

March 13, 2019

Virtual AskQC Office Hours

Introduction to VIAF

OCLC Metadata Quality



Housekeeping

- After the session you will be directed to a quick, optional survey

Virtual AskQC Office Hours
feedback survey

Please take a moment to provide feedback on today's office hour session. The responses are for informational purposes only and optional. Thank you for attending today's session!

1. Did you find today's session useful?

Yes
 No
 Sort of

2. Why did you choose the answer you did in question 1?

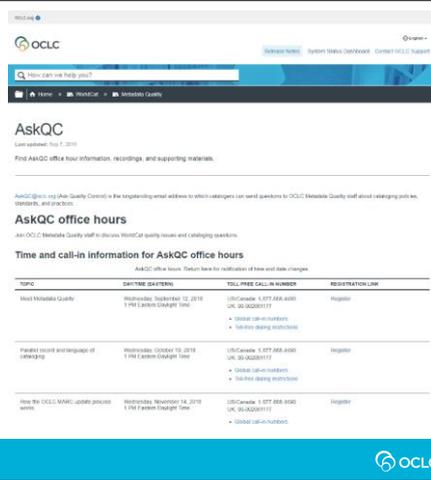
Enter your answer

3. Are there topics you would like us to cover in the future?

Enter your answer

Housekeeping

- After the session you will be directed to a quick, optional survey
- All session recordings, slides, and notes are available at oc.lc/askqc



The screenshot shows the AskQC website interface. At the top, there is a search bar with the text "How can we help you?". Below the search bar, the "AskQC" logo is visible, along with navigation links for "Home", "Product", and "Services Guide". The main content area is titled "AskQC" and includes a sub-header "Find AskQC office hour information, recordings, and supporting materials." Below this, there is a section titled "AskQC office hours" with a sub-header "Time and call-in information for AskQC office hours". A table provides details for three office hours sessions, including the date, time, call-in number, and registration link.

TOPIC	DATE/TIME (PST/PDT)	TOLL-FREE CALL-IN NUMBER	REGISTRATION LINK
Hour: Volunteer Queue	Wednesday, November 12, 2019 1 PM Eastern Standard Time	1-800-858-8850 URL: oc.lc/askqc/117 • Global call-in numbers • Toll-free dialing instructions	Register
Hour: Record and Manage IP Addressing	Wednesday, November 13, 2019 1 PM Eastern Standard Time	1-800-858-8850 URL: oc.lc/askqc/117 • Global call-in numbers • Toll-free dialing instructions	Register
How To: OCLC MARC update process	Wednesday, November 14, 2019 1 PM Eastern Standard Time	1-800-858-8850 URL: oc.lc/askqc/117 • Global call-in numbers	Register

On the call today



Laura Ramsey
Section Manager,
WorldCat Quality



Hayley Moreno
Database
Specialist II



Robin Six
Database
Specialist II



Jay Weitz
Senior Consulting
Database
Specialist



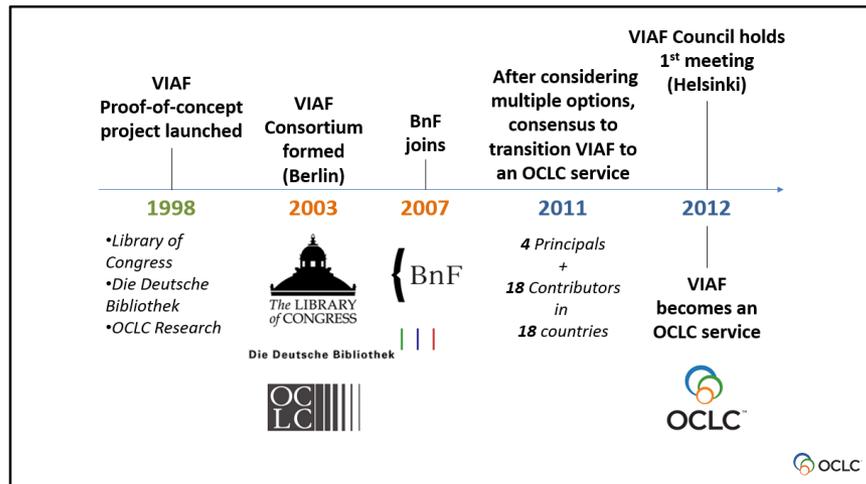
Shanna Griffith
Database
Specialist II



Robert Bremer
Senior Consulting
Database
Specialist

VIRTUAL INTERNATIONAL AUTHORITY FILE (VIAF)





In the 1990s, there was discussion about the possibility of having a common international authority file between the Library of Congress (LC) and Die Deutsche Bibliothek (DNB) but after this idea was unable to take off the libraries decide to take a different approach by shifting the idea of linking their authority files. Linking would present all the benefits of a common file without requiring a large investment of time and expense in creating a new file.

