

FAST HEADINGS PROJECT

at the Smithsonian Libraries

Heidy Berthoud
Head, Resource Description
BerthoudH[at]si.edu



Jackie Shieh
Descriptive Data Management
ShiehJ[at]si.edu

Smithsonian
Libraries

Outline

- Rationale
- Data selection criteria
- Tools
- Bib data enhancements
- Authority data enhancements
- Staff impacts

Rationale of applying FAST

- Compatibility with existing metadata
- Ease of assignment / more flexible
- Retrieval effectiveness of controlled vocabulary
- Cost of maintenance
- Semantic interoperability
- MARC data exposure*

Data selection

- Bibliographic records in Horizon
- Monographs (leader 06-07 = “am”)
- Published prior to 1923 (008/07-10 = “<1923”)

- Windows 10
 - Cataloging
 - Connexion
 - Horizon
 - Microsoft Office 365
(Excel, Powerpoint screen recording)
 - cURL (client URL)
- Open Source
 - MarcEdit
 - TextPad
 - OpenRefine
 - Draw.io

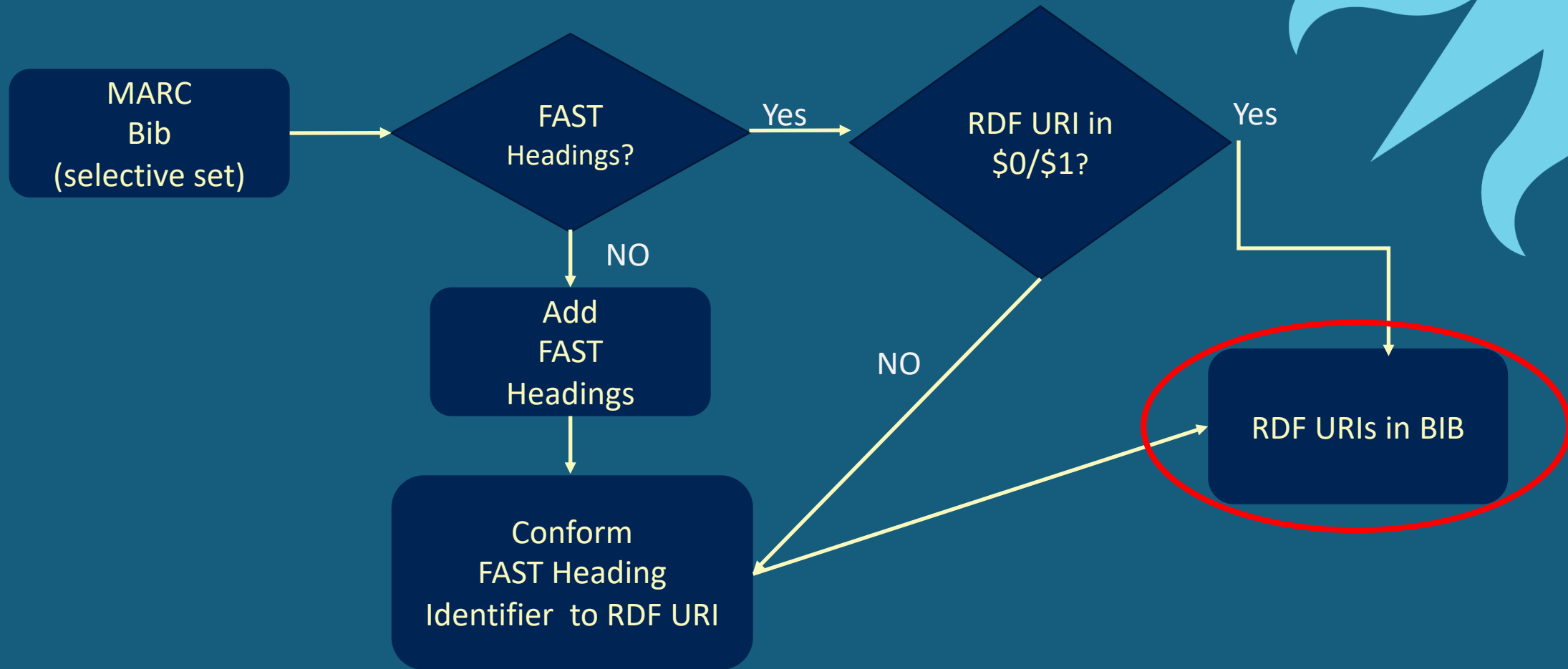
For Want of a Nail

- Linux
 - MARC::Record Perl package
 - Yaz toolkit (Index Data)
 - Perl and shell scripts
 - cURL (client URL)

