Nathan Putnam is the Director of Metadata Quality at OCLC and oversees the teams of specialists working on data quality for the WorldCat bibliographic database, knowledge base, and central index. He has over 15 years of cataloging experience working in multiple academic libraries, first as a cataloger and database maintenance specialist and more recently as the head of the metadata/cataloging department at the University of Maryland, College Park. Nathan joined OCLC in August 2015 to expand the scope and coverage of quality control and data clean-up.

As mentioned, I am over the teams working on WorldCat quality in the bibliographic database, knowledge base, and registry. The largest team works with the bibliographic database. Cynthia Whitacre oversees Metadata Policy in this area. The knowledge base and registry are areas where greater focus is going towards in terms of data quality. These areas will grow over time.
Libraries come together as OCLC to advance librarianship around the world – there is a commitment and a passion. Our cooperative is unique.

- Slide contains feedback from our Global Council members who were asked “why OCLC?”
- These quotes represent members’ appreciation for the uniqueness of the cooperative.

[Notes current as of 5/4/18]
OCLC libraries are part of an active and engaged global network of libraries with more than 16,000 members across 123 countries.

- Global network provides unparalleled access to the world’s collected knowledge in ways that cannot be achieved independently
- Ability to address shared challenges at scale
- Together, we connect people to the information they need to achieve their goals

[Per Ellen McCarthy, the 12/31/2017 update is the most current info available for this slide—notes current 5/4/18]
I would like to provide a brief update on the FirstSearch and WorldCat Discovery services.
One subscription = two interfaces

A FirstSearch/WorldCat Discovery subscription gives your library use of both the FirstSearch and WorldCat Discovery interfaces.

- Use FirstSearch for precise, full-featured searching of the WorldCat database.
- Use WorldCat Discovery to enable discovery of collections of your libraries and libraries worldwide through a single search of WorldCat and a central index of more than 2,500 e-content collections.

This subscription also makes your library’s collections visible on search engines and popular websites such as Wikipedia, Google Books and Goodreads, when you also represent your collections in WorldCat through an OCLC cataloging subscription. OCLC’s partnerships with websites where people usually begin their research mean most people reach WorldCat.org from search engine results in as few as two clicks. Once on WorldCat.org, they can see which libraries have the items they need.

[Notes current as of 6/5/18]
As you know, OCLC will continue to provide access to both FirstSearch and WorldCat Discovery.

In response to requests from users, we will release a restyled FirstSearch interface in 2018. These user-requested changes will give searchers the contemporary look they are accustomed to seeing in other web experiences while maintaining the reliability and precision search capabilities users have told us they value. Library staff will not need to change any settings or configurations to use the updated interface when it is released. Watch for an announcement from us with information about when the refreshed interface will be available.

[Notes current as of 6/5/18]
In WorldCat Discovery, we continue to investigate additional ways to deliver relevant search results that match searcher tasks. Searchers can select the task-based search algorithms and sort options that best match their specific search needs. Rigorous internal testing and piloting with member librarians resulted in the addition of two additional different relevance options: **Best Match and Recency**.

- Best Match, formerly called Relevance Only, retrieves known items that are good matches on visible fields.
- Recency retrieves more recent items, often considered more relevant for some academic disciplines.
- In addition, the previous **Library and Relevance** algorithm has been renamed **Library** and continues to surface local library holdings.

Users can also choose to sort results by: **Author, Date, Title, and Most Widely Held**. The **Most Widely Held** display does not cluster editions and formats, which is especially important to interlibrary loan staffs.

We appreciate the valuable insights from our members that have informed our work to continue search enhancements in WorldCat Discovery.

*Notes current as of 6/5/18*
Recent WorldCat Discovery enhancements

Search experience
- Group and ungroup editions and formats
- Transfer existing search query to Advanced Search

Streamline academic research
- Users can report broken links to help OCLC improve link quality
- Access to full-text links meet user expectations

Customization
- Re-use search filters across searches and in search boxes
- Include a custom message when emailing personal lists

• Users can now choose to display editions and formats as a group or as individual records.
• A query entered in the simple search box now auto populates the first box in Advanced search, when a user moves from the simple to Advanced search screen,
• Users can report broken links, to help OCLC improve link quality.
• Full-text links more clearly indicate the type of content available. For example, when the WorldCat knowledge base identifies an item as an e-book, the user will see a “View e-book” link.
• Users can reapply filters without manually selecting them for each search in a search session. Filters such as date or format can also be built into library search boxes, to save time for searchers.
• And library staff can include a custom message when emailing personal lists.

