

Ready...Set...URIs...Actionable!

Dan Tam Do
Metadata Services Manager
George Washington University
Libraries

Jackie Shieh
Coordinator, Resource Description
George Washington University
Libraries

Identifiers

Why *identifiers*?

First of Tim Berners-Lee's four principles of linked data:

“Use URIs as names for Things.”

<http://www.w3.org/DesignIssues/LinkedData.html>

Resource

The **THING** that has an identity

Uniformity

Why *uniform* identifiers?

Uniformity facilitates semantic interpretation

Why IRIs?

URIs: US-ASCII, subset of IRI

IRIs: fully Unicode-enabled

ASCII (URI) vs. UTF-8 (IRI)

	<i>encoding</i>		<i>utf-8 or %HH</i>	<i>us-ascii or %HH</i>
original	<====>	bytes	IRI	URI
March	utf-8	4D 61 72 63 68	March	March
März	utf-8	4D C3 A4 72 7A	März	M%C3%A4rz
März	iso-8859-1	4D E4 72 7A	M%E4rz	M%E4rz
März	macintosh	4D 8A 72 7A	M%8Arz	M%8Arz

URI: <http://xn--rksmrgs-5wao1o.josefsson.org>, <http://xn--99zt52a.w3.mag.keio.ac.jp>

IRI: <http://räksmörgås.josefsson.org>, <http://納豆.w3.mag.keio.ac.jp>

Actionability

Why *actionable* URIs/IRIs?

Second of Tim Berners-Lee's four principles of linked data:

“Use HTTP URIs so that people can look up those names.”

Actionability

Use HTTP scheme to query and retrieve:

```
"http:" "//" host [ ":" port ] [ abs_path [ "?" query ]]
```

Actionability

Not actionable:

(OCoLC)fst00903451

Actionable:

<http://id.worldcat.org/fast/903451>

Current environment

Where does linked data fit in?

First two principles of linked data already within reach

A mighty subfield

\$0


- Authority record control number or standard number
- Available for X00, X10, X11, X30, 240, most 65X, others
- Repeatable


Processing

Batch processing using MarcEdit

Build Linked Records

The Linked Data Tool has been designed to loop through a set of MARC records and resolve access points in the 1xx, 6xx, and 7xx data elements to their linked data end-points.

Source File: 

Save File: 

ID Services

- ☒ U.S. Library of Congress (id.loc.gov)
- ☐ AutoDetect Subject ID
- ☐ OCLC VIAF

Process

Cancel

Processing

Batch processing with Perl scripting

```
## FAST fst 2nd indicator value 7
for my $f6XX ( $record->field('65.') ) {
    my $sub2 = $f6XX->subfield('2') if ($f6XX->subfield('2'));

    if ( $sub2 ) {
        if ( $sub2 =~ /fast/ ) {
            my $sub0 = $f6XX->subfield('0') if ($f6XX->subfield('0'));
            if ($sub0 =~ /fst/ ) {
                my ($rem, $ID) = split /fst/, $sub0;

                my $newSUB = qq{http://id.worldcat.org/fast/$ID};
                $f6XX->update( 0 => $newSUB);}
            }
        }
    }
}
```

Processing

Single record updating using URIs provided by
id.loc.gov, FAST, VIAF, other services

Costs and benefits

Costs and drawbacks

- time spent learning, training
- time spent performing additional step, quality control, supplemental editing
- lack of immediate visible results

Costs and benefits

Benefits

- improved knowledge of MARC
- acquaintance with id.loc.gov, FAST, VIAF
- improved MarcEdit skills
- “things” vs. “strings”