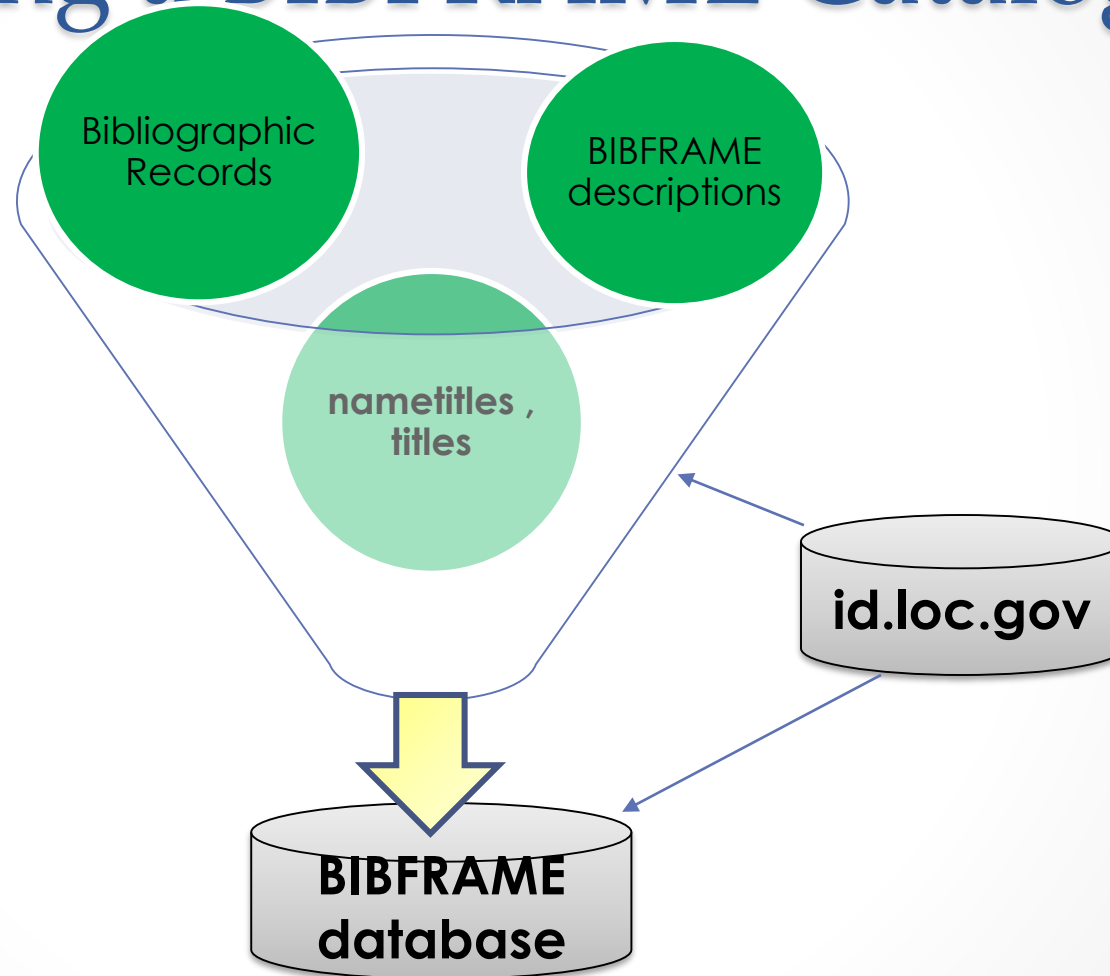