Details are available in Release Notes, on the WorldCat Discovery Support and Training website. A roadmap of recent and planned enhancements is on the WorldCat Discovery Community Center.

[Notes current as of 6/5/18]
Enhancements for WMS libraries

• For library staff:
  – More configuration options for Place Hold form
  – Create custom shelving location messages and shelving notes

• For library users:
  – Request multiple and specific copies in a single hold request
  – Provide details about preferred pickup location for holds
  – Submit a custom note to give library staff more information about hold requests

Recent user-driven enhancements specifically for libraries that use WMS have focused on the Place Hold feature.

In response to user requests, library staff now have more control over configuration of the information and choices that display on Place Hold forms. Staff can also help users more clearly understand the locations of needed items, when they create custom shelving location messages and shelving notes.

Library users can save time by requesting multiple copies and designating specific copies when placing hold requests. They can also send additional details to library staff about items they place on hold, and can designate a preferred pickup location.
Next WorldCat Discovery enhancements

- Shareable public lists
- Shelving location facet
- CONTENTdm thumbnails in search results
- Option to cite theses and dissertations
- Focus on new fulfillment workflow to help people more easily GET the items they DISCOVER

Details on WorldCat Discovery Roadmap:
https://www.oclc.org/community/discovery/roadmap.en.html

Ongoing enhancements continue to reflect strategic testing and extensive community input. Watch for communication from us about the timing for these upcoming changes.
Moving to Collection Manager...
Collection types

✓ WorldCat knowledge base collections
✓ WorldCat cataloging partner collections
✓ WorldCat data sync collections
✓ WorldCat query collections
✓ WorldCat updates

...there are several different types of collections available for your use. These include knowledge base, cataloging partner, data sync, query, and updates.
There is a lot of interest in URIs in MARC records to help facilitate a transition to linked data. You’ve likely seen many URIs while cataloging. In addition to the MARC update we are working on for the future, the Record Manager team has been adding functionality to Record Manager so that you are able to export URIs within individual MARC records.
WorldShare® Record Manager

✓ Create and enrich WorldCat bibliographic records
✓ Set and delete WorldCat holdings
✓ Access the online save file from both Record Manager and Connexion
✓ Search and display Dutch, German, Library of Congress, Maori, and MeSH authority records
✓ Incorporate auto-suggest FAST and Library of Congress/NACO Names

Learn more at oc.lc/getRM
Exporting URIs

- Receive Uniform Resource Identifiers when exporting from Record Manager and Collection Manager
- Embeds URIs in the exported record(s) when the heading is controlled and a single subfield $0 is available for that heading

The controlling function with the bibliographic record masks the information that makes up the URI, so when you export a record, you only have the textual heading and no URI. Record Manager, however, now lets you export URIs for certain vocabularies from individual MARC records. The URI is embedded in a subfield $0 after the subject string or name heading. The URI only appears when the heading is controlled and a single subfield $0 is available for that heading.

To clarify a little further, some LCSH uses free-floating subdivisions and those subjects do not have an authority record for the entire subject string. In this case, even though the heading is controlled, the make-up of that string results in two or more URIs. We aren’t supplying multiple URIs at this time.
Under the Record Manager Preferences button, you can set which URIs you would like to receive on export.

So what does this look like? Starting last December, Record Manager users could set their preferences to export URIs. You can see from this example that I would like URIs from the LC/NACO Authority File and LCSH. Other options include the German GND authorities and MeSH.
Under the Collection Manager Settings button, you can set which URIs you would like to receive on export.

In the spring, Collection Manager users gained the ability to export URIs in their bulk downloads of records.
Exporting URIs

In the Record Manager record, check that the headings are controlled.

After export, controlled headings will include the subfield $0 URI.

After the settings are saved, you can pull up any bibliographic record. Check that the headings you would like URIs for are controlled.

Record Manager works the same as in Connexion for controlling NACO and LCSH. The blue link to show it’s controlled. These headings will have the subfield $0 once they are exported.
Normal export either in Connexion or Record Manager looks like this. You only get the text strings.
Using Record Manager, you can include the URIs in the export.

In this record, each heading was controlled and the entire subject string could be represented with a subfield $0 URI.
NACO functionality (coming soon)

- Creating new authority records
  - Templates for RDA for names, series, geographics
  - Derive from existing authority record
- Editing/replacing existing records
- Duplicate detection (upon creating/replacing)
- Locking in save file (online only, no local save file)
- Uses WorldShare roles and permissions instead of Connexion authorizations

As of 6/15, slated for install on 7/28.
Fabulous new feature!!

