MARC to BIBFRAME: Take 2

SWIB17 - Semantic Web in Libraries

4 December 2017, Hamburg, DE

Wayne Schneider, Index Data
wayne@indexdata.com
http://www.indexdata.com

whoami

- Library software developer
- 21 years experience in libraries, academic and public
- Primary developer of marc2bibframe2
- Based in Minneapolis, MN

- Information toolmakers
- Standards-based solutions
- Long experience and practice in open source
- 23 years providing solutions for libraries and others

MARC to BIBFRAME: RDF

```xml
<record xmlns="http://www.loc.gov/MARC21/slim">
  <leader>01094cam a2200229 a 4500</leader>
  <controlfield tag="001">13600108</controlfield>
  <controlfield tag="005">20040915141353.0</controlfield>
  <controlfield tag="008">040520s2001    dk a   j      000 1 dan  </cont
  <datafield tag="010" ind1=" " ind2=" ">
    <subfield code="a">  2004436018</subfield>
  </datafield>
  <datafield tag="020" ind1=" " ind2=" ">
    <subfield code="a">8759517352</subfield>
  </datafield>
  <datafield tag="040" ind1=" " ind2=" ">
    <subfield code="a">DLC</subfield>
    <subfield code="c">DLC</subfield>
  </datafield>
  <datafield tag="050" ind1="0" ind2="0">
    <subfield code="a">PZ54</subfield>
    <subfield code="b">.A498 2001</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Andersen, H. C.</subfield>
    <subfield code="q">(Hans Christian),</subfield>
    <subfield code="d">1805-1875.</subfield>
  </datafield>
  <datafield tag="245" ind1="1" ind2="0">
    <subfield code="a">Ole Lukøie /</subfield>
    <subfield code="c">H.C. Andersen ; illustrationer af Otto Dickmeiss.<
  </datafield>
  <datafield tag="260" ind1=" " ind2=" ">
    <subfield code="a">[København?] :</subfield>
    <subfield code="b">Lindhardt og Ringhof,</subfield>
    <subfield code="c">c2001.</subfield>
  </datafield>
  <datafield tag="300" ind1=" " ind2=" ">
    <subfield code="a">1 v. (unpaged) :</subfield>
    <subfield code="b">col. ill. ;</subfield>
    <subfield code="c">23 cm.</subfield>
  </datafield>
  <datafield tag="500" ind1=" " ind2=" ">
    <subfield code="a">Text based on the original work published under th
originale eventyrhæfter med varianter ved Erik Dal og kommentarer ved Er
```

