

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Prepare to Be Linked

*Enhancing MARC Data with URI
on a Shoestring*

ALCTS CaMMS
Catalog
Management
Interest Group

January 21, 2017

Jackie Shieh
Coordinator Resource Description
jshieh [AT] gwu.edu



Prepare to be Linked: Is It for Mine?

You heard about linked data. You read about inserting HTTP URI in \$0.

You were asked to investigate and propose plans to transition your local data to be linked data.

- What does this mean?
- Where do I go from here?
- Will it make a difference?



A Use Case: GW Libraries

- Ca 1. M Bib
- Split into 390 Files
- 2,500 Bib Records per File
- File Size: 4.5mb-8.8mb
- Runtime per file ranging from 30 mins to 1 hour
- Timeframe: May-August 2015
- Three Staff Members
- URI Added in \$0 (Bib): **3,977,819**





**Before anything else, preparation
is the key to success.**

Alexander Graham Bell

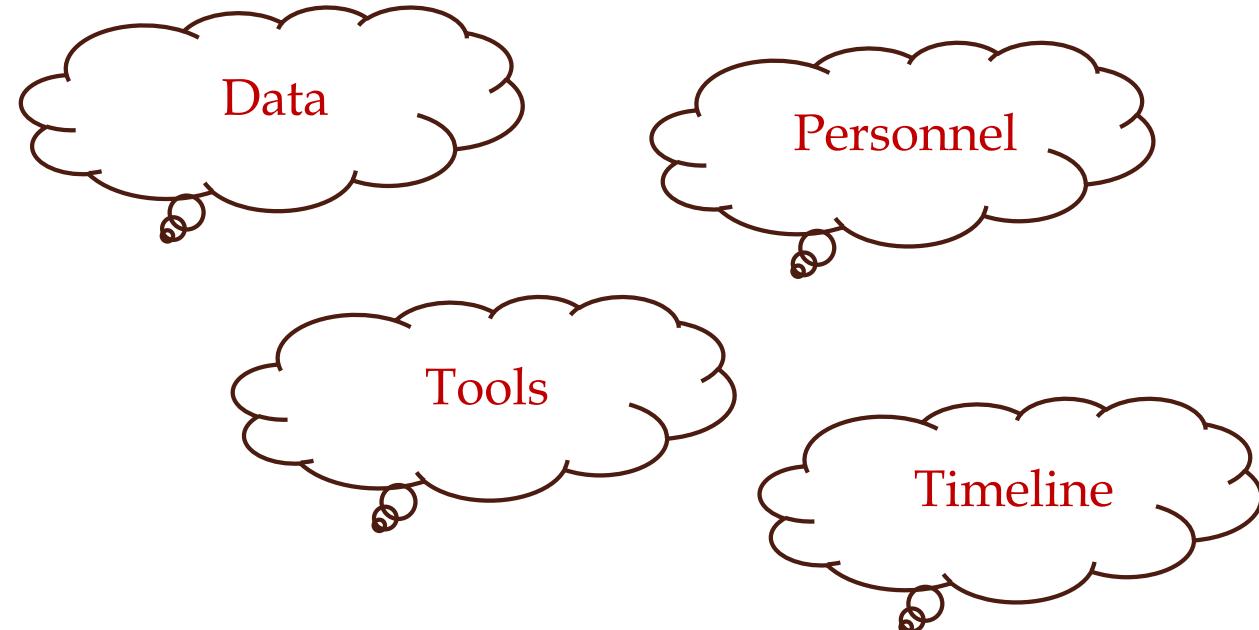


THE GEORGE
WASHINGTON
UNIVERSITY

WASHINGTON, DC

Decision Points

- Why
- What
- Where
- When
- Who
- How



Why and What

Linked data describes **a method of publishing structured data so that it can be interlinked and become more useful.**

The Semantic Web isn't just about putting data on the web. It is about **making links**, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, **related, data.**