After creating or editing an authority record, the institution can…

…make additional changes after submitting the record…

…before it’s sent in the daily contribution file
Vocabularies in Record Manager

What’s available or coming

- LC/NACO, NAF and LCSH
- Medical Subject Headings (MeSH)
- Māori Subject Headings
- Dutch Names Authority file (NTA Names)
- Integrated Authority File (GND)
- Canadiana Names *(coming soon)*
- Canadiana Subjects *(coming later)*

Note: look-up services and controlling functionality will only be available in Record Manager for new vocabularies.
Next up, the WorldCat bibliographic database and cataloging issues.
It is easy for us to take WorldCat for granted....but it is a truly unique asset to the library community and one of the few assets that continues to grow in scale and in richness every year.

Did you know that two new records are added to WorldCat every second?

*Notes current as of 5/8/2018*
In terms of languages, this means that WorldCat has increased its support from 15 scripts to more than 135 that represent 6,500+ languages around the world.

This positions the number of languages represented in WorldCat (currently 491) to grow exponentially!

**ADDITIONAL DETAILS IF NEEDED:**

**Language Count as of Apr 2018**

1. English 161,411,028  
2. German 52,334,614  
3. Undetermined 38,746,132  
4. French 38,100,124  
5. Spanish 20,337,025  
6. Chinese 13,116,809  
7. Japanese 12,612,541  
8. Italian 10,533,362  
9. Dutch 7,093,015
<table>
<thead>
<tr>
<th>Rank</th>
<th>Language</th>
<th>Number of Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Russian</td>
<td>6,271,496</td>
</tr>
<tr>
<td>11</td>
<td>Polish</td>
<td>6,266,102</td>
</tr>
<tr>
<td>12</td>
<td>No linguistic content</td>
<td>6,066,166</td>
</tr>
<tr>
<td>13</td>
<td>Latin</td>
<td>5,419,789</td>
</tr>
<tr>
<td>14</td>
<td>Swedish</td>
<td>4,863,714</td>
</tr>
<tr>
<td>15</td>
<td>Danish</td>
<td>3,706,047</td>
</tr>
<tr>
<td>16</td>
<td>Portuguese</td>
<td>2,906,621</td>
</tr>
<tr>
<td>17</td>
<td>Hebrew</td>
<td>2,490,515</td>
</tr>
<tr>
<td>18</td>
<td>Arabic</td>
<td>2,413,083</td>
</tr>
<tr>
<td>19</td>
<td>Slovenian</td>
<td>2,328,246</td>
</tr>
<tr>
<td>20</td>
<td>Czech</td>
<td>1,733,667</td>
</tr>
<tr>
<td>21</td>
<td>Thai</td>
<td>1,391,217</td>
</tr>
<tr>
<td>22</td>
<td>Korean</td>
<td>1,381,278</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(previously spot 24)</td>
</tr>
<tr>
<td>23</td>
<td>Hungarian</td>
<td>1,292,499</td>
</tr>
<tr>
<td>24</td>
<td>Catalan</td>
<td>1,124,682</td>
</tr>
<tr>
<td>25</td>
<td>Finnish</td>
<td>1,019,321</td>
</tr>
<tr>
<td>26</td>
<td>Romanian</td>
<td>809,582</td>
</tr>
<tr>
<td>27</td>
<td>Turkish</td>
<td>801,575</td>
</tr>
<tr>
<td>28</td>
<td>Indonesian</td>
<td>723,443</td>
</tr>
<tr>
<td>29</td>
<td>Multiple languages</td>
<td>646,086</td>
</tr>
<tr>
<td>30</td>
<td>Norwegian</td>
<td>624,598</td>
</tr>
<tr>
<td>31</td>
<td>Greek, Modern (1453-)</td>
<td>571,425</td>
</tr>
<tr>
<td>32</td>
<td>Croatian</td>
<td>557,714</td>
</tr>
<tr>
<td>33</td>
<td>Malay</td>
<td>513,812</td>
</tr>
<tr>
<td>34</td>
<td>Ukrainian</td>
<td>484,082</td>
</tr>
<tr>
<td>35</td>
<td>Serbian</td>
<td>442,296</td>
</tr>
<tr>
<td>36</td>
<td>Persian</td>
<td>414,180</td>
</tr>
<tr>
<td>37</td>
<td>Vietnamese</td>
<td>360,557</td>
</tr>
<tr>
<td>38</td>
<td>Afrikaans</td>
<td>345,736</td>
</tr>
<tr>
<td>39</td>
<td>Bulgarian</td>
<td>300,477</td>
</tr>
<tr>
<td>40</td>
<td>Hindi</td>
<td>297,775</td>
</tr>
<tr>
<td>41</td>
<td>Yiddish</td>
<td>292,580</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(previously spot 42)</td>
</tr>
<tr>
<td>42</td>
<td>Tamil</td>
<td>276,627</td>
</tr>
<tr>
<td>43</td>
<td>Greek, Ancient (to 1453)</td>
<td>251,987</td>
</tr>
<tr>
<td>44</td>
<td>Frisian</td>
<td>220,582</td>
</tr>
<tr>
<td>45</td>
<td>Urdu</td>
<td>217,392</td>
</tr>
<tr>
<td>46</td>
<td>Slovak</td>
<td>179,647</td>
</tr>
<tr>
<td>47</td>
<td>Norwegian (Bokmål)</td>
<td>178,007 (New to Top 50)</td>
</tr>
<tr>
<td>48</td>
<td>Basque</td>
<td>170,088</td>
</tr>
<tr>
<td>49</td>
<td>Lithuanian</td>
<td>147,830</td>
</tr>
<tr>
<td>50</td>
<td>Bengali</td>
<td>133,811</td>
</tr>
</tbody>
</table>
[Notes updated 5/8/2018]
There are 39 non-Latin scripts throughout WorldCat with CJK taking the largest share with 23 million records containing the respective scripts. Number of records for all 39 scripts are available in the speaker notes.

