

HYBRIDIZING MARC

USING LINKED DATA IN YOUR ILS NOW

Created by [Galen Charlton](#) / [@gmcharlt](#)

Implementation Manager at [Equinox Software](#)

CHANGE IS HARD

- Change is unpredictable
- Change is expensive
- Change is time-consuming

THE WEIGHT OF HISTORY

Time and time again, libraries have found themselves in the position of being early adopters of technology.

In many cases, *TOO* early.

MARC

1966 First distribution of MARC records

1969 Invention of GML

1978 Start of SGML standardization effort

1998 XML becomes a W3C recommendation

1999 First RDF recommendation by W3C

NCIP

2000 SOAP 1.1 submitted to W3C

2002 NCIP 1.0 published

2003 SOAP becomes a W3C recommendation

2004-2005 JSON starts becoming popular

Oops

LEGACY OF LEGACY DATA FORMATS

If we were starting from scratch, we wouldn't invent MARC...

- Less ambiguity, easier to parse
- Design for sharing
- Design for easily communicating updates
- Design for easier tool-building

But there is also a positive legacy.

YES, WE ACTUALLY DO
KNOW SOMETHING ABOUT
METADATA

... AND WE HAVE A LOT OF IT

BURN THE WORLD DOWN!!!

Let's try explaining that to our auditors...

WAIT A FEW YEARS FOR SOMEBODY TO TELL US HOW TO SOLVE ALL OUR PROBLEMS

... then wait for the software to be implemented

- A play in two acts..
- Library: fix it!
- Vendor: tell us what you want!
- A play in infinite acts..

CONSTRAINTS

- Nobody has an infinite appetite for standards froth
- ... and increasingly, many libraries can't afford it
- The half-life of the MARC record is going to be measured in years... if not decades
- Centralizing all the metadata has its own issues

A PRACTICAL WAY FORWARD

- Embrace incrementalism
- Take advantage of the renaissance in tools to slice and dice MARC records
- Let the authority records save us

LINKING HEADINGS IN THE ILS

Both Koha and Evergreen support linking bibs to authorities the hard way.

I.e., normalized heading + thesaurus in bib record to corresponding authority record.

Subfield \$9 in the case of Koha, subfield \$0 in the case of Evergreen.

String-matching of headings is ugly, but works.

LINKING OUTSIDE THE ILS

For LC Names and LCSH, given the authority record, it's easy to compute the identifier.

NOT A NEW CONCEPT

... but it gives us leverage.

AUTHORITY RECORDS AND IDENTIFIERS

Authority records are a gateway to Linked Data

```
LDR 01135cz  a2200205n  4500
001      35
005      20060512103327.0
008      790705n|  acannaabn          |n  aaa
010      _an  79056654
035      _a(OCOLC)oca00289745
053  0  _aPS3552.U827
100  1  _aButler, Octavia E.
```

And here's a simple link, via an identifier....

```
LDR 01019cam a2200349 i 4500
005      20140627012826.0
008      790320s1979      nyu      000 1 eng
010      _a      78022597
020      _a0385150598 :
          _c$8.95
100 1    _aButler, Octavia E.
          _935
245 10   _aKindred /
          _cby Octavia E. Butler.
250      _alst ed.
...
999      _c2139
          _d2139
```

Also, a missed opportunity, at least in MARC21-land...

100	1	_	935
-----	---	---	-----

100	1	_	0(DLC)n	79056654
-----	---	---	---------	----------

IDENTIFIERS == LEAD

- **n 79056654**

IDENTIFIERS == GOLD

- <http://id.loc.gov/authorities/names/n79056654>

SO WHAT?

THE VIRTUAL INTERNATIONAL AUTHORITY FILE (VIAF)

- viaf.org
- Octavia Butler, again

BRIDGING THE GAP

First, let's figure out the VIAF identifier.

- [http://viaf.org/viaf/search?
query=cql.any="n 79056654"&httpAccept=application/xml](http://viaf.org/viaf/search?query=cql.any='n 79056654'&httpAccept=application/xml)

No SPARQL endpoint — data dumps are available, though

```
<ns2:Document about="http://viaf.org/viaf/34453955/">  
<ns2:inDataset resource="http://viaf.org/viaf/data"/>  
<ns2:primaryTopic resource="http://viaf.org/viaf/34453955"/>  
</ns2:Document>
```


THIS LOOKS FAMILIAR

- <http://viaf.org/viaf/34453955>
- <http://viaf.org/viaf/34453955/rdf.xml>

BRINGING IT BACK HOME

- Killing off dead authors
- Adding cross-references
- and we need not restrict our enrichment to the RDF -
[more is available](#)
- ... and do it automatically

LET'S RECAP

bib heading → authority record → LC control number of
authority record → id.loc.gov identifier → VIAF → RDF →
enrichment in existing ILS

POINTS OF ENRICHMENT

- Displays and links in the discovery interface
- ... and if you're not fortunate enough to be running an ILS that you have access to the source code for, if you can at least inject some JavaScript, you can do a lot.
- Put back in the MARC record for enrichment purposes.

GOING FROM VIAF TO KOHA

A little bit of JavaScript goes a long way...

```
$(document).ready(function(){
if (document.location.href.indexOf('_NAME') == document.location.href.length-5 || // match /cgi-bin/koha/authorities/aut
document.location.href.indexOf('authorities/authorities.pl?authid='>-1) {
if (document.location.href.indexOf('authorities/authorities.pl?authid='>-1) {
var title = $('h1').text();
if (title.indexOf('Name')!=title.length-4) return; // not available for non-NAME headings
}
function myUpperCase(s) {
if (s.indexOf('nli')>-1) return 'NLI'+s.substr(3);
return s.toUpperCase();
}
function autoComplete() {
$('head').append('<link rel="stylesheet" href="http://ajax.googleapis.com/ajax/libs/jqueryui/1.8.19/themes/base/

// Koha 3.12.x
$('#authoritytabs').prepend('\
<div class="ui-widget">\
<label for="viaf">VIAF Autosuggest:</label> <input type="text" style="font-size:0.9em" id="viaf" size="8
</div>\
');
```

FOR MORE DETAILS

- The full code is available on the [Koha wiki](#)
- Credit goes to Stefano Bargioni of the Pontificia Università della Santa Croce

UPSHOT

- With VIAF autosuggest, the only excuse for not authorizing name heading — or we should say, linking to names qua Linked Data, is time.

HYBRIDS CAN EXIST NOW

No need to wait on...

- BIBFRAME
- A whole new ILS
- Everybody else to adopt LD techniques first

HYBRIDS CAN EXIST NOW

What you do need...

- Willingness to iterate
- Authority control -- but Linked Data gives you more ammunition for the business case to do it in the first place.

THE HYBRID CAN PUBLISH LINKED AND STRUCTURED DATA AS WELL

- Recall Dan's presentation
- MARC to RDF conversions are much more meaningful if you take advantage of authority control.

TOOLS

- Catmandu::RDF and Catmandu::MARC
- OpenRefine
- Linked Data support in Koha RFC
- Forthcoming - record enhancement

THANKS!

GALEN CHARLTON