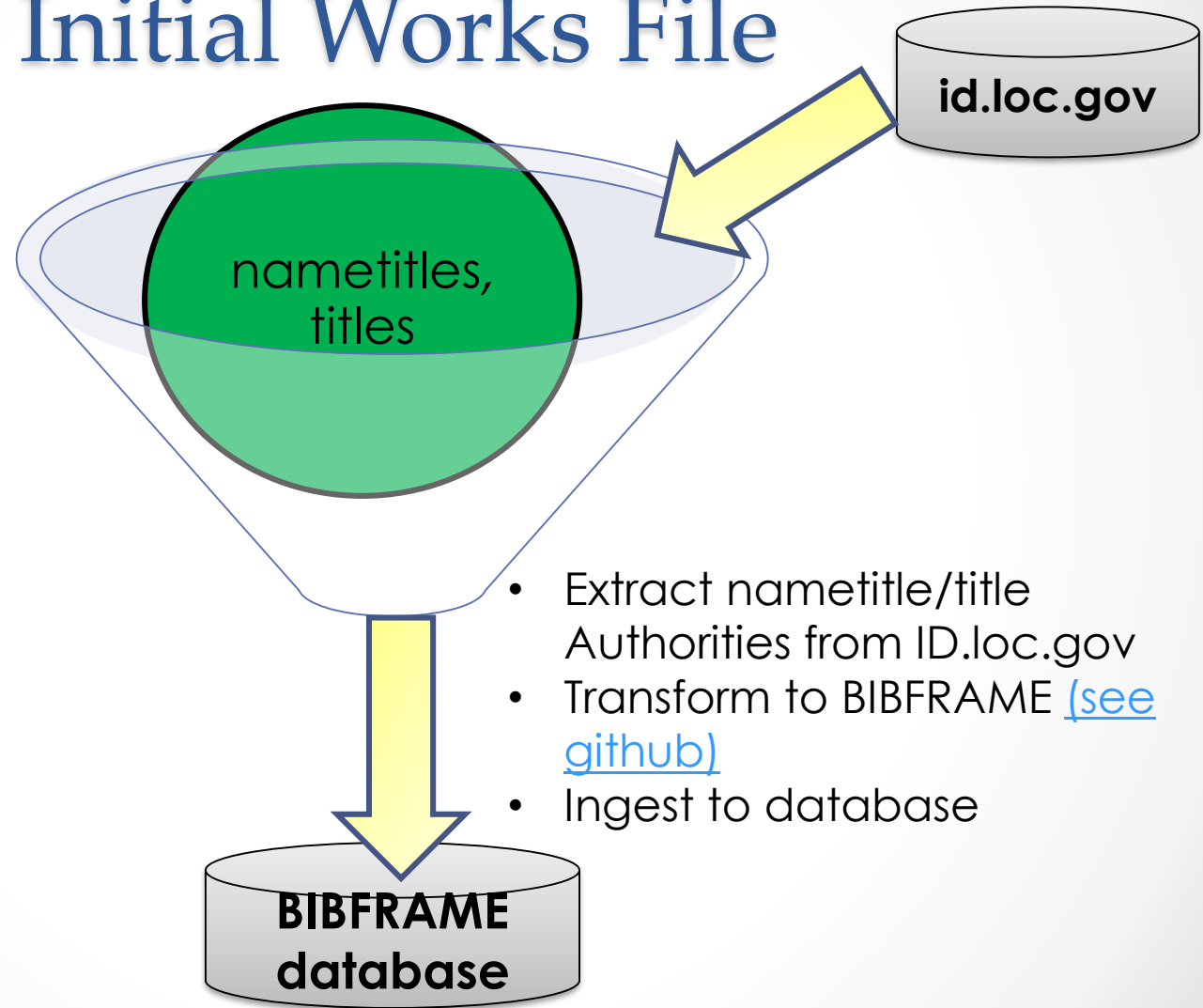