So in April 1998 LC, GND, and OCLC embarked on a proof-of-concept project to test linking each other’s authority records for personal names. The project was a success and in August 6, 2003 the VIAF Consortium was formed after LC, DNB, and OCLC signed an agreement during the International Federation of Library Associations and Institutions (IFLA) conference in Berlin, Germany. A few years later the National Library of France (Bibliothèque nationale de France, or BnF) joined the consortium with an agreement effective October 5, 2007.

In 2011 the four organizations (LC, GND, OCLC, and BnF) assumed the role of Principals in the consortium, having joint responsibility for VIAF with OCLC hosting VIAF and supplying the software, and the participating institutions

supplying the content (in this case authority and bibliographic data). Additional organizations later joined the consortium as Contributors, providing source files and expertise to advance the state of VIAF.

With the successful proof of concept of VIAF, discussions begin in earnest among the Principals about a suitable long-term organizational arrangement for VIAF. After considering various options, the Principals and Contributors agreed in 2012 to transition VIAF to an OCLC service. This transition from an interim, shared-governance arrangement to OCLC having primary responsibility for maintenance of VIAF would allow the program to continue as a well-positioned, long-term, cooperative activity. OCLC would provide the appropriate infrastructure to respond to rising levels of traffic as VIAF gained momentum and popularity as a resource for library authority work and linked data activities.

VIAF Council

- Created in 2011/2012
- Self-governing advisory body
- Provides guidance on policies, practices and operation of VIAF



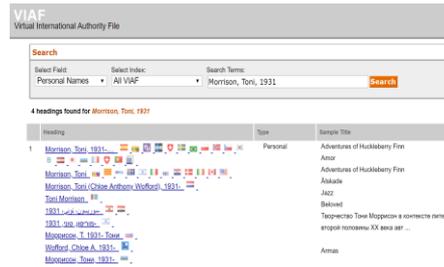
With VIAF now being an OCLC service we still wanted those contributing authority files to the program to help guide the future of VIAF. The VIAF Council was formed in 2011/2012 and is a self-governing advisory body, where membership consist of one voting representative for each contributor in the VIAF program. The 2018/2019 current Chair is Bill Leonard from Library and Archives Canada and Chair-Elect is Vincent Boulet from BnF. OCLC provides council a listserv, simple logistical support for meetings & appoints an OCLC Liaison (which is me!). VIAF Council provides guidance on the policies, practices, and operation of VIAF, including developing policies on the eligibility and selection of new [VIAF Contributors](#). The Council advises OCLC on VIAF's service features and supported data formats, and it promotes the effective and appropriate use of VIAF.



Now that we know the origin stories of VIAF. What is it?

What is VIAF anyway?

- An authority file
- An identifier
- Data hub



- VIAF is an authority file where identical records from different authority files are being linked together in a cluster. VIAF does not replace the source authority data, but creates something new built upon the relations mined from it. This gives VIAF users convenient access to the world's major name authority files. We have matched, link, and group our contributors authorities, allows researchers to identify names, locations, works, and expressions while preserving regional preferences for language, spelling, and script.
- Each VIAF cluster has an identifier.
- VIAF can be considered somewhat of a data hub as it is a collection of data from multiple sources (e.g., VIAF contributors (e.g., national libraries) and data providers (e.g., Wikidata and ISNI). VIAF organizes the data which then can be downloaded for anyone to use.

Who participates?

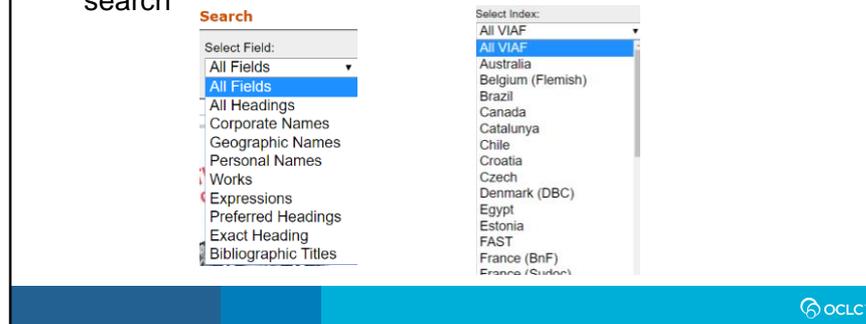
- National Libraries
 - National Library of Brazil
 - National Library of Lithuania
- Special Organizations
 - International Inventory of Musical Sources (RISM)
 - Union List of Artist Names (Getty Research Institute)
- Other data providers
 - ISNI
 - Syriac Reference Portal