Bib data enhancements

- Added FAST identifiers (then converted to URIs)
- Updated MARC encoding standards and data validation
- RDA content standards
- Added URIs to AAPs and alternates (1xx, 6xx, 7xx)
- Replaced/added relationship URI (\$4) to relator code
- Removed ISBD punctuation (trailing and between subfields)

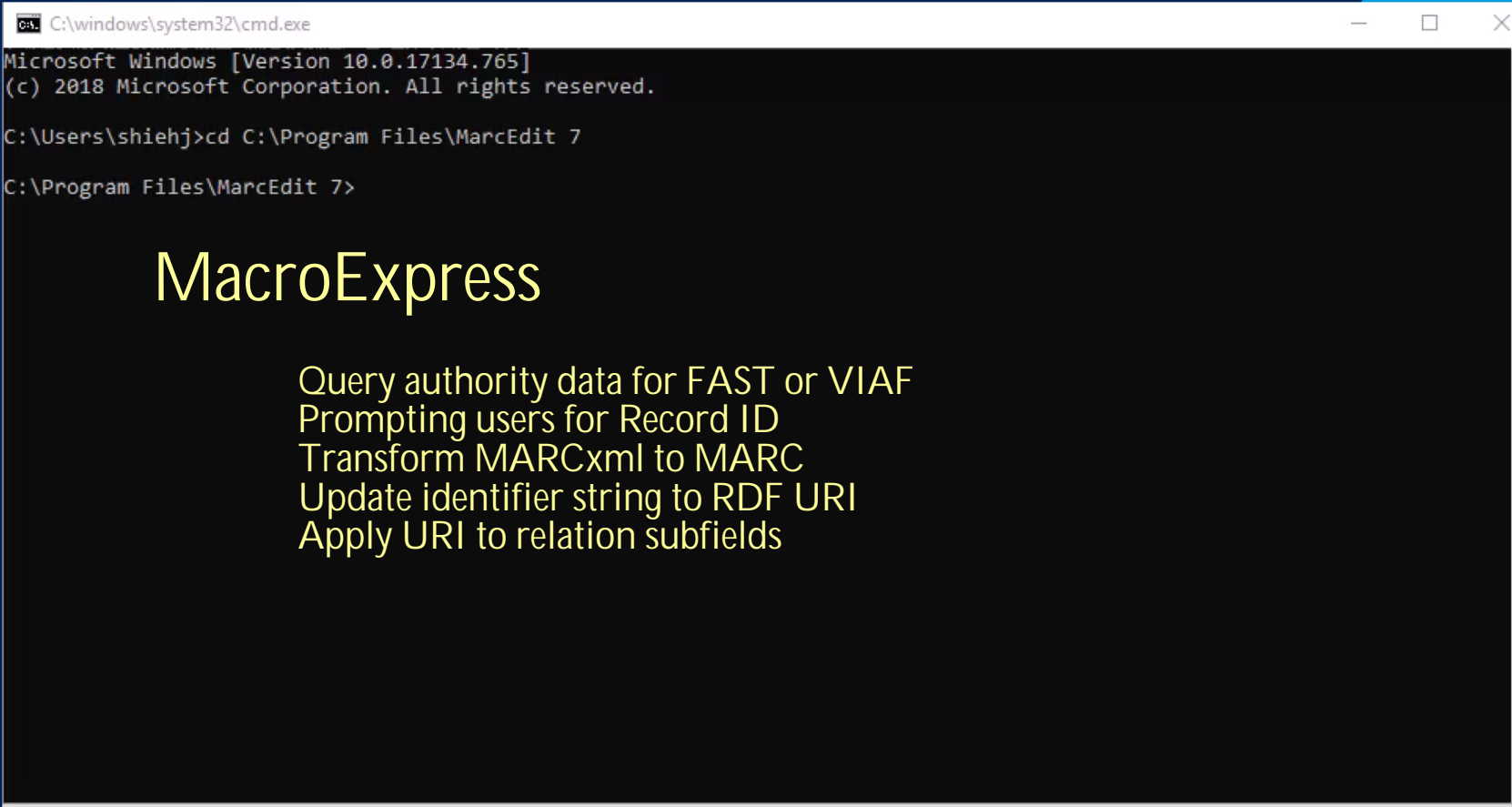
Enhancement process



Authority data enhancements : challenges

- System challenges in loading multiple authority files
- Local practice: authority control at point of cataloging
 - Single item authority workflow
 - Batch process authority workflow
- Post-load reconciliation process