<table>
<thead>
<tr>
<th>Non-Latin Script</th>
<th>Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJK</td>
<td>23,929,308</td>
</tr>
<tr>
<td>Hebrew</td>
<td>2,563,105</td>
</tr>
<tr>
<td>Arabic</td>
<td>1,739,758</td>
</tr>
<tr>
<td>Cyrillic</td>
<td>1,260,598</td>
</tr>
<tr>
<td>Thai</td>
<td>1,084,481</td>
</tr>
<tr>
<td>Greek</td>
<td>194,246</td>
</tr>
<tr>
<td>Tamil</td>
<td>67,039</td>
</tr>
<tr>
<td>Symbols</td>
<td>47,646</td>
</tr>
<tr>
<td>Devanagari (Nagari)</td>
<td>22,838</td>
</tr>
<tr>
<td>Armenian</td>
<td>11,639</td>
</tr>
<tr>
<td>Bengali</td>
<td>3,770</td>
</tr>
<tr>
<td>Ethiopic (Ge'ez)</td>
<td>1,597</td>
</tr>
<tr>
<td>Gujarati</td>
<td>1,081</td>
</tr>
<tr>
<td>Georgian (Mkhedruli)</td>
<td>641</td>
</tr>
<tr>
<td>Language</td>
<td>Code</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Kannada</td>
<td>538</td>
</tr>
<tr>
<td>Telugu</td>
<td>525</td>
</tr>
<tr>
<td>Gurmukhi</td>
<td>505</td>
</tr>
<tr>
<td>Bopomofo</td>
<td>223</td>
</tr>
<tr>
<td>Malayalam</td>
<td>200</td>
</tr>
<tr>
<td>Syriac</td>
<td>135</td>
</tr>
<tr>
<td>Myanmar (Burmese)</td>
<td>71</td>
</tr>
<tr>
<td>N’Ko</td>
<td>71</td>
</tr>
<tr>
<td>Tibetan</td>
<td>48</td>
</tr>
<tr>
<td>Khmer</td>
<td>35</td>
</tr>
<tr>
<td>Coptic</td>
<td>29</td>
</tr>
<tr>
<td>Unified Canadian Aboriginal Syllabics</td>
<td>23</td>
</tr>
<tr>
<td>Lao</td>
<td>10</td>
</tr>
<tr>
<td>Mongolian</td>
<td>10</td>
</tr>
<tr>
<td>Tifinagh (Berber)</td>
<td>5</td>
</tr>
<tr>
<td>Balinese</td>
<td>3</td>
</tr>
<tr>
<td>Sinhala</td>
<td>3</td>
</tr>
<tr>
<td>Yi</td>
<td>3</td>
</tr>
<tr>
<td>Cherokee</td>
<td>2</td>
</tr>
<tr>
<td>Kayah Li</td>
<td>2</td>
</tr>
<tr>
<td>Tai Tham (Lanna)</td>
<td>2</td>
</tr>
<tr>
<td>Lisu (Fraser)</td>
<td>2</td>
</tr>
<tr>
<td>Oriya</td>
<td>2</td>
</tr>
<tr>
<td>Vai</td>
<td>2</td>
</tr>
<tr>
<td>Batak</td>
<td>1</td>
</tr>
</tbody>
</table>

Total 30,930,197
Moving on to specific cataloging activities and announcements...
I want to first talk about a data project Metadata Quality is actively working on, the conversion of 260 fields to 264 fields in RDA cataloging records.

