The Libhub Initiative: Increasing the Web Visibility of Libraries

DCMI / ASIS&T Webinar
January 7, 2015

Eric Miller
em@zepheira.com
https://www.linkedin.com/in/erimille
@erimille
Columbus, OH
Wednesday 7:00 AM
Mostly Cloudy

Temperature: 11°F
Humidity: 67%
Wind: 3 mph
Precipitation: 20%

Day Weather:
- Wed: 16° -9°
- Thu: 18° 12°
- Fri: 19° -6°
- Sat: 23° 5°
- Sun: 45° 25°
- Mon: 37° 10°
- Tue: 43° 25°
- Wed: 43° 21°
My Goal

1. Sharing finds from the field and practitioners perspective over the past year

2. Description (Bibframe) and Discovery (schema.org) standards as complements

3. What do we mean by Web visibility?

4. Introduce The Libhub Initiative designed to accelerate the visible library.
A webinar in 3 acts
Act 1:
Background and Context
Hi, We are Zepheira

We are a Passionate, Experienced, Strategic, Technical, Entrepreneurial Team. We love Libraries, Data, The Web, Standards, Open Source, and Learning.

The team’s background:

http://linkedin.com/company/zepheira/
http://zepheira.com/
Zepheira’s Perspectives on Leadership and Practice
Libraries, Web, Data & Technology

Standards
- RDF
- Semantic Web
- W3C
- OCLC
- Library of Congress

Data and Tech Strategies
- BIBFRAME
- Schema.org

Vocabularies

Projects
- NLB
- National Library of Medicine

Activities
- Libhub

Large Scale Projects
- BIBFLOW
  - An IMLS Project of the UC Davis Library & Zepheira
Perspective

• The Web enables many perspectives

• The utility of those perspectives is another perspective

• And that's by design

• And this is what we've heard and coupled with our professional experiences
Shared observations

1. A Library is more than its collection

2. “Linked Data” is Linked first, Data second

3. The Web of Data is here

4. And Libraries are not participating to their full potential
Minimum Viable Product, Incremental Value, and Continuous Learning

Not like this....

1 2 3 4

Like this!

1 2 3 4 5

Henrik Kniberg
BIBFRAME Vocabulary
What is Schema.org?

This site provides a collection of schemas that webmasters can use to markup HTML pages in ways recognized by major search providers, and that can also be used for structured data interoperability (e.g. in JSON). Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right Web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.

We invite you to get started!

View our blog at blog.schema.org.
Thing > CreativeWork > Book
A book.

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bookFormat</td>
<td>BookFormatType</td>
<td>The format of the book.</td>
</tr>
<tr>
<td>illustrator</td>
<td>Person</td>
<td>The illustrator of the book.</td>
</tr>
<tr>
<td>numberOfPages</td>
<td>Integer</td>
<td>The number of pages in the book.</td>
</tr>
<tr>
<td>about</td>
<td>Thing</td>
<td>The subject matter of the content.</td>
</tr>
<tr>
<td>accessibilityAPI</td>
<td>Text</td>
<td>Indicates that the resource is compatible with the referenced accessibility API (<a href="https://schema.org/accessibilityAPI">WebSchemas wiki lists possible values</a>).</td>
</tr>
<tr>
<td>accessibilityControl</td>
<td>Text</td>
<td>Identifies input methods that are sufficient to fully control the described resource (<a href="https://schema.org/accessibilityControl">WebSchemas wiki lists possible values</a>).</td>
</tr>
<tr>
<td>accessibilityFeature</td>
<td>Text</td>
<td>Content features of the resource, such as accessible media, alternatives and supported enhancements for accessibility (<a href="https://schema.org/accessibilityFeature">WebSchemas wiki lists possible values</a>).</td>
</tr>
<tr>
<td>accessibilityHazard</td>
<td>Text</td>
<td>A characteristic of the described resource that is physiologically dangerous to some users. Related to WCAG 2.0 guideline 2.3. (<a href="https://schema.org/accessibilityHazard">WebSchemas wiki lists possible values</a>).</td>
</tr>
<tr>
<td>accountablePerson</td>
<td>Person</td>
<td>Specifies the Person that is legally accountable for the CreativeWork.</td>
</tr>
<tr>
<td>aggregateRating</td>
<td>AggregateRating</td>
<td>The overall rating, based on a collection of reviews or ratings, of the item.</td>
</tr>
<tr>
<td>alternativeHeadline</td>
<td>Text</td>
<td>A secondary title of the CreativeWork.</td>
</tr>
<tr>
<td>associatedMedia</td>
<td>MediaObject</td>
<td>A media object that encodes this CreativeWork. This property is a synonym for encoding.</td>
</tr>
<tr>
<td>audience</td>
<td>Audience</td>
<td>The intended audience of the item, i.e. the group for whom the item was created.</td>
</tr>
<tr>
<td>audio</td>
<td>AudioObject</td>
<td>An embedded audio object.</td>
</tr>
<tr>
<td>author</td>
<td>Person or Organization</td>
<td>The author of this content. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably.</td>
</tr>
<tr>
<td>award</td>
<td>Text</td>
<td>An award won by this person or for this creative work. Supersedes awards.</td>
</tr>
</tbody>
</table>
Linked Data

