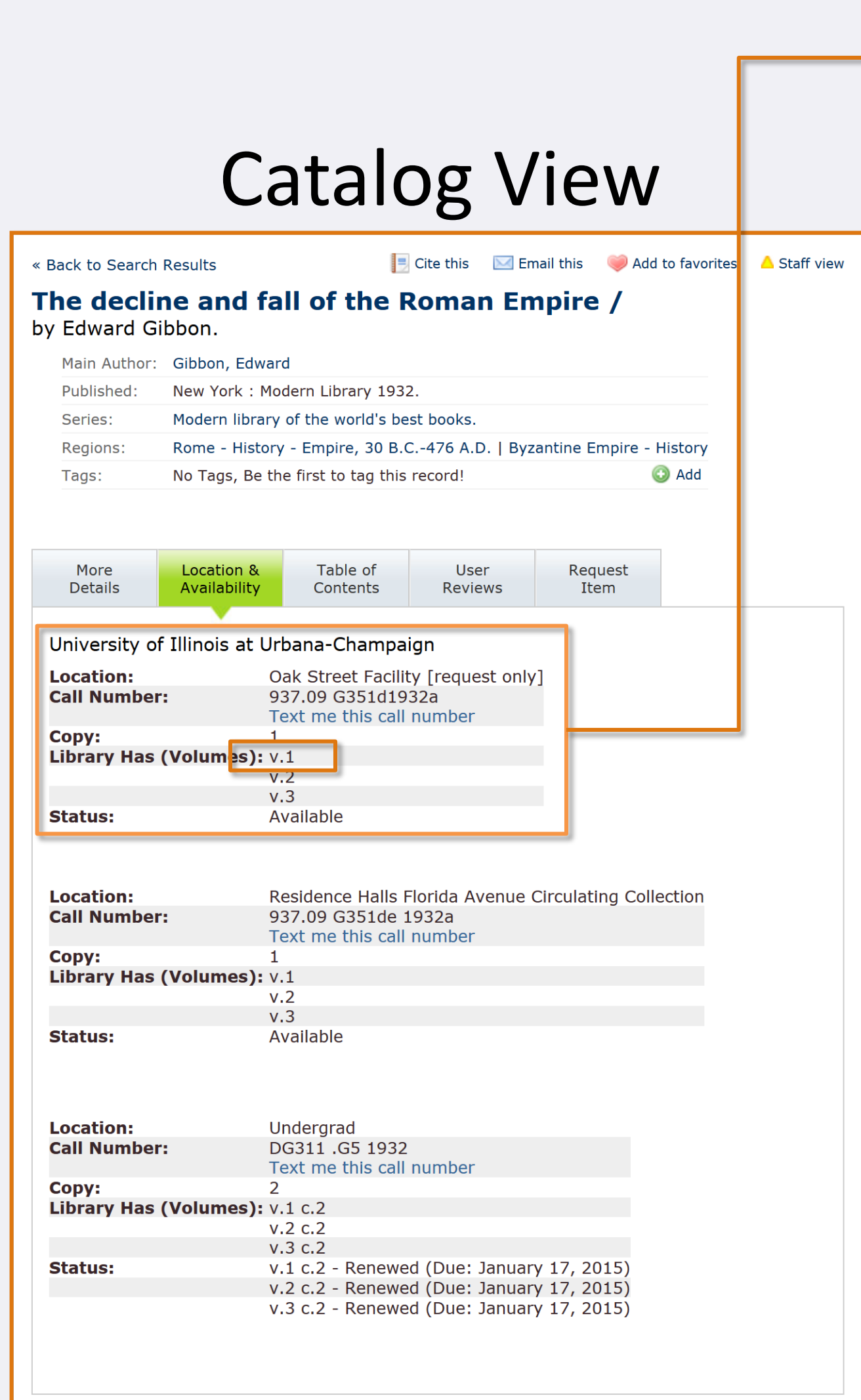
SEARCH ISSUES

1. Performance
 - Downloading data from VIAF and LCSH is faster than using APIs
2. Multiple Matches
 - Links are only added if there is exactly one match (*)
3. Data Quality (Records)
 - Acronyms, abbreviations, terms of address (Mrs., Sir, etc)
4. Data Normalization
 - Punctuation is removed from both MODS and VIAF names when searching
5. Name Type in MODS vs. VIAF
6. LCSH Authorized Headings vs. Topic Subdivisions

SEARCH MATCHES



HOLDINGS TRANSFORMATION STEPS



1. MARCXML
 - MARCXML bibliographic descriptions were first created for each holding with selected volume-specific information (e.g., barcode) recorded in the 955 local data field
 - A VB.NET program collapsed volume-level records associated with a single bibliographic entity into one record containing all holding information in repeated MARC 852 data fields
2. MARCXML to MODS
 - Each 852 data field is mapped to a MODS <location> element
 - Subfield \$a is mapped to sub-element <physicalLocation>; all other 852 subfields map to sub-elements of a single <copyInformation> element, within the <holdingSimple> sub-element
3. MODS to RDF with Schema.org
 - The MODS metadata enriched with links to name and subject authorities are transformed into schema.org semantics
 - Holdings information is transformed using a modified mapping to Offer based on recommendation from the W3C Schema Bib Extend Group

```
MARCXML
<!--marc:datafield tag="852" ind1="0" ind2="" -->
<marc:subfield code="3">v.1</marc:subfield>
<marc:subfield code="a">IU</marc:subfield>
<marc:subfield code="b">Oak Street Facility [request only]</marc:subfield>
<marc:subfield code="h">937.09</marc:subfield>
<marc:subfield code="j">G351d1932a</marc:subfield>
<marc:subfield code="p">30112048874520</marc:subfield>
<marc:subfield code="t">1</marc:subfield>
</marc:datafield>
```

```
MODS
<!--location-->
<physicalLocation displayLabel="Institution Code">IU</physicalLocation>
<!--holdingSimple-->
<!--copyInformation-->
<subLocation>Oak Street Facility [request only]</subLocation>
<shelfLocator>937.09 G351d1932a</shelfLocator>
<note displayLabel="Copy Number">1</note>
<note displayLabel="Barcode">30112048874520</note>
<enumerationAndChronology>v.1</enumerationAndChronology>
</copyInformation>
</holdingSimple>
</location>
```

```
Schema.org
<!--sc:hasPart-->
<!--sc:PublicationVolume-->
<sc:volumeNumber>v.1</sc:volumeNumber>
<!--sc:offers-->
<!--sc:Offer-->
<sc:seller rdf:resource="http://id.loc.gov/authorities/names/n79066210"/>
<sc:availableAtOrFrom-->
<sc:Place-->
<sc:name>Oak Street Facility [request only]</sc:name>
</sc:Place-->
<sc:availableAtOrFrom-->
<!--sc:itemOffered-->
<!--sc:IndividualProduct-->
<sc:sku>937.09 G351d1932a</sc:sku>
<sc:serialNumber>30112048874520</sc:serialNumber>
<sc:productID>Copy Number: 1</sc:productID>
</sc:IndividualProduct-->
</sc:itemOffered-->
</sc:Offer-->
</sc:offers-->
</sc:PublicationVolume-->
</sc:hasPart-->
```