12+ million records are coded RDA and approximately 1.5 million, or 12.5%, lack a 264 field. You may remember that the MARC update establishing the 264 field with the different second indicators for type of publication and copyright date went into affect after the start of RDA. These 1.5 million records are mostly from that time period.

To improve the quality and consistency of these record, Metadata Quality decided to start a project to update these. In the fall of 2017, a survey went out to the PCC email list to give an overview of the plan and gather feedback. Most of the concerns that were raised were about complex production/publication situations. We expect there will need to be some manual fixes after we run automated clean-up.
Here is an example of the majority of changes that will be made. Changes will not be made programatically for ambiguous or complex data. Here, two 264 fields were added because we could determine the publication date and the copyright date in the original 260.
Duplicate record clean-up

WorldCat Metadata Quality automated & staff activities

- Automated merging – Duplicate Detection & Resolution (DDR)
  - Continuously running for data sync
  - Can be set to target sets for records
  - Duplicates removed May 2018 – 1,081,157 from 14,899,765 records processed
- Manual merging
  - Merge requests – average of 2,500 per month
  - Merge transactions in May 2018 – 62,004 of 128,949 records

We continually get questions regarding duplicate records and what we are doing to minimize them. There are several ways that we go about this. The first way, with the biggest impact, is the Duplicate Detection & Resolution, or DDR, process that continually runs on incoming records. Just last month, DDR alone removed almost 425 thousand records from the 7.7 million that it looked at. We often will run targeted sets through DDR as well to remove additional duplicates.

DDR, of course, leaves some duplicates behind, so we still have some manual clean-up. This is where the Metadata Quality staff who focus on the bibliographic database work their magic. This team receives on average, 25 hundred merge requests per month. In May, the group completed almost 62 thousand merge transactions. Keep in mind that one tractions can affect any number of records.
Since we receive a high volume of merge requests, you will understand that we have a little backlog for merging. Currently the backlog contains about 25 thousand requests, with roughly 2/3s for book records.
Backlog of non-book duplicate requests

Here is a break down of the non-book formats with sound recordings, continuing resources, and scores taking the biggest chunks.
Member Merge Project

Train members to manually merge bib records

• 2 cohorts have gone through the training since 2015
• 3rd cohort being formed, training starts in August
• As of May 2018, merge transactions = 20,612
• We provide documentation and review
• Training starts with books, moves to additional formats

One of the ways we have used to help reduce the backlog is by training members to merge records instead of reporting them. Who better to merge when they have the resource in hand and have done extensive searching of WorldCat to find the best record?

Since training began in 2015, there have been 2 cohorts that have conducted over 20 thousand transactions for merging. Each cohort had members from several institutions.

Members are provided training and documentation, as well as a review all of the proposed merge transactions. Members start with learning how to merge book records and then have the option to learn additional formats if interested.
You can see that since starting this program, the participants have been actively merging records. We are only 7 months into FY18 and already have close to 4,000 transactions. This is the first year where the two cohorts are working on merging. Right now all of the institutions are independent on books and several have additional independence in other formats.
“It’s a power trip!”

Casey Mullin
Head of Cataloging and Metadata Services, Western Washington University Libraries, Member Merge Project Cohort 2

“Merging bibliographic records is the most fun you can have without laughing.”

Jay Weitz
Senior Consulting Database Specialist, Metadata Quality, OCLC

Some fun comments about the merging! 😊

Casey Mullin who has been participating in the second cohort, gave an unsolicited review at ALA Midwinter in February. Jay Weitz, who many of you know, also provided his feedback during this session. Casey also went on to say that he learned quite a bit about how WorldCat is put together and really enjoyed learning the nuances. Many of the internal OCLC staff find merging records to be one of their favorite work things to do.
Moving away from merging, just a general comment that we are now under a continuous revision of Bibliographic Formats and Standards. Most of the chapters for guidance on specific MARC fields have been updated. Chapter 5 on Quality Assurance was just revised and released. The team is now working on chapter 3, Special Cataloging Guidelines.