"a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."

http://en.wikipedia.org/wiki/Linked_Data
Phases of Linked Data / BIBFRAME Adoption

THE CHASM

Technology Enthusiasts

Visionaries

The Early Market

Pragmatists

Conservatives

The Mainstream Market

Skeptics
Phases of Linked Data / BIBFRAME Adoption

Experimenters
Phases of Linked Data / BIBFRAME Adoption

The CHASM

- Technology Enthusiasts
- Visionaries
- Pragmatists
- Conservatives
- Skeptics

The Early Market

Experimenters

Early Implementers
Phases of Linked Data / BIBFRAME Adoption

THE CHASM

- Technology Enthusiasts
- Visionaries
- Pragmatists
- Conservatives
- Skeptics

The Early Market

Experimenters
Early Implementers

Data Publishers & Connectors

The Mainstream Market

bf: BIBFRAME
Phases of Linked Data / BIBFRAME Adoption

THE CHASM

The Early Market
- Technology Enthusiasts
- Visionaries

The Mainstream Market
- Pragmatists
- Conservatives
- Skeptics

Experimenters
- Early Implementers

Data Publishers & Connectors

Mainstream Workflow
Phases of Linked Data / BIBFRAME Adoption

THE CHASM

The Early Market

Technology Enthusiasts
Visionaries

Early Implementers

The Mainstream Market

Data Publishers & Connectors
Mainstream Workflow

Pragmatists
Conservatives
Skeptics

Back Office Systems
Phases of Linked Data / BIBFRAME Adoption
Phases of Linked Data / BIBFRAME Adoption
Phases of Linked Data / BIBFRAME Adoption
Phases of Linked Data / BIBFRAME Adoption
Phases of Linked Data / BIBFRAME Adoption
Act 2:
Speak like the Web
Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management control
The Web (Redux)

- Identifiers (URI)
- Relationships (Links)
- Services
MARC as "Things not Strings"
MARC as “Things not Strings”
MARC as “Things not Strings”

```
001 26132811
008 920608s1992 nyu
010 $a92021087
020 $a1555701205
040 $aDLC$cDLC$dAGL
049 $aHNKA
050 00 $aZ678.9$b.C58 1992
082 00 $a025/002855220
10 $aCohn, John M
260 $aNew York: $bNeal Schuman Publishers, $cc1992
300 $a116 p. ; $c28 cm
440 0 $aHow-to-do it manuals for libraries; $uno. 25
504 $aIncludes bibliographical references and index
650 0 $aLibraries$xAutomation$xManagement$xHandbooks, manuals, etc
650 0 $aLibrary planning$xHandbooks, manuals, etc
700 10 $aKelsey, Ann L
700 10 $aFiel, Keith Michael
```
MARC as “Things not Strings”
MARC as “Things not Strings”
BIBFRAME
Core model for defining Web control points of bibliographic data for more effective sharing, navigation and collaboration
Simple, replicable linked data patterns
And we can replicate these simple patterns to define as many control points we need.
Practical Practitioners

