Semi-Automated FAST Reconciliation for Hyrax Repositories

Tiffany Chan

Senior Developer/Analyst | University of Victoria Libraries 🇨🇦

14 June 2023 | Open Repositories Conference

Link to Slides: uviclibrary.github.io/OR2023Dev

Link to Slides: uviclibrary.github.io/OR2023Dev

Some Background

Migrated to Hyku (Hyrax-based) from ContentDM

Decided to use OCLC FAST (controlled vocabulary) for several metadata fields

Textual values converted to URIs whenever possible

More details on our metadata migration

Link to Slides: uviclibrary.github.io/OR2023Dev

Textual Values

Zenzile Miriam Makeba

Makeba, M. (Miriam)

Makeba, Miriam

Human-readable

Can be ambiguous, inconsistent

URI

http://id.worldcat.org/fast/201144

Label

Makeba, Miriam

Machine-readable URIs and
human-readable labels

Uniquely identifiable

How Hyrax handles URIs/Labels in Metadata

URIs and Labels

```
"subject_tesim":
["http://id.worldcat.org/fast/1050538",
  "http://id.worldcat.org/fast/78887",
  "http://id.worldcat.org/fast/2012667",
  "http://id.worldcat.org/fast/1050534"]

"subject_label_tesim":["Painters--
Correspondence",
  "Orpen, William, Sir, 1878-1931",
  "Glenavy, Beatrice Moss Campbell,
Baroness, 1883-1970",
  "Painters--Biography"
```

indexed on the backend

Labels

Creator
: Orpen, William, Sir, 1878-1931

Contributor
: Glenavy, Beatrice Moss Campbell, Baroness, 1883-1970

Subject
: Painters--Correspondence
  Orpen, William, Sir, 1878-1931
  Glenavy, Beatrice Moss Campbell, Baroness, 1883-1970
  Painters--Biography

displayed in the interface

FAST is constantly evolving...

New URIs are being created

New URIs need to be created

Many of our subjects/metadata are of especially local interest, so they don't already exist in FAST

We enter data as textual values first, then submit a request to OCLC to create/assign URIs

When FAST approves our request, we need to replace the textual values with the new URI(s)
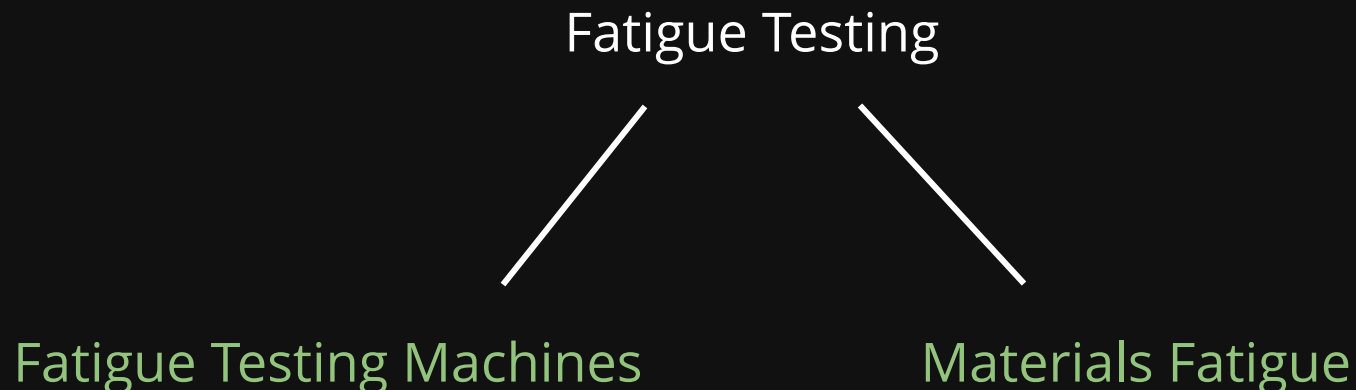
FAST is constantly evolving...

Existing URIs can still change

Examples of Changes

Dates of birth or death are added to the label

Pelé, 1940 - 2022

Headings can be deprecated or split

Fatigue Testing

Fatigue Testing Machines   Materials Fatigue

URI Maintenance in Hyrax

OCLC doesn't contact you when they add a URI that you requested

No easy way to update automatically

Hyrax cannot use the MARC file to update
There is no API (at least not yet)