If you are unsure as to whether or not a page has been updated, please checkout the revised date at the bottom of the page.
OCLC MARC Update

Update no. 25, Dec 2017 and no. 26, Apr 2018

- Addition of subfield $1 = Real World Object URI to over 90 bibliographic fields
- Addition of field 758 = Resource Identifier such as the FRBR works, expressions, manifestation, and items
  - Working with PCC on implementation
- Changes to 257, 382, 730

LC released the no. 25 update to MARC last December. The update allows for an increase in URIs in the MARC record and a place to add a Work id.

Many fields will see the addition of subfield $1, for the Real World Object URI. The 758 has also been added for an identifier for a resource that is either the resource described in the bibliographic record or a resource to which it is related. Resources thus identified may include, but are not limited to, FRBR works, expressions, manifestations, and items.

As we work internally to implement these new fields, we are also working closely with the PCC on best practices. Stay tuned for more information on this.

Full OCLC-MARC Update 2018

The 2018 OCLC-MARC Update will implement MARC 21 Bibliographic and Holdings format changes announced in MARC 21 Updates No. 25 (December 2017) and No. 26 (April 2018) including:

In the Bibliographic and Holdings field 007 for Maps, a new code “x” (Not Applicable) is defined in subfield $e (007/04, Physical Medium) for remote digital
resources. Documentation-only changes have also been made in subfields $b (007/01, Specific Material Designation) and $g (007/06, Production/Reproduction Details).

Bibliographic field 257 (Country of Producing Entity) has had its scope broadened to include areas not legally recognized as countries.

Bibliographic field 382 (Medium of Performance) has had its subfield $r (Total Number of Individuals Performing Alongside Ensembles) redescribed.

Bibliographic field 730 (Added Entry – Uniform Title) has had subfield $4 (Relationship) added.

New Bibliographic field 758 (Resource Identifier) has been defined.

New subfields $t (Report Number), $u (Standard Technical Report Number), and $z (ISBN) have been added to the Bibliographic field 777 (Issued With Entry).

Subfield $d (Date of Meeting or Treaty Signing) has been made Repeatable in all Bibliographic Meeting Name fields.

Subfield $s (Version) has been made Repeatable in 31 Bibliographic fields.

Bibliographic subfield $0 (Authority Record Control Number or Standard Number) has been slightly redefined.

Subfield $1 (Real World Object URI) has been defined in over 90 Bibliographic fields and four Holdings fields.

Subfield $3 (Materials Specified) has been added to Bibliographic fields 377 (Associated Language), 380 (Form of Work), 381 (Other Distinguishing Characteristics of Work or Expression), 383 (Numeric Designation of Musical Work), and 384 (Key).

OCLC will also validate MARC codes announced in eight Library of Congress Technical Notices (http://www.loc.gov/marc/marcinf.html#naa) issued since November 2017.

This OCLC-MARC Update will also:

Invalidate the OCLC-defined Encoding Level (Leader/17) value “L”.

More strongly enforce the mandatory input standard for the presence of field 040 subfield $b for Language of Cataloging in most bibliographic records.

Invalidate First Indicator values 7, 8, and 9 in Bibliographic field 243 (Collective Uniform Title).

Invalidate the OCLC-defined Second Indicator value “8” (Sears List of Subject Headings) in all applicable 6XX fields.

All details will be available in an upcoming OCLC Technical Bulletin. We plan to install the OCLC-MARC Update 2018 during the second half of calendar year 2018 and will make announcements widely through the usual discussion lists and Connexion logon greetings.

Elements from the MARC 21 Authority Format Updates No. 25 and No. 26 will be implemented not at this time but instead at a future date in coordination with the Library of Congress and the Name Authority Cooperative (NACO) of the Program for
Cooperative Cataloging (PCC). LC, NACO, and OCLC will make announcements at that future date.
OCLC MARC Update

Changes to OCLC MARC

• Invalidate the OCLC-defined Encoding Level (Leader/17) value “L”.
• More strongly enforce the mandatory input standard for the presence of field 040 subfield $b$ for Language of Cataloging in most bibliographic records.
• Invalidate First Indicator values 7, 8, and 9 in Bibliographic field 243 (Collective Uniform Title).
• Invalidate the OCLC-defined Second Indicator value “8” (Sears List of Subject Headings) in all applicable 6XX fields.
Encoding levels