Content

Conversation

Community
As a cornerstone of Zepheira's Practical Practitioner series, we believe participants need to get hands on with their own data. To that end, our training includes access to an introductory set of utilities to guide participants through data import, analysis and end user discovery.

The following MARC diagnostics and BIBFRAME transformation service is provided to participants of this training curriculum. To get started link to, paste in, or upload a MARC XML file (up to 100 records). Feel free to add a 'title' to the data to help describe what you've uploaded.

Don't have access to a set of MARC records? Here are some examples others have uploaded. Drag or copy the following links to the URL Input form above to transform these records.

- George Washington sample collection (small)
- National Library of Medicine sample collection (medium)
- George Mason University DVD Backstage sample (large)
- Sample Rare Book / Manuscript collection (small)
- Sample Music collection (small)
Excellent! The transformation and diagnostics process was successful. Feel free to bookmark this page to come back to it in the future. Now that you have, you're ready to begin!

1. To better understand your uploaded data first look at the MARC Field Diagnostics Viewer.

2. Now take a look at the BIBFRAME Resource Viewer to help better understand the MARC Records to BIBFRAME Resources transformation process.

3. And finally here is a BIBFRAME Work-focused Navigator or a BIBFRAME Collection Navigator to illustrate how your data might look like in a traditional discovery application.

4. Don't forget to Download your RDF BIBFRAME data.

5. And since the future direction of these tools rely on your input Help Shape What Comes Next.

---

**Summary**

- Processing started on **2014-12-02 at 15:38**
- **193** MARC records analysed in **6.99** seconds
- Input Format:

**Observations**

- **193** MARC records converted to **1214** BIBFRAME Resources
- **230** BIBFRAME Resources folded indicating this data is moderately connected
NLM SAMPLE

1423 Resources

sorted by: labels; then by... • grouped as sorted

1 • 2 • 3 ... 72 Next »

1. oXD85ZJ7 (link)
   label: oXD85ZJ7
   type: Topic
   URI: https://linksmith.ze ... mple_3/item#oXD85ZJ7
   name: Hospitals, Military
   sf-x: history
   sf-a: Hospitals, Military
   bf-id: oXD85ZJ7
generalSubdivision:history

2. qNmTh3yY (link)
   label: qNmTh3yY
   type: Work
   URI: https://linksmith.ze ... mple_3/item#qNmTh3yY
   bf-id: qNmTh3yY
   mediaCategory: unmediated
   tag-995a: AUTH
   language: rus
   tag-880c: [2012], выпускающий редактор, А.В. Харевич.
   contentTypeCode: txt
   nlmCallNumber: WX 27 GR9, 2013 B-503
   tag-0359: 101603036
   subject: OZy2W9Qb, r59xqPrn, PlrWz1ep, oXD85ZJ7
   tag-995b: 20130909
197 Works

1. **305 let pervomu gosudarstvennomu lechebnomu uchrezhdeniuiRossii : 1707-2012 / vypuskaishchii redaktor, A.V. Kharevin.**

   **Contributor(s)**  Kharevin, A. V.,  

   **Subject(s)**  Military Medicine  

   **Language**  rus  

   **NLM Call Number**  WX 27 GR9 and 2013 B-503

   **Item(s)**  

   **Publisher(s)**  Éko-Press, [2012] Moskva :  

   **ISBN**  9785904301880  

   **Extent**  121 pages

2. **Put’ vverkh k podnozhiui vishchin : avtobiograficheskie zametki / Nikolai Mushkambarev.**
29 Items

1. **Disasters of war.**
   - Member(s): Aun podrán servir Goya.