Hard to search for the same value in multiple fields using the interface

FAST Changes

Log of updates to FAST Subject Headings

RECENT FAST CHANGES

Date	Notes	MARC Change File
2023-04-26	2439 Records	FAST Change File
2023-04-25	1648 Records	FAST Change File
2023-03-30	33 Records	FAST Change File
2023-03-22	15645 Records	FAST Change File
2023-02-17	1183 Records	FAST Change File
2023-02-09	125 Records	FAST Change File
2023-02-01	18 Records (form/genre)	FAST Change File
2023-02-01	588 Records	FAST Change File
2023-01-18	2802 Records	FAST Change File
2022-12-21	288 Records	FAST Change File
2022-12-02	4 Records (form/genre)	FAST Change File
2022-12-02	391 Records	FAST Change File
2022-11-22	2081 Records	FAST Change File
2022-11-21	455 Records	FAST Change File
2022-11-03	308 Records	FAST Change File
2022-11-03	13 Records (form/genre)	FAST Change File
2022-10-28	3813 Records	FAST Change File
2022-10-14	75 Records	FAST Change File
2022-10-11	4238 Records	FAST Change File
2022-09-30	5 Records (form/genre)	FAST Change File

FAST Updater

Download recent Microsoft Excel spreadsheet(s)
(You can specify a date range)

Parse the spreadsheets and categorize changes according to type (New heading, modified heading, obsolete/deprecated, split)

Periodic automatic updates via Sidekiq

Automatic Changes

New Headings

Search the repository for items with fields containing the textual value, then replace them with the URI

Modified Headings

Search the repository for the URI and reindex affected items, using Hyrax (and associated gems) to fetch and index the updated label

New Headings

Before

```
"subject_tesim":[
  "Spreitz, Karl, 1927-",
  "http://id.worldcat.org/fast/837358"
],

"subject_label_tesim":[
  "Spreitz, Karl, 1927-",
  "Boys"
],
```

After

```
"subject_tesim":[
  "http://id.worldcat.org/fast/2013151",
  "http://id.worldcat.org/fast/837358"
],

"subject_label_tesim":[
  "Spreitz, Karl, 1927-",
  "Boys"
],
```

Modified Headings

Before

```
"subject_tesim":[
  "http://id.worldcat.org/fast/2013151",
  "http://id.worldcat.org/fast/837358"
],

"subject_label_tesim":[
  "Spreitz, Karl, 1927-",
  "Boys"
],
```

After

```
"subject_tesim":[
  "http://id.worldcat.org/fast/2013151",
  "http://id.worldcat.org/fast/837358"
],

"subject_label_tesim":[
  "Spreitz, Karl, 1927-2016",
  "Boys"
],
```

Deprecated or Split Headings

Require manual intervention

Email a configured address describing changes and suggestions (if any)

Staff can use a custom interface to search for and replace the old URI with new one(s)

Dashboard

Registered Users	Total Visitors	Returning Visitors	New Visitors
Current	Past 30 days	Past 30 days	Past 30 days
92	n/a	n/a	n/a

Administrative Sets

Administrative Set	Works	Files
Default admin set		
Default Admin set 2		

User Activity

Repository Growth

Repository Objects

Published
12,067

Replace or Delete a URI

Previous Changes

Replace or Delete a URI

Search Results

Tiffany Chan

ACTIVITY

📈 Repository Activity   ›

📈 Your activity   ›

🗒 Profile

🔔 Notifications

↔ Transfers

👥 Manage Proxies

📊 Reports

REPOSITORY CONTENTS

📁 Collections

📄 Works

TASKS

☰ Batch Processing   ›

🔧 Replace or Delete FAST URIs

🚩 Review Transfers

👤 Manage Users

Replace or Delete a URI

Replace one FAST URI with another across the entire repository, or scoped to a collection.

URI to replace or delete:

○ Delete   ● Replace with:

● in Everything   ○ in a Collection:

[Search for an entity ▾]   [Search for an entity ▾]   ✖ Remove   [Search for a collection ▾]

OR

➕ Add a URI

[Paste a URI]

Search Preview   **Submit**

Search Results   **Previous Changes**

Previous Changes

Old Label / URI	New Labels / URIs	Action	Scope / Collection	Timestamp	Complete?
Tiffany family (http://id.worldcat.org/fast/371299)	**Tiffany family** (http://id.worldcat.org/fast/371299) **Tiffany and Company** (http://id.worldcat.org/fast/549011)	Replace	Test Garbage Collection	March 21 2023, 06:09 PM	**Success** 1 replacement(s) made.
Refuse and refuse disposal--Environmental aspects USE Refuse and refuse disposal (http://id.worldcat.org/fast/1092912)		Delete	Everything	March 15 2023, 02:42 PM	**Success**

What's Next?

Convenient way to add or remove URIs/textual values from the repository

Make it a gem, valkyrize it

FAST Changes API Survey

Thanks!

Tiffany Chan (tjychan@uvic.ca)

Code: github.com/UVicLibrary/fast_update

Link to slides: uviclibrary.github.io/OR2023Dev

Slides made with slides.com / reveal.js