# Building a BIBFRAME Catalog



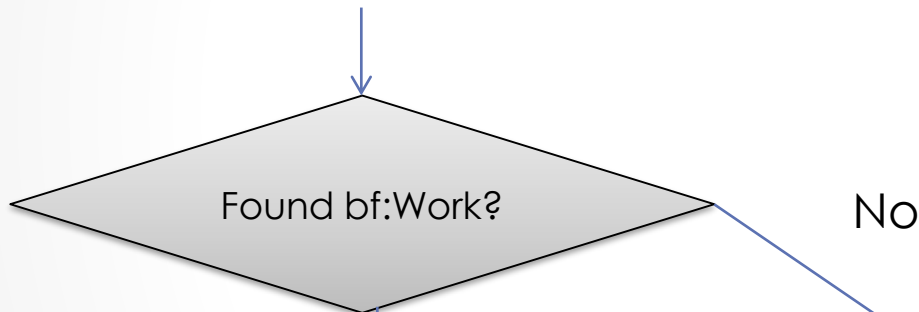
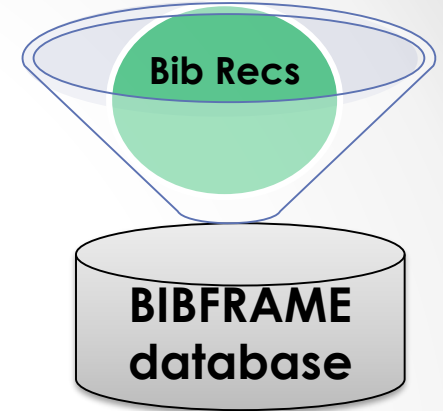
# Initial Works File



# Bibliographic Conversion



- MARC2bibframe2 transform ([see github](#))
- Match to existing bf:Works with same nametitle



Yes

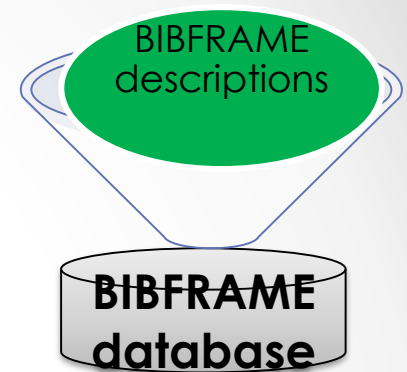
- Merge, Dedup Subjects, Classifications
- Store in Found Work
- Adjust uris to found Work,
- Store new Instances, Items

- Store as new bf:Work
- Store new Instances, Items



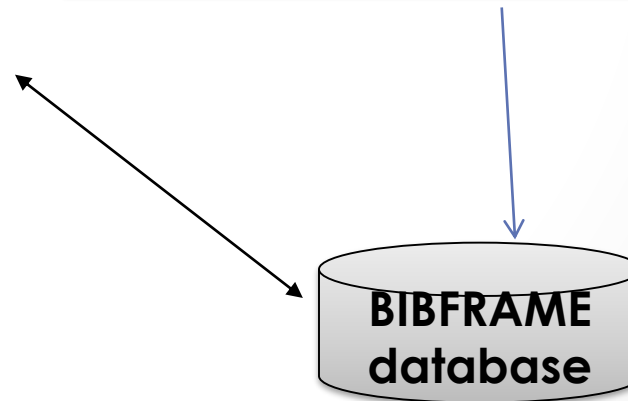
# BIBFRAME Descriptions

BFE BIBFRAME  
Editor



- Create Instance, Items(s)
- Look up a bf:Work in BIBFRAME database
- Ingest with link to the Work

- Create new bf:Work, Instances, Items
- Ingest (what is the uri?)