- A roadmap to the Semantic Web by Tim Berners-Lee



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

RDF Triples (Statements): IRI/URI

Subject is usually a
IRI

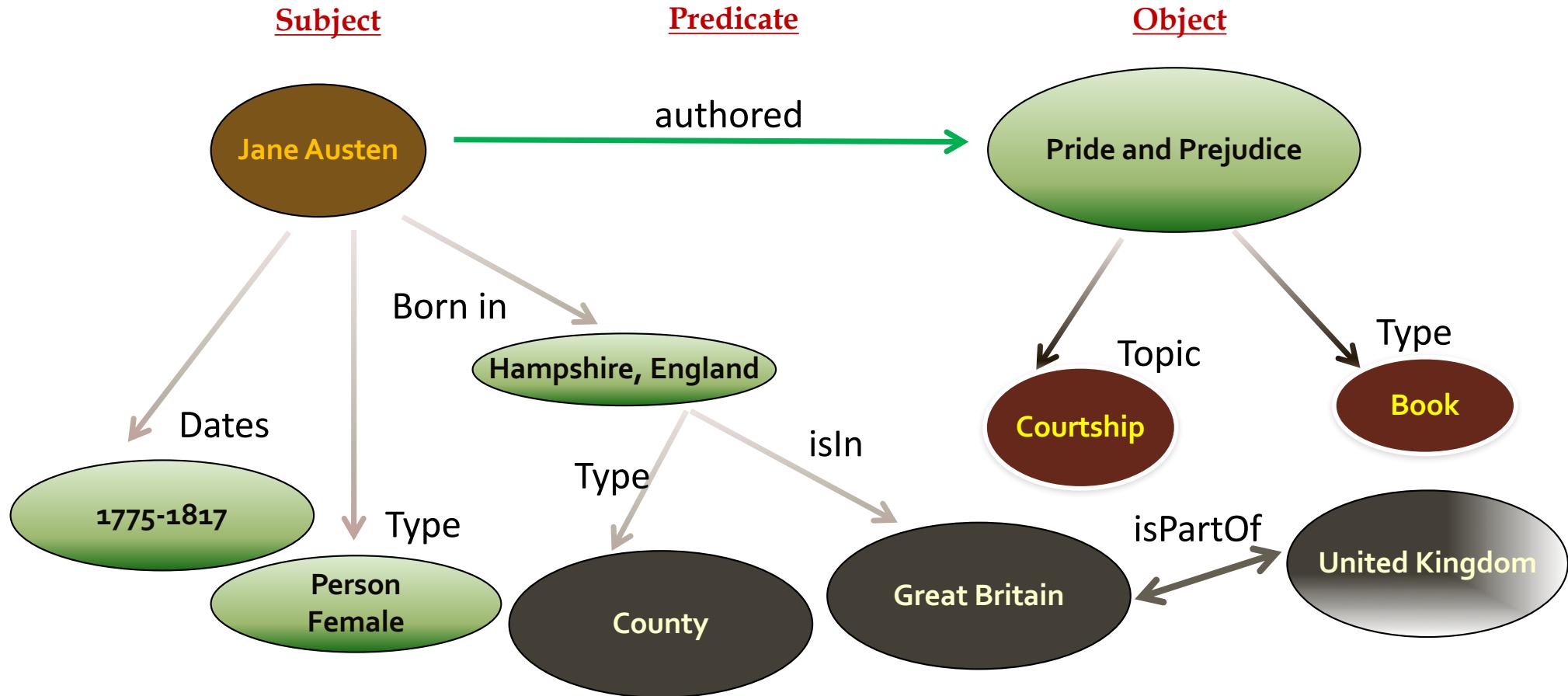
Object can be a IRI

Predicate is always a
IRI

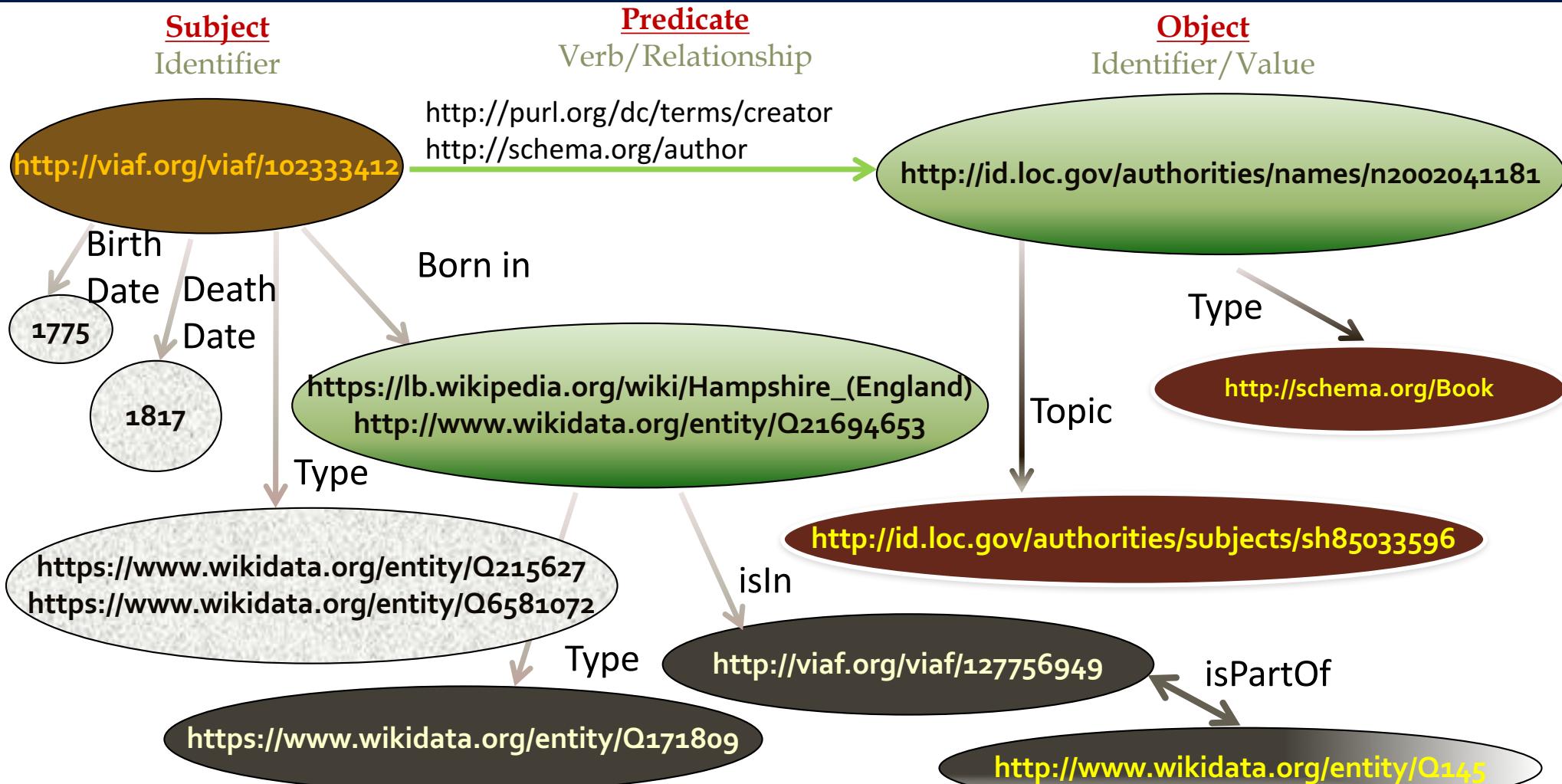
Object can also be a
literal
(string)



Entitification of Resource Data



HTTP IRIs in Entity



MARC Bibliographic Data

Where Do IRIs for RDF Objects Go?

- \$u URL → Uniform Resource Identifier
- \$0 Authority record control number or standard number



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

\$0 in Connexion

100 1#\$aBach, Johann Sebastian.\$4aut\$0(DE-101c)310008891

100 1#\$aBach, Johann Sebastian.\$4aut\$0 <http://d-nb.info/gnd/310008891>

100	1	‡a Bach, Johann Sebastian, ‡d 1685-1750. ‡0 http://id.loc.gov/authorities/names/n79021425
-----	---	---



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Expected Results: RDF URI in \$0

001 3792771
005 20130822135633.0
008 910418s1991 nyu 000 1 eng