2. **Medical subject headings.**
   - Member(s): Qā'īlamat ru'ūs al-mawḍū'ūt al-ṭībīyah.

3. **Duus' neurologisch-topische Diagnostik.**
   - Member(s): Duus' topical diagnosis in neurology: Mathias Baehr, Michael Frotscher; with contributions by Wilhelm Kueker; founding author, Peter Duus; 400 illustrations by Gerhard Spitzer and Barbara Gay.
   - Member(s): Duus' topical diagnosis in neurology: Mathias Baehr, Michael Frotscher; with contributions by Wilhelm Kueker; founding author, Peter Duus; translated by Ethan Taub; 400 illustrations, most in color, by Gerhard Spitzer and Barbara Gay.

4. **Physica.**
   - Member(s): Livre des subtilités des créatures de diverses natures = Sainte Hildegarde de Bingen; traduits, présentés et annotés par Bernard Verten.
BIBFRAME RESOURCE VIEWER

NLM SAMPLE

1423 Resources

sorted by: labels; then by... • grouped as sorted

1 · 2 · 3 ... 72 Next »

1. oXD85Zj7 (link)
   label: oXD85Zj7
   type: Topic
   URI: https://linksmith.ze ... mple_3/item#oXD85Zj7
   name: Hospitals, Military
   sf-x: history
   sf-a: Hospitals, Military
   bf-id: oXD85Zj7
   generalSubdivision: history

2. qNmTh3yY (link)
   label: qNmTh3yY
   type: Work
   URI: https://linksmith.ze ... mple_3/item#qNmTh3yY
   bf-id: qNmTh3yY
   mediaCategory: AUTH
   language: rus
   tag-880c: [2012], выпускающий редактор, А.В. Харевин.
   contentTypeCode: txt
   nlmCallNumber: WX 27 GR9, 2013 B-503
   tag-0359: 101603036
   subject: OZY2W9Qb, rS9xQPmm, PlrWzlep, oXD85Zj7
   tag-995b: 20130909
A link is worth a 1000 words
Summary

- 193 MARC records converted to 1423 BIBFRAME Resources
- 498 BIBFRAME Resources folded indicating this data is highly connected
- 168 MARC tags did not match applied transformation rules indicating this data is moderate candidate for custom rule design
- 38 MARC tags identified as local extensions indicating this data is strong candidate for local rule design

Resource Types

1423 BIBFRAME Resources were materialized from 193 MARC records
1423 BIBFRAME Resources were materialized from 193 MARC records

The following is a breakdown of Resource Types by total count of materialized resources.
Folding

Of the 1423 BIBFRAME Resources generated, 498 resources folded. The following is a percentage breakdown by Resource type.

The following is a breakdown of Resource Types by Total count of materialized resources and associated Folding counts.
Opportunity
Then we took a turn:

Brown, Margaret Tobin, 1867-1932

Molly Brown: sharing her good fortune

Molly Brown House

Politics and Government

Suffragists—United States—Colorado

The unsinkable Molly Brown: a musical comedy

Titanic (Steamship)

Shipwreck Survival

Molly Brown House Museum

National Trust for Historic Preservation in the United States

Denver (Colo.)

Local history

Distinctive Destinations

BIBFRAME Lite Icon Key

Link

Annotation

Work

Instance

Authority

Icons made by Freepik at www.flaticon.com and licensed under Creative Commons BY 3.0
In Summary

1. Reflecting data into Bibframe provides a highly connected web of data

2. Enormous potential for increased connectivity

3. Control points exist to connect this even more
   1. No other community does authorities like we do

4. We're making it extremely difficult to connect
   1. Lower costs to linking is critical to improved visibility

5. As of now, (relatively) completely dark to the Web
Act 3 : Visibility
“When my community searches the Web for something we have, we better show up as an option.”