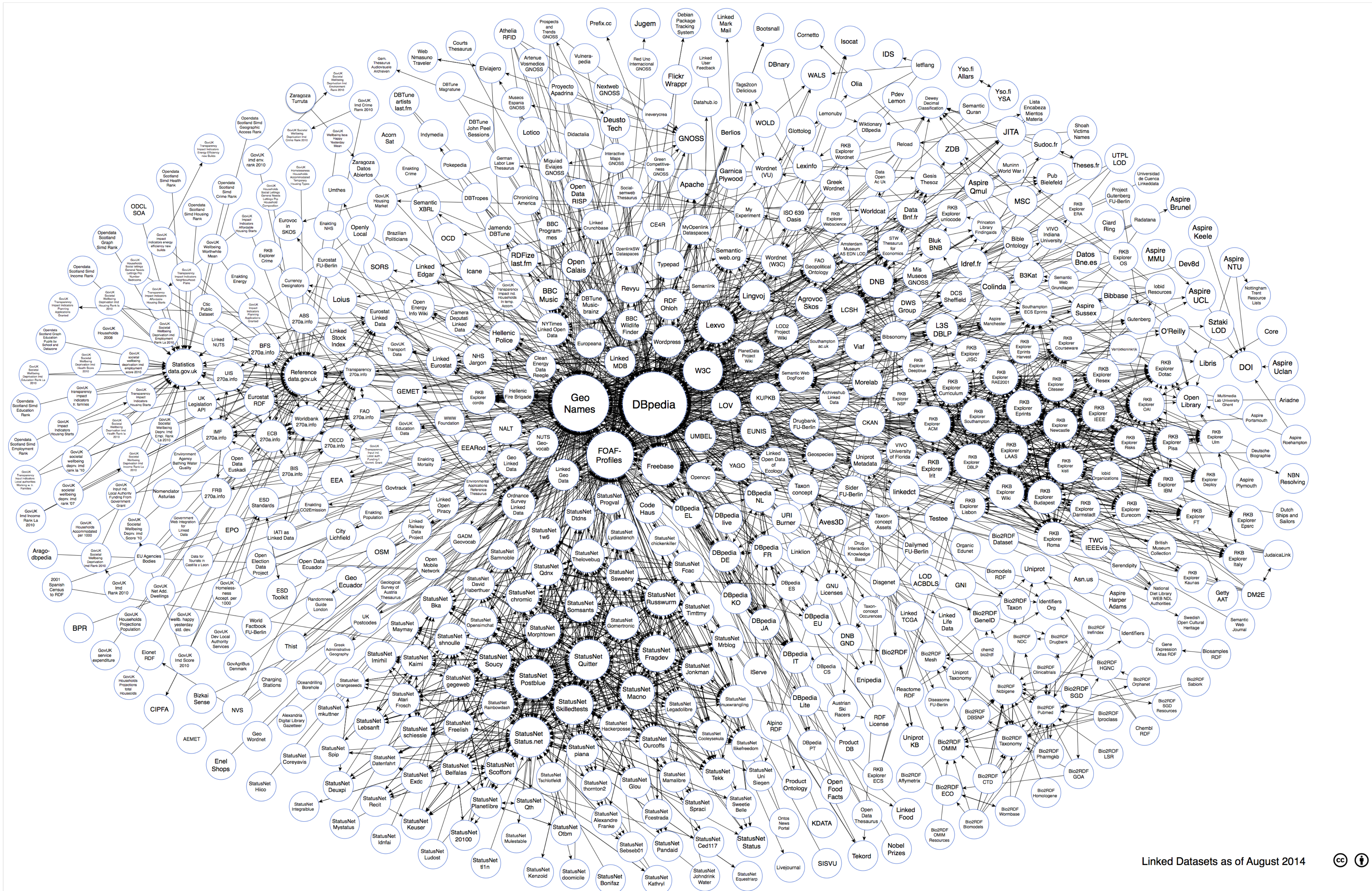
```
<http://example.org/13600108#Work>
  a <http://id.loc.gov/ontologies/bibframe/Work>, <http://id.loc.gov/ontologies/bibframe/
  ns0:adminMetadata [
    a ns0:AdminMetadata ;
    ns0:status [
      a ns0:Status ;
      ns0:code "c"
    ] ;
    ns1:encodingLevel [
      a ns1:EncodingLevel ;
      ns0:code "f"
    ] ;
    ns0:descriptionConventions [
      a ns0:DescriptionConventions ;
      ns0:code "aacr"
    ] ;
    ns0:identifiedBy [
      a ns0:Local ;
      rdf:value "13600108"
    ] ;
    ns0:changeDate "2004-09-15T14:13:53"^^xsd:dateTime ;
    ns0:creationDate "2004-05-20"^^xsd:date ;
    ns0:source [
      a ns0:Source, ns0:Agent ;
      rdfs:label "DLC"
    ], [
      a ns0:Source, ns0:Agent ;
      rdfs:label "DLC"
    ]
  ] ;
  ns0:language <http://id.loc.gov/vocabulary/languages/dan> ;
  ns0:intendedAudience <http://id.loc.gov/vocabulary/maudience/juv> ;
  ns0:genreForm <http://id.loc.gov/vocabulary/marcgt/fic> ;
```