Elimination of OCLC-defined alphabetic encodings levels in favor of standard MARC 21 codes

- Continues effort to move to standard MARC 21 coding and eliminate OCLC-defined elements
- Simplifies documentation and training
- At least 6 months advance notice to members
- How will this affect your workflows?
Encoding levels

Five OCLC-defined codes

- I = full level – duplicates MARC 21 code blank
- K = minimal level – duplicates MARC 21 code 7
- J = deleted record – loss of encoding level info
- M = record loaded via data sync collections - loss of encoding level info
- L (already eliminated from database)
FAST (Faceted Application of Subject Terminology) -
https://www.oclc.org/research/themes/data-science/fast.html

FAST (Faceted Application of Subject Terminology) is derived from the Library of Congress Subject Headings (LCSH). The broad purpose of adapting the LCSH with a simplified syntax to create FAST is to retain the very rich vocabulary of LCSH while designed to support faceted retrieval. The schema maintains upward compatibility with LCSH, and any valid set of LC subject headings can be converted to FAST headings.
FAST

Productionization

• Move from experimental research systems to production systems
• Create sustainable governance structure
• For the past 6 months, have gathered input from FAST users
• Currently intalks with OCLC development staff
FAST

Needs and desires

- Ability to request/add new and alternative terms
- Long-term support with continued access, preferably open
- Continued ease of use, including maintained and improved tools
- Regular updates, including better content and coverage
- Batch conversion, APIs, and integration into cataloging tools
- Additional usage guidance
- Clearly defined relationship with LCSH
FAST

Community engagement and governance

- **Editorial** – approving new terms, establishing term relationships, mapping requirements, etc.

- **Operational** – application guidelines, training, documentation, guidance, etc.

- **Governance** – recommend directions and goals, inform strategic directions, establish working groups, etc.

- **Communications** – outreach, promotion, marketing, talking points, working with the community, etc.
No other library organization has created and shared more linked data than OCLC

So what has OCLC been doing for the last 10+ years with respect to linked data?

Our strategic approach is threefold:
• Build on the foundation of WorldCat.
• Collaborate and partner with community leaders.
• Make iterative progress rather than wait for perfection.

We know that linked data brings new contextual relationships to WorldCat with opportunities to modernize and improve both library and patron workflows. And it also positions our community to be hubs in the global information network beyond libraries.
Like other linked data pilot projects we’ve conducted with member libraries, exploring problems that linked data can solve is not new. This particular project has the goal of providing services for continued improvement of library workflows. Looking at the current environment, we have workflows for: searching, copy cataloging, original cataloging, and authorities.

In the future, we see amplified searching, adding relationships, entity management, and library-sourced vocabularies.
Since last fall, we’ve been working on a Wikibase prototype to describe library resources. For the initial phase for this project, we worked with two library partners. We also created a cross-functional team within OCLC to ensure sustainability as applicable. Members from research, IT, and product have all worked together to make this prototype a reality.

The partners and OCLC staff used the following editor to create, share and edit entity descriptions. They could also contribute additional contextual relationships between entities, beyond those that can be found by mining structured data in bibliographic and authority records. A second goal for the project was to provide an infrastructure to reconcile names for people, organizations, concepts, places, and events against an index based on entities, returning language-tagged headings and persistent identifiers.
Key technologies

For this, we used two out-of-the-box technologies: MediaWiki and the MediaWiki extension, Wikibase.

*MediaWiki logo: https://www.mediawiki.org/wiki/Promote_MediaWiki
Wikibase logo: https://commons.wikimedia.org/wiki/File:Wikibase_logo.png*
The Wikibase search and display/API is used for a view of an entity. An entity's page includes labels, statements, and identifiers. [click] The search box, shown here in the lower right hand corner, provides additional detail to help a searcher choose the needed entity.

Once this was functional, the goal was for the partner libraries to tell us what was lacking and what they needed to move forward. We found that for library workflows, there are certainly improvements that would facilitate the reconciliation process.

One noted improvement, especially when coming from a bibliographic-centric view, is refinements to auto-suggest. Auto-suggest will only match names entered into the entity description, but not other forms. A cataloger used to searching in inverted order would only see results in the auto-suggest if the inverted form of the name was added as a label or alias [click]. Retrieving Davis comma Ann in the auto-suggest box is only possible because that form is included in the alias.
It’s also limiting matches to the default label and “also known as” aliases for your currently selected language [click], and limiting matches to labels and aliases that begin with your search terms. [click] A search on Bradford will not retrieve Ann’s entity.