- Chuck Gibson, Director & CEO
Worthington Public Library
Can’t ignore the problem...
Start with Agreement and Purpose

“Everyone benefits from the visibility of libraries and their content on the Web.”
Learning through action together
But we aren’t speaking in a way the Web understands

We have a wealth of content and resources locked behind legacy, closed technology systems and niche vocabularies
The traditional, **Visible** Web focuses on Harvesting and Links to Pages
The emerging **Invisible Web** focuses on Data, Resources, Vocabulary, and Connections.
New Vocabularies and Characteristics
Retail – items, reviews, geo, descriptions, inventory, hours, social, events
New Vocabularies and Characteristics
Movies – Geo, reviews, ratings, images, previews, times, tickets

I found quite a number of movies playing today near San Francisco:

Now Playing
San Francisco

The Spectacular
1:25
88%
R

The Wolverine
10:05
68%
P5-13

Monsters University
10:45
79%
G

Rave Motion Pictures
Address: 1071 Gemini Pi, Columbus, OH 43240
Phone: (614) 781-8228

Reviews
4.2 ★★★★ 34 Google reviews

"Love the seats so comfortable nice bathrooms two always kept clean." 8 reviewers

"The screen is big and the sound is incredible." 3 reviewers

View all Google reviews

More reviews: yahoo.com, abumba.com, yellowpagea.com, superpages.com

People also search for

AMC Easton Town Center 30
Marcus Crosswoods Cinema
Strand Theatre
Regal Geogesv... Square Stadium 16
New Vocabularies and Characteristics
Restaurants – locations, reviews, hours, reservations, menus
How does the Web see Libraries?
Libraries = Community Businesses
Location, photos, hours, reviews, social, events
If at all....
External Perspectives

• Are websites and systems harvestable?
• Is there a unified and accessible industry vocabulary?
• Are there strong connections and relationships?
• What is the consistency and reliability of the user experience and available data?
This is the now

60+ Pages later.... still not even one entry that had anything to do with Libraries
FACEPALM
For when epic fail just doesn't cut it.
This is what a search engine harvester sees. Unconnected data results in poor page rank.
Good

Hardcover

isHeldBy

isHeldBy

isHeldBy

isHeldBy

isHeldBy
Better
Best!

And Linked Data is a key
A link is worth a 1000 words
This pattern is already happening in many localized markets as we speak
Resulting in front page consolidation
But we’re very close

• MARC To BIBFRAME (social)

• Frustration with consolidation in marketplace (economic)

• Web is increasingly actionable / semantic e.g. schema.org (technical)
BIBFRAME

Purpose and Promise

- **Purpose**: Replacing MARC
- **Promise**: So much more

- **Purpose**: Serving Libraries
- **Promise**: Related memory organizations and the users they serve

- **Purpose**: Leverage existing Web standards to speak with a consistent voice
- **Promise**: Visibility, Discovery and Effectiveness
Library Description and Capacity

Systems and Services

Content and Collections
Link Accelerator
Link Accelerator
In the summer of 2015, Zepheira will introduce the Libhub Network, a Web-based data network designed to surface and connect libraries and their resources in a visible and actionable fashion.
• In the Network, each library will have their own set of data optimized for linking across the library resources, across other organizations with shared assets, and within the context of the broader Web to ensure that library content weaves into the Web in a usable and connected fashion.
A link is worth a 1000 words

PageRank
Don’t be too clever

- This isn’t about gaming a system
- It’s not about optimizing to any one service
- It’s about speaking in a way the Web understands
- And benefiting from what this enables
Quentin Tarantino - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Quentin_Tarantino
His films include Reservoir Dogs (1992), Pulp Fiction (1994), Jackie Brown (1997), .... Tarantino's contribution to the Grindhouse project was titled Death Proof. ... Tarantino's 2009 film Inglourious Basterds is the story of a group of guerrilla U.S. ..... co.uk/film/filmblog/2011/may/05/quentin-tarantino-django-unchained-script/
Quentin Tarantino filmography - Death Proof - Jackie Brown - New Beverly Cinema