Partial RDF graph for *Ole Lukøie*

100 1#$a Andersen, H. C.**$q**(Hans Christian)**,$d**1805-1875.

```
<bf:contribution>
  <bf:Contribution>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bflc/PrimaryContribution"/>
    <bf:agent>
      <bf:Agent rdf:about="http://example.org/13600108#Agent100-2">
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
        <bflc:name00MatchKey>
          Andersen, H. C. (Hans Christian), 1805-1875.
        </bflc:name00MatchKey>
        <bflc:primaryContributorName00MatchKey>
          Andersen, H. C. (Hans Christian), 1805-1875.
        </bflc:primaryContributorName00MatchKey>
        <bflc:name00MarcKey>
          1001 $aAndersen, H. C.$q(Hans Christian),$d1805-1875.
        </bflc:name00MarcKey>
         <rdfs:label>
           Andersen, H. C. (Hans Christian), 1805-1875.
         </rdfs:label>
      </bf:Agent>
    </bf:agent>
    <bf:role>
      <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/ctb"/>
    </bf:role>
  </bf:Contribution>
</bf:contribution>
```

The LOD Cloud



Conversion Processes

- Static Conversion
- Active Conversion
- Reconciliation

Really?

W3C Recommendation

XSL Transformations (XSLT)
Version 1.0

W3C Recommendation 16 November 1999

This version:
http://www.w3.org/TR/1999/REC-xslt-19991116
(available in XML or HTML)
Latest version:
http://www.w3.org/TR/xslt
Previous versions:
http://www.w3.org/TR/1999/PR-xslt-19991008
http://www.w3.org/1999/08/WD-xslt-19990813
http://www.w3.org/1999/07/WD-xslt-19990709
http://www.w3.org/TR/1999/WD-xslt-19990421
http://www.w3.org/TR/1998/WD-xsl-19981216
http://www.w3.org/TR/1998/WD-xsl-19980818
Editor:
James Clark <jjc@jclark.com>

Copyright © 1999 W3C® (MIT, INRIA, Keio), All Rights Reserved. W3C liability, trademark, document use and software licensing rules apply.

Abstract

This specification defines the syntax and semantics of XSLT, which is a language for transforming XML documents into other XML documents.

XSLT is designed for use as part of XSL, which is a stylesheet language for XML. In addition to XSLT, XSL includes an XML vocabulary for specifying formatting. XSL specifies the styling of an XML document by using XSLT to describe how the document is transformed into another XML document that uses the formatting vocabulary.

Conversion Specifications

http://www.loc.gov/bibframe/mtbf/

210 - ABBREVIATED TITLE (R)	##=Instance - title - VariantTitle - AbbreviatedTitle	
Indicators		
# - Abbreviated key title	## - source - Source - rdf:value - "issnkey"	
0 - Other abbreviated title	ignore	
Subfield Codes	## - rdfs:label [Make title label by extracting $ab, keep order in field, substitute blank for each subfield code]	
$a - Abbreviated title (NR)	## - mainTitle - literal	
$b - Qualifying information (NR)	## - qualifier - literal	
$2 - Source (R)	## - source -Source - rdfs:label - "content of $2"	
$6 - Linkage (NR)	See Subfield $6 spec	