# Infrastructure

- MarkLogic NoSQL Server (3 node cluster) for ID
  - Storage, search/display, RDF triplestore
- MarkLogic 3 node cluster
  - for BIBFRAME and ID ingest, processing, testing
- Apache/Varnish Web Cache
  - (2 VMs for load balancing)
- Xquery, SPARQL code base for ingest, search/display
- Javascript codebase for BIBFRAME editor
- XSL for MARCXML, ONIX data transformations

# Infrastructure Updates

- Added new node to MarkLogic production cluster for ID
- Added 1 varnish web cache server
- Added 2 new nodes for BIBFRAME processing MarkLogic cluster
- Upgraded from MarkLogic version 5 to version 8
- MarkLogic Semantics replaces 4store triplestore
  - Document-based triples for ease of updates
- New BIBFRAME database added to id database
  - Still not public
- HTTPS support just added (not mandated)

# Software updates I

- New MARC Conversion in xsl instead of xquery
- Installation of conversion in Metaproxy, yaz
- New Authorities transform for nametitles
- Comparison program online to show MARC and BIBFRAME side by side in rdfxml and ttl serializations.
- Merge/ingest programs (nametitles and bibliographic records) updated for BIBFRAME2 vocabulary
- New search/display interface

# Software updates II

- Use SPARQL to show links to parent Work/Instance, sibling Instances, Item titles
- New templates for BIBFRAME2 vocabulary in Editor, new lookups for controlled vocabularies
- Editor now has lookups to BIBFRAME database for attaching Instances to Works
- Storing “published” BIBFRAME descriptions in database
- Daily nametitle and bib ingests from ILS to database to simulate the real catalog



# Some Numbers

## ID.loc.gov: 10.5M Names, Subjects, vocabularies

- 300M triples
- subjects: 21M
- predicates: 768
- objects: 25M

## BIBFRAME Database: 65M Works, Instances, Items

- 4 Billion Triples
- subjects : 500M
- predicates: 14,615
- objects: 800M

# Merge/Match Specs

- Based on 130/240 uniform titles indexed as “nametitle”
- New bf:Works stored with “nametitle” index and so become match point for future records
- For each new work from MARC, concatenate primary contributor and title (not from MARC 880)
  - <bflc:name00MatchKey>**Twain, Mark, 1835-1910.**</bflc:name00MatchKey>
  - <bflc:title00MatchKey>**Adventures of Huckleberry Finn**</bflc:title00MatchKey>
- (strip trailing slash)
- Match to existing database index entries.
- Suppressing “Untitled”, null etc., going forward

# Merge Stats

- 1.2M nametitles/titles as Works
- 17M Bibliographic descriptions
- 1.2M Works have merged instances
- 1.4M Instances merged altogether (onto nametitles/titles or other bibs)
- 530K Instances merged onto nametitle/title works
  - (still verifying these results)

# Merge Example I

LIBRARY OF CONGRESS    ASK A LIBRARIAN    DIGITAL COLLECTIONS    LIBRARY CATALOGS

The Library of Congress > BIBFRAME Database    **Detail**

Enter search word(s)    Everything    Search

[Print this item](#)

## Runaway mittens

Instance from Bib

<b>Instance</b>	<a href="resources/instances/c0014771920001">resources/instances/c0014771920001</a>
Title	Runaway mittens
LCCN	87012024
ISBN	0688070531
ISBN	068807054X (lib. bdg.)
Physical details	col. ill.
Creative responsibility statement	by Jean Rogers ; pictures by Rie Muñoz
Provider statement	New York : Greenwillow Books, c1988.
Mode of issuance	<a href="http://id.loc.gov/vocabulary/issuance/mono">http://id.loc.gov/vocabulary/issuance/mono</a>
<b>Publication</b>	<a href="http://id.loc.gov/ontologies/bibframe/Publication">http://id.loc.gov/ontologies/bibframe/Publication</a>
Date	(EDTF)1988
Place	<a href="http://id.loc.gov/vocabulary/countries/nvu">http://id.loc.gov/vocabulary/countries/nvu</a>
<b>Publication</b>	<a href="http://id.loc.gov/ontologies/bibframe/Publication">http://id.loc.gov/ontologies/bibframe/Publication</a>
Date	c1988
Place	New York
Agent	Greenwillow Books
Extent	[24] p.
Dimensions	26 cm.
Has Item	<a href="resources/items/c0014771920001">resources/items/c0014771920001</a>