Quentin Tarantino - IMDb
www.imdb.com/name/nm0000233/
In January of 1992, Reservoir Dogs appeared at the Sundance Film Festival, by ... up Dogs success with Pulp Fiction which premiered at the Cannes film festival. ... Quentin Tarantino Still of Quentin Tarantino in Django Unchained Still of ... Django Unchained - Biography - Death Proof - Kill Bill: Vol. 3

Quentin Tarantino movies, photos, movie reviews, filmography, and ...www.ainmovie.com/artists/quentin-tarantino-p113658
Word-of-mouth on Reservoir Dogs began to build at the 1992 Sundance Film Festival, ... During 1993, Tarantino wrote and directed his next feature, Pulp Fiction, ... Grindhouse into two portions: the first half, Death Proof, directed by Tarantino, ...

Quentin Tarantino Movies | Cracked.com
www.cracked.com/funny-1906-quentin-tarantino-movies/
Quentin Tarantino used to recommend movies while working in a video rental store, making him the most successful "annoying guy going on about 'good' "...

Everything Tarantino « The film works of Quentin Tarantino
www.everythingtarantino.com/
Jamie Foxx says that Quentin Tarantino rewrote the ending to Django Unchained in a matter of minutes after being unhappy with the previous conclusion.

Top Ten Quentin Tarantino Movies « Movies-Films-MotionPictures
moviesfilmsmotionpictures.com/.../top-ten-quentin-tarantino-...
Mar 27, 2010 – Reservoir Dogs is the first movie Quentin Tarantino ever directed and it is ... Death Proof is "the" guilty pleasure movie for me. ... It's really amazing how wildly entertaining and amazing a film that Pulp Fiction is. ... a doubt in my mind that Inglourious Basterds is Quentin Tarantino's greatest piece of cinema.

Quentin Tarantino

Born: March 27, 1963 (age 59), Knoxville

Upcoming movies: Django Unchained, Kill Bill: Vol. 3

Awards: Palme d'Or, National Society of Film Critics Award for Best Director, More

Education: Narbonne High School, Fleming Junior High School

<table>
<thead>
<tr>
<th>Song</th>
<th>Year</th>
<th>Album</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview</td>
<td>1997</td>
<td>Tarantino Connection</td>
</tr>
<tr>
<td>Pulp Fiction</td>
<td>1998</td>
<td>Interview Picture Disc</td>
</tr>
<tr>
<td>Jackie Brown</td>
<td>1998</td>
<td>Interview Picture Disc</td>
</tr>
</tbody>
</table>

Books

<table>
<thead>
<tr>
<th>People also search for</th>
<th>Name</th>
<th>Poster</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Romance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From dusk till dawn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quentin Tarantino</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death Proof</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grindhouse</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Moving the Needle and Transforming the Web

• NO NEED TO WAIT
• Build on existing investments
• Use Bibframe to reflect content in a Web of Data
• Project in to a vocabulary the Web understands (schema.org via RDFa)
• Leverage the Web’s cooperative infrastructure
• Link between shared & Web assets to test impact on results
• Help the Web understand library vocabularies
• Connect into legacy systems
What you can do Now

1. Explore your Web footprint
   1. Create a Wikipedia entry, Google+, Facebook, Yelp! etc. for your Library
   2. Understand the gap between your catalog and the Web
2. Explore Terry Reese’s MarcNext work (MarcEdit to Libhub pipeline)
3. Dan Scott’s Evergreen schema.org work in Evergreen
4. Express your support for library visibility at libhub.org
5. Get involved! Your perspective counts!
Not like this....

1
2
3
4

Like this!

1
2
3
4
5

Henrik Kniberg
Thank you

Eric Miller
em@zepheira.com