XSpec

https://github.com/xspec/xspec

TEST REPORT

Stylesheet: **Users/wayne/workspace/marc2bibframe2/xsl/marc2bibframe2.xsl**

Tested: 20 January 2017 at 14:06

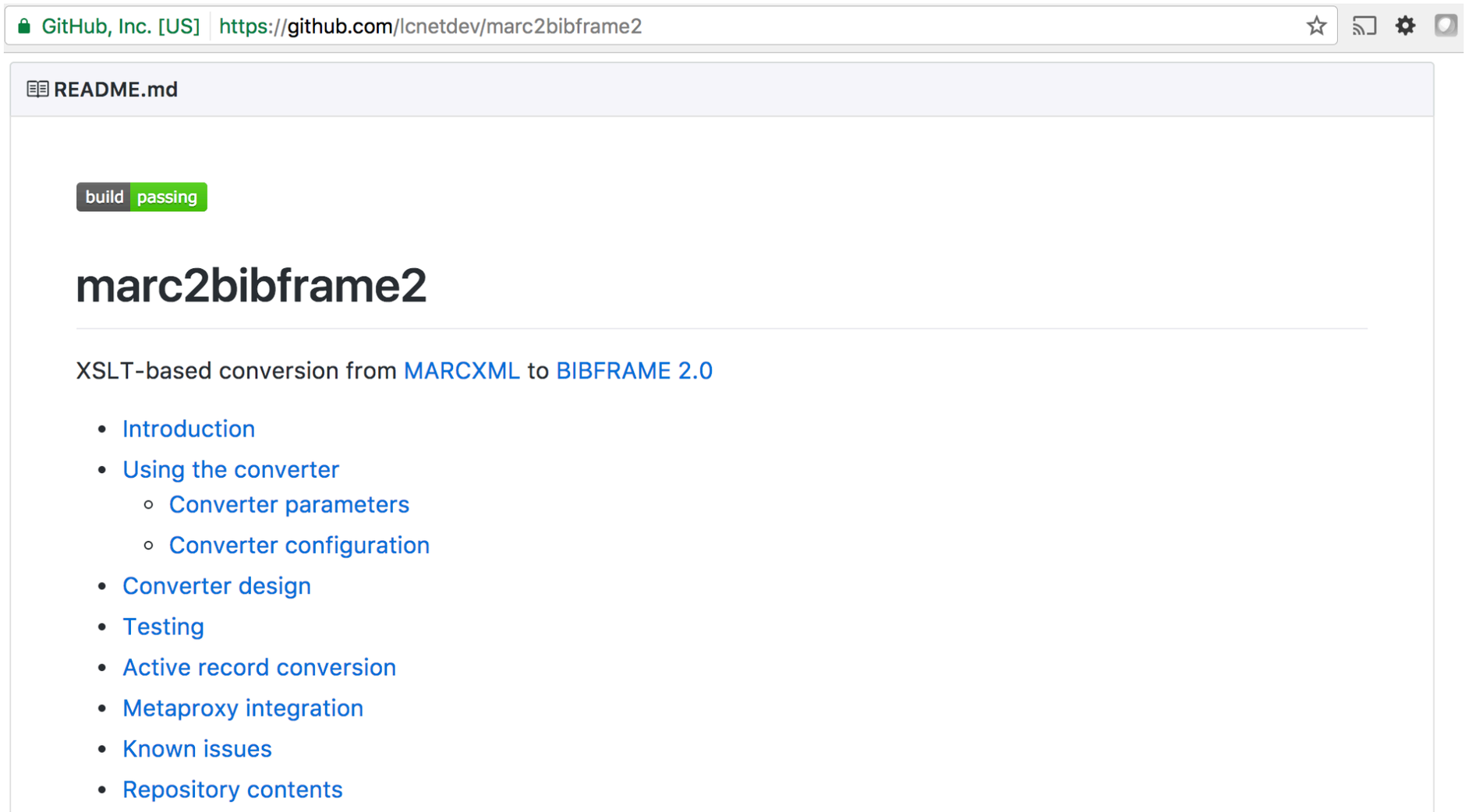
Contents

passed/pending/failed/total **903/0/0/903**

Regression	2/0/0/2
Record ID	1/0/0/1
LDR	8/0/0/8
001 - CONTROL NUMBER	1/0/0/1
003 - CONTROL NUMBER IDENTIFIER	1/0/0/1
005 - DATE AND TIME OF LAST TRANSACTION	1/0/0/1
007 - PHYSICAL DESCRIPTION FIXED FIELD	88/0/0/88
006 - FIXED-LENGTH DATA ELEMENTS-- ADDITIONAL MATERIAL CHARACTERISTICS	1/0/0/1
008 - FIXED-LENGTH DATA ELEMENTS--GENERAL INFORMATION	34/0/0/34
010 - LIBRARY OF CONGRESS CONTROL NUMBER	2/0/0/2
015 - NATIONAL BIBLIOGRAPHY NUMBER	4/0/0/4
016 - NATIONAL BIBLIOGRAPHIC AGENCY CONTROL NUMBER	4/0/0/4
017 - COPYRIGHT OR LEGAL DEPOSIT NUMBER	5/0/0/5
020 - INTERNATIONAL STANDARD BOOK NUMBER	5/0/0/5
022 - INTERNATIONAL STANDARD SERIAL NUMBER	6/0/0/6
024 - OTHER STANDARD IDENTIFIER	7/0/0/7
025 - OVERSEAS ACQUISITION NUMBER	1/0/0/1
026 - FINGERPRINT IDENTIFIER	4/0/0/4
027 - STANDARD TECHNICAL REPORT NUMBER	2/0/0/2
028 - PUBLISHER NUMBER	4/0/0/4
030 - CODEN DESIGNATION	2/0/0/2
032 - POSTAL REGISTRATION NUMBER	2/0/0/2
033 - DATE/TIME AND PLACE OF AN EVENT	7/0/0/7

marc2bibframe2

https://github.com/lcnetdev/marc2bibframe2



Active Conversion - Configuration

```
<backend syntax="xml">
  <xslt stylesheet="xsl/marc2bibframe2.xsl">
    <param name="idsource" value="http://id.loc.gov/vocabulary/organizations/dlc"/>
  </xslt>
  <rdf-lookup timeout="3">
    <namespace prefix="bf" href="http://id.loc.gov/ontologies/bibframe/"/>
    <namespace prefix="bflc" href="http://id.loc.gov/ontologies/bflc/"/>
    <namespace prefix="rdf" href="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