Authority query by identifier



```
C:\windows\system32\cmd.exe
Microsoft Windows [Version 10.0.17134.765]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\shiehj>cd C:\Program Files\MarcEdit 7
C:\Program Files\MarcEdit 7>

MacroExpress

Query authority data for FAST or VIAF
Prompting users for Record ID
Transform MARCxml to MARC
Update identifier string to RDF URI
Apply URI to relation subfields
```

Authority: LCSH & FAST

010	—	†a sh 85076047	
024	7	†a 00996258 2 fast	
040	—	†a DLC c DLC d DLC	
053	—	†a DS432.L4	
150	—	†a Lepcha (South Asian people)	
450	—	†w nne a <u>Lepchas</u>	
450	—	†a <u>Rong (South Asian people)</u>	
550	—	†w g †a <u>Ethnology</u> z <u>Bhutan</u>	
550	—	†w g †a <u>Ethnology</u> z <u>India</u>	
550	—	†w g †a <u>Ethnology</u> z <u>Nepal</u>	
670	—	†a Work cat.: Foning, A.R. Lepcha, my vanishing	
670	—	†a Sharma, J.S. Encyc. Indica, 1981.	
670	—	†a Gazetteer of India, 1965: b v. 1, p. 300.	
670	—	†a Risley, H.H. Tribes and castes of Bengal, 198	
670	—	†a Dalton, E.T. Tribal hist. of Eastern India, 1978	
670	—	†a Britannica, 1968: b v. 20, p. 507.	
670	—	†a Web. 3.	
997	—	†a CULTURETERM	

LDR 00000cz a2200037n 45 0		
001	fst00996258	Horizon import
003	OCOLC	moves 001+003
		to 035
005	20160727155209.0	
008	041024nn anznnbabn	ll ana d
016	7	\$a fst00996258 \$2 OCOLC
024	7	\$a http://id.worldcat.org/fast/996258 \$2 uri
040	—	\$a OCOLC \$b eng \$c OCOLC \$f fast
053	_ 0	\$a DS432.L4
150	—	\$a Lepcha (South Asian people)
450	—	\$a Lepchas
450	—	\$a Rong (South Asian people)
550	—	\$a Ethnology \$0 (OCOLC)fst00916106
551	—	\$a Bhutan \$0 (OCOLC)fst01207216
551	—	\$a India \$0 (OCOLC)fst01210276
551	—	\$a Nepal \$0 (OCOLC)fst01206102
688	—	\$a LC (2017) Subject Usage: 29
688	—	\$a WC (2017) Subject Usage: 120
750	_ 0	\$a Lepcha (South Asian people) \$0 (DLC)sh 85076047
750	_ 4	\$a Lepcha people \$4 RM \$0 http://en.wikipedia.org/wiki/Lepcha_people \$0 https://www.wikidata.org/wiki/Q972684 <input type="radio"/> Remove incorrect match
750	_ 7	\$a Lepcha (peuple d'Asie du Sud) \$2 ram \$4 EQ \$0 http://data.bnf.fr/ark:/12148/cb119411544

035 \$a ([OCOLC](#))fst0096258

Identity management : opportunities

- Semantic relations: e.g., “sameAs”, “relatedMaterials”, “closeMatch”
- Coordinates and GeoNames
- Creating new identities using Wikidata

Staff impacts : breaking old habits

- Cataloging silos
 - Multiple cataloging sites with diverging policies
 - Communication struggles
- Training inconsistencies
- Reinventing workflow
 - Development for *all* staff levels
 - Connecting new technologies to daily tasks

Staff impacts : learning new skills

- Regular Expression concept
- MarcEdit
 - MARCJoin
 - Validate MARC records
 - Find/replace
 - Manage/apply tasks
 - Export tab delimited records
 - Export/import for/from OpenRefine
 - Remove blank lines
 - RDA helper
 - Build linked records
 - Compile file into MARC
- OpenRefine
 - Clustering
 - Using “blank down” function to remove duplicates
- TextPad
 - Free evaluation copy
 - Has block select functionality
 - Sort and dedupe

Resources

- FAST Linked data API

<https://www.oclc.org/developer/develop/web-services/fast-api/linked-data.en.html>

- MarcEdit Command line

<https://marcedit.reeset.net/cmarcedit-exe-using-the-command-line>

- Index Data Yaz

- <https://www.indexdata.com/resources/software/yaz/>

- TextPad

- <https://www.textpad.com/>

- MetaCpan Perl package

- <https://metacpan.org/pod/MARC::Record>

- cURL://

- <https://curl.haxx.se/>