Administrative metadata

### Instance Of

- > [Runaway mittens /\(c001477192\)](#)

### Sibling(s)

- > [Runaway mittens / c00147719200](#)
- > [Runaway mittens /\(c0118455440001\)](#)

### Has Items(s)

- > [c0014771920001](#)

### Other Views for This Description

- > [BIBFRAME RDF](#)
- > [BIBFRAME Turtle](#)
- > [BIBFRAME JSON](#)
- > [BIBFRAME JSON-LD](#)
- > [METS](#)

### Bookmark This Description

# Merge Example II

## Bach, Johann Sebastian, 1685-1750. Concertos. Selections

Work from Authority

<b>Work</b>	<a href="/resources/works/n81133661">/resources/works/n81133661</a>
<b>Title</b>	Concertos
<b><u>VariantTitle</u></b>	3 violin concerti
<b><u>VariantTitle</u></b>	6 Solo-Konzerte in Rekonstruktion
<b><u>VariantTitle</u></b>	9 concertos pour flûte, hautbois, violons, clavecins
<b><u>VariantTitle</u></b>	Cembalokonzerte
<b><u>VariantTitle</u></b>	Complete harpsichord

### Has Instance(s)

- › [Three concerti for three solo instruments and string orchestra\(c0055796930001\)](#)
- › [The concertos for 3 & 4 claviers\(c0055797000001\)](#)
- › [Concerto for two violins in D minor, BWV. 1043 ; Violin concerto in E, BWV. 1042 ; Concerto for violin & oboe in D minor, BWV. 1060\(c0055808710001\)](#)
- › [Concerto for violin and strings in A minor, BWV 1041 ; Concerto for violin and strings in E, BWV 1042 ; Concerto for 2 violins and strings in D minor, BWV 1043\(c0055810530001\)](#)
- › [9 concertos pour flûte, hautbois, violons,](#)

Title authority collocating mechanism, probably not a pure bf:Work. But results from cataloging decisions.

# SPARQL Use I

Display Instance parent, sibling title info using SPARQL

LIBRARY OF CONGRESS | ASK A LIBRARIAN | DIGITAL COLLECTIONS | LIBRARY CATALOGS

The Library of Congress > BIBFRAME Database | Detail

Enter search word(s) | Everything | Search

[Print this item](#)

**Runaway mittens**  
Instance from Bib

Instance	<a href="#">resources/instances/c0014771920001</a>
Title	Runaway mittens
LCCN	87012024
ISBN	0688070531
ISBN	068807054X (lib. bdg.)
Physical details	col. ill.
Creative responsibility statement	by Jean Rogers ; pictures by Rie Muñoz
Provider statement	New York : Greenwillow Books, c1988.
Mode of issuance	<a href="http://id.loc.gov/vocabulary/issuance/mono">http://id.loc.gov/vocabulary/issuance/mono</a>
<b>Publication</b>	<a href="http://id.loc.gov/ontologies/bibframe/Publication">http://id.loc.gov/ontologies/bibframe/Publication</a>
Date	(EDTF)1988
Place	<a href="http://id.loc.gov/vocabulary/countries/nyu">http://id.loc.gov/vocabulary/countries/nyu</a>
<b>Publication</b>	<a href="http://id.loc.gov/ontologies/bibframe/Publication">http://id.loc.gov/ontologies/bibframe/Publication</a>
Date	c1988
Place	New York
Agent	Greenwillow Books
Extent	[24] p.
Dimensions	26 cm.
Has Item	<a href="#">resources/items/c0014771920001</a>

Administrative metadata

Options

GO

[Search Tips](#)

**Instance Of**

- > [Runaway mittens /\(c001477192\)](#)

**Sibling(s)**

- > [Runaway mittens / c0014771920001](#)
- > [Runaway mittens /\(c0118455440001\)](#)

**Has Items(s)**

- > [c0014771920001](#)