    <lookup xpath="//bf:Agent[contains(@rdf:about,'example.org')]">
      <key field="bflc:name11MatchKey"/>
      <key field="bflc:name00MatchKey"/>
      <key field="bflc:name10MatchKey"/>
      <server url="http://id.loc.gov/authorities/names/label/%s" method="HEAD"/>
    </lookup>
  </rdf-lookup>
  <rdf-lookup timeout="3">
    <namespace prefix="bf" href="http://id.loc.gov/ontologies/bibframe/"/>
    <namespace prefix="bflc" href="http://id.loc.gov/ontologies/bflc/"/>
    <namespace prefix="rdf" href="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
    <namespace prefix="rdfs" href="http://www.w3.org/2000/01/rdf-schema#"/>
    <lookup xpath="//bf:Work/bf:subject/bf:Temporal[contains(@rdf:about,'example.org')] |
                   //bf:Work/bf:subject/bf:Topic[contains(@rdf:about,'example.org')] |
                   //bf:Work/bf:subject/bf:Place[contains(@rdf:about,'example.org')] |
                   //bf:Work/bf:genreForm/bf:GenreForm[contains(@rdf:about,'example.org')]">
      <key field="rdfs:label"/>
      <server url="http://id.loc.gov/authorities/subjects/label/%s" method="HEAD"/>
    </lookup>
  </rdf-lookup>
  <rdf-lookup timeout="3">
    <namespace prefix="bf" href="http://id.loc.gov/ontologies/bibframe/"/>
    <namespace prefix="bflc" href="http://id.loc.gov/ontologies/bflc/"/>
    <namespace prefix="rdf" href="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
    <namespace prefix="rdfs" href="http://www.w3.org/2000/01/rdf-schema#"/>
    <lookup xpath="//bf:Role[not(@rdf:about)] | //bflc:relation/bflc:Relation[not(@rdf:about)]">
      <key field="bflc:relatorMatchKey"/>
      <key field="rdfs:label"/>
      <server url="http://id.loc.gov/vocabulary/relators/label/%s" method="HEAD"/>
    </lookup>
  </rdf-lookup>
</backend>
```