A reminder that this is only for the auto-suggest feature. If you click on “containing” at the bottom of the search suggestions, you’ll get all of the results that use Bradford and will fine Ann’s entity.
We also create a separate search interface, a part from the editor, that pulls in entity information and data from other sources. It is also much more user friendly when it comes to searching. The dbpedia information [click] is being pulled in to enhance the description for Ann, including her picture. It also displays the statements and identifiers from our entity interface, those you see here on the right.
Phase II library partners

1. American University
2. Brigham Young University
3. Cleveland Public Library
4. Cornell University
5. Harvard University
6. Michigan State University
7. National Library of Medicine
8. North Carolina State University
9. Northwestern
10. Princeton
11. Smithsonian Institution
12. Temple University
13. University of California, Davis
14. University of Minnesota
15. University of New Hampshire
16. Yale University

In Phase II, we greatly expanded the number of participants. Our goals were:

• to better understand large entity databases;
• to identify the sticky problems around entities that will require additional research;
and,
• to gain an improved understanding of community needs for entity management especially managing local authorities.

Phase II is wrapping now and at ALA Annual we will be giving progress and status reports.
Phase III objectives

- Phase III partners being finalized
- Top enhancements requests by library partners
  - Improve indexing
  - Batchload new entities provided by partner libraries
  - Require reference sources for statements
  - Offer property-based constraints for editing
  - Search by non-prototype identifiers (in the UI)
- Partner request for the project’s focus
  - Describing cultural objects and their digital representations

We will soon be starting phase III and the list of partners is being finalized. Based on the outcomes of phase II, our goal is to work on the top enhancement requests: improved indexing, batch-loading of new entities, require reference sources for statements, offer property-based constrains when editing, and searching by non-prototype identifiers. The partners have also requested a focus on describing cultural objects and their digital representations.
If you are interested in more information, including contact information, check out the linked data Wikibase prototype website under OCLC Research. For general information on OCLC and linked data, check out oc.lc/linkeddatasummary.
Finally, ways of connecting with the community.
Virtual AskQC Office Hours

**Discuss** cataloging policy with OCLC metadata experts and **receive updates** on current WorldCat metadata quality projects.

Next session: Wednesday, June 27 at 1 PM Eastern
See oc.lc/askqc for more info and links to previous sessions

Something we are trying during the first half of the year are office hours for AskQC. AskQC@oclc.org is a longstanding email address to which catalogers can send questions to OCLC Metadata Quality staff about cataloging policies, standards, and practices. Each office hour starts with an update on a cataloging topic and is followed by a question and answer period. Feedback thus far has been positive and we are working on getting the recording links and notes posted online.

We plan to do this through June and then review the outcomes to see what will go forward.

**Topics**

January: 260 to 264 conversion
February: cataloging defensively with the edition statement to prevent accidental merging
March: an overview of processing change requests, how Metadata Quality staff sort, analyze, categorize, and handle error reports
April: URLs in a shared cataloging, the importance of the 856 2nd indicator and when to use a 956
May: validation in WorldCat cataloging
June: overview of the Expert Community
The OCLC Community Center is the place for making community connections, sharing best practices, staying up to date on product releases and contributing ideas.

- Currently twelve (12) product-based communities – visit oclc.org/community
- Collaboration with peers to improve products
- Unified access to ALL WorldShare applications

[Notes current as of 5/8/18]
Another way to actively engage with the cooperative is through our social and digital channels.

- If you’re not doing so already, we encourage you to follow OCLC on these social channels and share our content with your personal and professional networks.

[Notes current as of 5/4/18]
OCLC Next, our blog, shares our work and engages with members. Contributors include OCLC leaders and industry experts. You can subscribe to the Next blog and receive weekly updates direct to your inbox.

In the post you see here, I’ve written about the 1 billionth OCN, which incidentally was for a digital image from the Chiba University Library (YA@) in Chiba, Japan.
The 2018 OCLC Americas Regional Council Meeting will be held on October 25–26 in Chicago, Illinois, at The Drake hotel.

In addition to the main conference agenda, we will host a pre-conference meeting for the OCLC Research Library Partnership on Wednesday, October 24. Registration is open to all in the OCLC Research Library Partnership and institutions are welcome to send more than one delegate.

Early bird registration for ARC 18 is now open and the Call for Participation opened on May 31.

[Notes updated 5/23/18]
Thank you for your kind attention. I’m happy to take questions.