**Other Views for This Description**

- > [BIBFRAME RDF](#)
- > [BIBFRAME Turtle](#)
- > [BIBFRAME JSON](#)
- > [BIBFRAME JSON-LD](#)
- > [METS](#)

**Bookmark This Description**

# SPARQL Use I

Display Instance parent, sibling title info using SPARQL

LIBRARY OF CONGRESS ASK A LIBRARIAN DIGITAL COLLECTIONS LIBRARY CATALOGS

The Library of Congress > BIBFRAME Database Detail

Enter search word(s) Everything Search

[Print this item](#)

**Runaway mittens**  
Instance from Bib

Instance	<a href="#">resources/instances/c0014771920001</a>
Title	Runaway mittens
LCCN	87012024
ISBN	0688070531
ISBN	068807054X (lib. bdg.)
Physical details	col. ill.
Creative responsibility statement	by Jean Rogers ; pictures by Rie Muñoz
Provider statement	New York : Greenwillow Books, c1988.
Mode of issuance	<a href="http://id.loc.gov/vocabulary/issuance/meng">http://id.loc.gov/vocabulary/issuance/meng</a>
<b>Publication</b>	<a href="http://id.loc.gov/ontologies/bibframe/Publication">http://id.loc.gov/ontologies/bibframe/Publication</a>
Date	(EDTF)1988
Place	<a href="http://id.loc.gov/vocabulary/countries/nyu">http://id.loc.gov/vocabulary/countries/nyu</a>
<b>Publication</b>	<a href="http://id.loc.gov/ontologies/bibframe/Publication">http://id.loc.gov/ontologies/bibframe/Publication</a>
Date	c1988
Place	New York
Agent	Greenwillow Books
Extent	[24] p.
Dimensions	26 cm.
Has Item	<a href="#">resources/items/c0014771920001</a>

Administrative metadata

Options

GO

[Search Tips](#)

1 Instance Of

- > [Runaway mittens / \(c001477192\)](#)
- > [Runaway mittens / c0014771920001](#)
- > [Runaway mittens / \(c0118455440001\)](#)

Has Items(s)

- > [c0014771920001](#)

Other Views for This Description

- > [BIBFRAME RDF](#)
- > [BIBFRAME Turtle](#)
- > [BIBFRAME JSON](#)
- > [BIBFRAME JSON-LD](#)
- > [METS](#)

Bookmark This Description

# SPARQL Use I

Display Instance parent, sibling title info using SPARQL

The screenshot shows the Library of Congress BIBFRAME Database interface. At the top, there are navigation links: "LIBRARY OF CONGRESS", "ASK A LIBRARIAN", "DIGITAL COLLECTIONS", and "LIBRARY CATALOGS". Below this is a search bar with the text "The Library of Congress > BIBFRAME Database" and a "Detail" link. The search bar contains "Enter search word(s)", a dropdown menu set to "Everything", and a "Search" button. A "Print this item" link is visible on the right. The main content area is titled "Runaway mittens" and "Instance from Bib". It displays a list of metadata fields:

- Instance** [resources/instances/c0014771920001](#)
  - Title Runaway mittens
  - LCCN 87012024
  - ISBN 0688070531
  - ISBN 068807054X (lib. bdg.)
- Physical details** col. ill.
- Creative responsibility statement** by Jean Rogers ; pictures by Rie Muñoz
- Provider statement** New York : Greenwillow Books, c1988.
- Mode of issuance** <http://id.loc.gov/vocabulary/issuance/mone>
- Publication** <http://id.loc.gov/ontologies/bibframe/Publication>
  - Date (EDTF) 1988
  - Place <http://id.loc.gov/vocabulary/countries/nyu>
- Publication** <http://id.loc.gov/ontologies/bibframe/Publication>
  - Date c1988
  - Place New York
  - Agent Greenwillow Books
- Extent** [24] p.
- Dimensions** 26 cm.
- Has Item** [resources/items/c0014771920001](#)

Administrative metadata

This screenshot shows a search options bar with a text input field, a "GO" button, and a "Search Tips" link.