Metaproxy Integration


```
▼<zs:searchRetrieveResponse xmlns:zs="http://www.loc.gov/zing/srw/">
  <zs:version>1.1</zs:version>
  <zs:numberOfRecords>1</zs:numberOfRecords>
 ▼<zs:records>
  ▼<zs:record>
    <zs:recordSchema>bibframe2</zs:recordSchema>
    <zs:recordPacking>xml</zs:recordPacking>
   ▼<zs:recordData>
    ▼<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     xmlns:bf="http://id.loc.gov/ontologies/bibframe/" xmlns:bflc="http://id.loc.gov/ontologies/bibframe/lc-extensions/"
     xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#">
     ▼<bf:Work rdf:about="http://example.org/13600108#Work">
      ▶<bf:adminMetadata>...</bf:adminMetadata>
       <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Text"/>
      ▼<bf:language>
         <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/dan"/>
       </bf:language>
      ▼<bf:intendedAudience>
       ▼<bf:IntendedAudience rdf:about="http://id.loc.gov/vocabulary/maudience/juv">
          <rdfs:label>juvenile</rdfs:label>
       </bf:IntendedAudience>
  </bf:intendedAudience>
 ▼<bf:genreForm>
  ▼<bf:GenreForm rdf:about="http://id.loc.gov/vocabulary/marcgt/fic">
     <rdfs:label>fiction</rdfs:label>
   </bf:GenreForm>
  </bf:genreForm>
 ▼<bf:classification>
  ▼<bf:ClassificationLcc>
   ▼<bf:source>
      <bf:Source rdf:about="http://id.loc.gov/vocabulary/organizations/dlc"/>
    </bf:source>
    <bf:classificationPortion>PZ54</bf:classificationPortion>
    <bf:itemPortion>.A498 2001</bf:itemPortion>
   </bf:ClassificationLcc>
  </bf:classification>
 ▼<bf:contribution>
  ▼<bf:Contribution>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/lc-extensions/PrimaryContribution"/>
   ▼<bf:agent>
    ▼<bf:Agent rdf:about="http://example.org/13600108#Agent100-12">
      <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
      <bflc:name00MatchKey>Andersen, H. C. (Hans Christian), 1805-1875.</bflc:name00MatchKey>
      <bflc:primaryContributorName00MatchKey>Andersen, H. C. (Hans Christian), 1805-1875.</bflc:primaryContributorName00MatchKey>
     ▼<bflc:name00MarcKey>
        1001 $aAndersen, H. C.$q(Hans Christian),$d1805-1875.
       </bflc:name00MarcKey>
       <rdfs:label>Andersen, H. C. (Hans Christian), 1805-1875.</rdfs:label>
     </bf:Agent>
```


http://www.indexdata.com/metaproxy

Conversion Challenges

- Deliberate duplication of data in control fields and data fields

- Data normalization vs. presentation

255 ##$aScales vary$d(Zones +90° to +81° to 63°, -81° to 98° ;$eeq. 1950).

- Multiscript records

- Item information

- Pace of change

Future Directions

- Refine specifications to prevent duplication of elements
- Improve handling of multi script records
- Configuration driven
- Mapping schema (http://rml.io)
- More refined active conversion

MARC to BIBFRAME Sample Workflows

- National Library of Finland RDF pipeline

  https://github.com/NatLibFi/bib-rdf-pipeline

- Index Data YAZ toolkit

  http://www.indexdata.com/yaz

MARC to BIBFRAME: Take 2

SWIB17 - Semantic Web in Libraries

4 December 2017, Hamburg, DE

Wayne Schneider, Index Data
wayne@indexdata.com
http://www.indexdata.com