010 __ |a 91052940
035 __ |a (OCOLOC)ocm15525100
040 __ |a DLC |c DLC |d DGW
019 __ |a 29774405
020 __ |a 0679405429
050 00 |a PR4034 |b .P7 1991
082 00 |a 823/.7 |2 20
049 __ |a DGWW
100 1__ |a Austen, Jane, |d 1775-1817. |0 http://id.loc.gov/authorities/names/n79032879 |0 (DLC)n79032879 |0 http://viaf.org/viaf/102333412 |0 (viaf)102333412
245 10 |a Pride and prejudice / |c Jane Austen ; with an introduction by Peter Conrad.
260 __ |a New York : |b Knopf : |b Distributed by Random House, |c c1991.
300 __ |a xxix, 368 p. ; |c 22 cm.
490 0__ |a Everyman's library ; |v 1
651 _0 |a England |v Fiction.
650 _0 |a Young women |z England |v Fiction. |0 http://id.loc.gov/authorities/subjects/sh2008124372 |0 (DLC)sh2008124372
650 _0 |a Courtship |v Fiction. |0 http://id.loc.gov/authorities/subjects/sh2008100298
650 _0 |a Sisters |v Fiction. |0 http://id.loc.gov/authorities/subjects/sh2008111400
650 _0 |a Domestic fiction, English. |0 http://id.loc.gov/authorities/subjects/sh86004355
655 _7 |a Love stories. |2 rccfd