The screenshot shows a SPARQL result visualization for the query. It displays a list of results under the heading "Instance Of". The results are:

- > [Runaway mittens /\(c001477192\)](#)
- > [Runaway mittens / c0014771920001](#)
- > [Runaway mittens /\(c0118455440001\)](#)

Below the results, there is a section titled "Other Views for This Description" with links to:

- > [BIBFRAME RDF](#)
- > [BIBFRAME Turtle](#)
- > [BIBFRAME JSON](#)
- > [BIBFRAME JSON-LD](#)
- > [METS](#)

At the bottom, there is a "Bookmark This Description" link.



# SPARQL Use II

Display Item title, parent info from other docs using SPARQL

```
- <rdf:RDF>
  - <bf:Item rdf:about="                /resources/items/c0014771920001">
    - <bf:shelfMark>
      - <bf:ShelfMarkLcc>
        <rdfs:label>PZ7.R6355 Ru 1988</rdfs:label>
      - <bf:source>
        <bf:Source rdf:about="http://id.loc.gov/vocabulary/organizations/dlc"/>
        </bf:source>
      </bf:ShelfMarkLcc>
    </bf:shelfMark>
    <bf:derivedFrom rdf:resource="                /resources/bibs/1477192.marxml.xml"/>
    <bf:itemOf rdf:resource="                /resources/instances/c0014771920001"/>
  </bf:Item>
</rdf:RDF>
```

# SPARQL Use II

Display Item title, parent info from other docs using SPARQL

LIBRARY OF CONGRESS    ASK A LIBRARIAN    DIGITAL COLLECTIONS    LIBRARY CATALOGS

The Library of Congress > BIBFRAME Database    **Detail**

Enter search word(s)    Everything    Search

[Print this item](#)

**Runaway mittens**  
Item

<b>Item</b>	<a href="/resources/items/c0014771920001">/resources/items/c0014771920001</a>
<b>LCC call number</b>	PZ7.R6355 Ru 1988
<b>Source</b>	<a href="#">DLC</a>
<b>Source metadata</b>	<a href="/resources/bibs/1477192.marcxml.xml">/resources/bibs/1477192.marcxml.xml</a>

## Item Of

› [Runaway mittens / \(c0014771920001\)](#)

## Sibling(s)

› **c0014771920001**

## Other Views for This Description

› [BIBFRAME RDF](#)

› [BIBFRAME Turtle](#)

› [BIBFRAME JSON](#)

› [BIBFRAME JSON-LD](#)

› [METS](#)

## Bookmark This Description

# SPARQL Use II

Display Item title, parent info from other docs using SPARQL

The screenshot shows the Library of Congress BIBFRAME Database interface. At the top, there are navigation links: LIBRARY OF CONGRESS, ASK A LIBRARIAN, DIGITAL COLLECTIONS, and LIBRARY CATALOGS. Below this is a breadcrumb trail: The Library of Congress > BIBFRAME Database. A search bar is present with the text 'Enter search word(s)', a dropdown menu set to 'Everything', and a 'Search' button. The main content area is titled 'Runaway mittens' and 'Item'. It contains a table with the following data:

<b>Item</b>	<a href="/resources/items/c0014771920001">/resources/items/c0014771920001</a>
<b>LCC call number</b>	PZ7.R6355 Ru 1988
<b>Source</b>	<a href="#">DLC</a>
<b>Source metadata</b>	<a href="/resources/bibs/1477192.marcxml.xml">/resources/bibs/1477192.marcxml.xml</a>

On the right side, there is a sidebar with the heading 'Item Of' and a red circle around the link '> [Runaway mittens / \(c0014771920001\)](#)'. Below this, there is a section 'Other Views for This Description' with links for '> [BIBFRAME RDF](#)', '> [BIBFRAME Turtle](#)', '> [BIBFRAME JSON](#)', '> [BIBFRAME JSON-LD](#)', and '> [METS](#)'. At the bottom of the sidebar is a section 'Bookmark This Description'.

# SPARQL Use II

Display Item title, parent info from other docs using SPARQL

The screenshot shows the Library of Congress BIBFRAME Database interface. At the top, there are navigation links: LIBRARY OF CONGRESS, ASK A LIBRARIAN, DIGITAL COLLECTIONS, and LIBRARY CATALOGS. Below this is a breadcrumb trail: The Library of Congress > BIBFRAME Database > Detail. A search bar is present with the text 'Enter search word(s)', a dropdown menu set to 'Everything', and a 'Search' button. The main content area displays the item title 'Runaway mittens' in a large blue font, which is circled in red. Below the title, there is a table with the following rows:

Item	
Item	<a href="/resources/items/c0014771920001">/resources/items/c0014771920001</a>
LCC call number	PZ7.R6355 Ru 1988
Source	<a href="#">DLC</a>
Source metadata	<a href="/resources/bibs/1477192.marcxml.xml">/resources/bibs/1477192.marcxml.xml</a>

To the right of the main content, there is a sidebar with the heading 'Item Of' and a link to 'Runaway mittens /(c0014771920001)', which is also circled in red. Below this, there is a section 'Other Views for This Description' with links for BIBFRAME RDF, BIBFRAME Turtle, BIBFRAME JSON, BIBFRAME JSON-LD, and METS. At the bottom of the sidebar, there is a section 'Bookmark This Description'.

# Issues Already Encountered

- Serializations are an ongoing issue:
  - `<rdf:Description><rdf:type rdf:resource="bf:Work"/></rdf:Description> == <bf:Work/>`
- Huge number of triples: how to limit, dedup on the way in, cache labels, etc.
- Merge: MARC 130s are problematic for title authorities; too many "Untitled" etc.
  - eg., photographs
- Merge: Record load sequence affects matching on initial build and reload. (Daily records okay)
- BIBFRAME conversion spec changes affect existing descriptions: need update mechanisms that don't affect merges
- Plenty of interesting examples of merging, conversion, or inadequate data in so many descriptions from varying cataloging rules over the years.

# Still to come I

- Open BIBFRAME data to public in some form
  - Bulk download? Searchable interface?
- Analyze data structures for Editor, vocabulary, conversion specs. improvements
- Loading BIBFRAME from ILS or elsewhere into Editor
  - eg., “copy cataloging”
- Ingest CIP and ONIX records
- Implement offset and limit in SPARQL queries

# Still to come II

- More SPARQL queries for related works, translations
- Link MARC 7xx related works to existing descriptions.
- More flexible Editor
- New RDF display interface: pure SPARQL display?
- Nametitle authority Works: link translations on ingest
- Services at ID to support external users: picklists etc.

# Useful Links

Compare side-by-side MARC/BIBFRAME

bib: <http://id.loc.gov/tools/bibframe/compare-id/full-ttl?find=5226>

authority: [Work conversion](#)

SRU BIBFRAME in Metaproxy

- BY Voyager bib id: (rec.id) [Metaproxy for Snoopy on Wheels](#)
- Add some Entity resolution : ["bibframe2a" recordSchema](#)
- by LCCN: (bath.lccn) [Lookup using LCCN](#)

ID label lookup for any authority/vocabulary

- <http://id.loc.gov/authorities/names/label/Twain,%20Mark,%201835-1910.%20Adventures%20of%20Huckleberry%20Finn>

Find docs by rdf:type in ID: <http://id.loc.gov/search/?q=rdf:type:NameTitle&q=>

Documentation:

- <http://www.loc.gov/bibframe>
- <https://github.com/lcnetdev>



# Questions?

- Nate Trail
- LS/ABA/NDMSO
- Library of Congress
- ntra@loc.gov