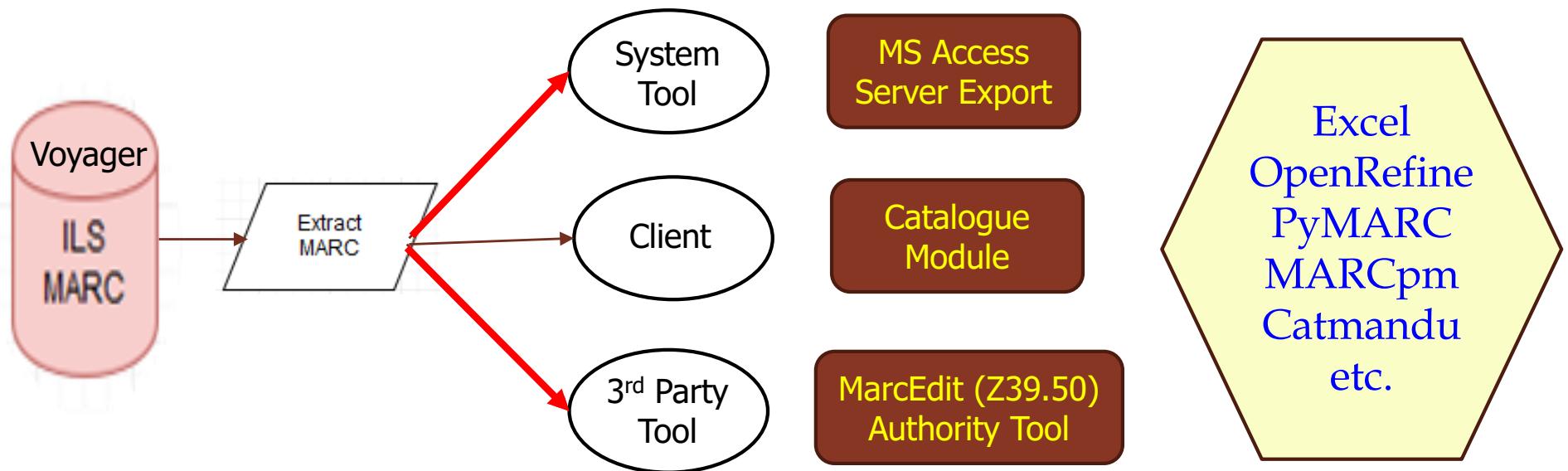
Let's Get to Work

- Team Formation
- Tasks Realignments
- Snapshot MARC Data
- System Scalability (Files Size)
- Extraction Paths
- Trustworthy Tools for RDF URI (URL) ~~(URL)~~
- Data Analytics
- Workflow Design (including Trial Runs)



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

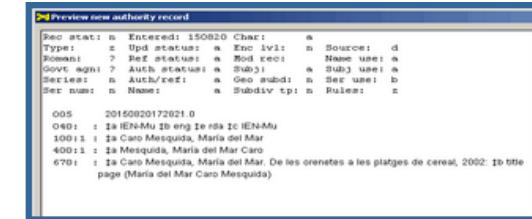
Data Extraction Paths



Tools

MarcEdit -- Your complete *free* MARC editing utility

<http://marcedit.reeset.net/downloads>



**Authority toolkit:
create and modify
authority records**

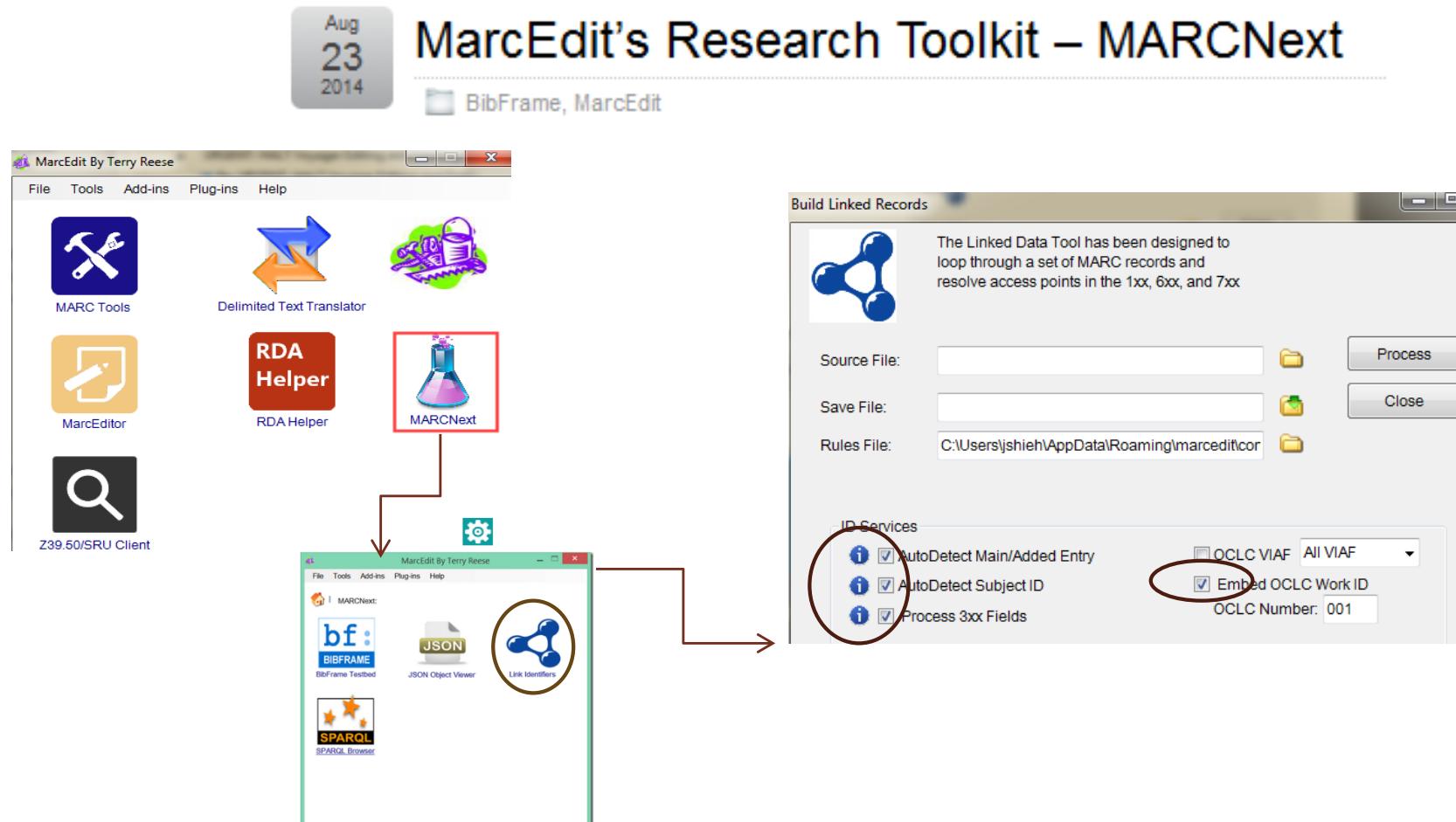
Northwestern University, Evanston, IL USA. December 31, 2016

<http://files.library.northwestern.edu/public/oclc/documentation/>



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

MarcEdit



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Where Does MarcNext Put URI/IRI?

100 field_\$0	110 field_\$0	111 field_\$0	130 field_\$0	240 field_\$0				
700 field_\$0	710 field_\$0	711 field_\$0	730 field_\$0	830 field_\$0				
600 field_\$0	610 field_\$0	611 field_\$0	630 field_\$0	650 field_\$0	651 field_\$0	655 field_\$0		



MarcNext Converts \$0 Text String to IRI

=650 \7\$aMedical policy.\$2fast\$0(OCoLC)fst01014505



=650 \7\$aMedical policy.\$2fast
\$0<http://id.worldcat.org/fast/1014505>



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Workflow Changes

- URI Embedding (routine)
- Quality Control (easier)
- Quality Assurance (better)
- Ongoing Maintenance (unknown)
- Evidence-based Metadata Service (hopeful)



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Commercial Alternative

- URI in Bib and Authority
- Reconciliation
- Maintenance
- Canonical Community Vocabularies and Ontologies
- Multiple \$0s



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Can You Do This?

Undoubtedly so!

T~~Q~~uesti~~o~~n?



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

THE GEORGE
WASHINGTON
UNIVERSITY

WASHINGTON, DC