VIAF is truly an international program with contributors coming from all continents of the world (with the exception of Antarctica). We have different types of sources that add their data. National libraries make up the bulk of our contributors with National Library of Brazil being over a year ago and the National Library of Lithuania who is currently in test. Potential contributors are places in test to make sure there is no issues with the data and as we completing their application package. Once they past a certain time period they become officially added to VIAF. Special organizations that have unique collections like RISM was recently add to VIAF and Union List of Artist Names have been in VIAF for quite some time. We finally have the category of other data providers like ISNI and the Syriac Reference Portal who do not have a representative in VIAF Council. We continue to add new participants to the VIAF program and you can always find out who is new by looking at the bottom of the page to see who is in test.

How to search?

- VIAF users can select the field and index they wish to search



- Narrowing down the field you wish to search helps in reducing the amount of results you retrieve in VIAF. This is imperative as VIAF does not provide any way for users reorder, or modify the results they have. As shown in the first image you can search for a particular name in the corporate, geographic, or personal index. You can also search for form of the name that appears at the authorized access point 1xx field by selecting preferred headings. All headings will retrieve any 1xx, 4xx, or 5xx in the authorities so your results would be broader.
- The index can also be narrowed down to whatever file you wish to only retrieve results from, whether that be Wikidata or LC/NACO. If the entity is not in the file you select to index, then no results are retrieved.

What does a VIAF webpage contain?

- VIAF ID
- Permalink
- Preferred forms of names from all authority files that contain the entity
- Alternate forms of names
- Related names (e.g., geographic places associated with an entity, pseudonyms)
- Works
- Co-authors
- Countries and regions of publication
- Publication statistics
- Selected publishers
- About section (e.g., occupation, language, external links to the entity)



The VIAF page for an entity can vary on the amount of information it has based on how much metadata VIAF contributors have on it. For example, the snippet of the VIAF page I showed in an earlier slide was for the author Toni Morrison's, which has a lot of rich information as she is a very notable person. There are several VIAF contributors who have her name in their authority files. There is an extensive list of works by the author due to VIAF contributors, not just sending us authorities, but also bibliographic records which help us in the matching process for VIAF. The VIAF page continues with co-authors who have written with Toni Morrison, where she has published, who those publishers are, publication statistics, and an about section (which is all coming from the authority and bibliographic files of VIAF contributors as well as other sites like Wikidata).

Due to the wealth of information that you can find in VIAF it's not surprising that it has become a research tool for library professionals who create authority records and add some of the information they find in VIAF to the description of the authority. Wikipedia editors also use VIAF to help enhance their description of an entity.

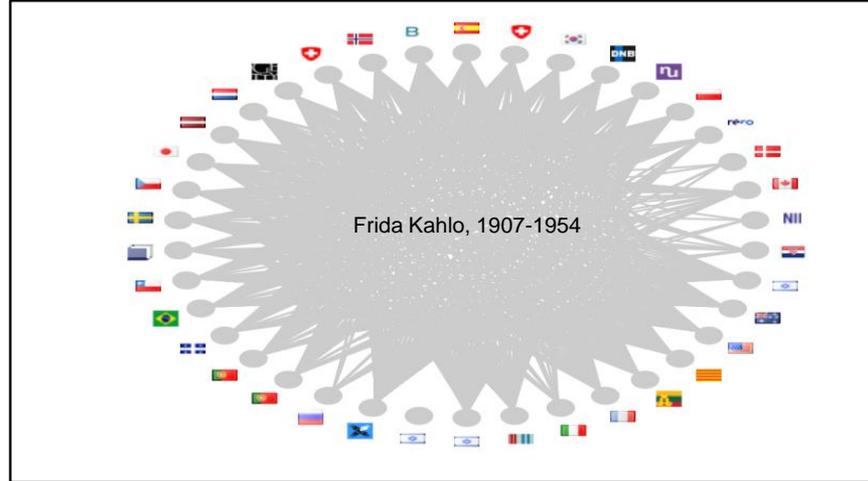
What are processed records?

- The authority records of VIAF sources found in the cluster

Leader	00000cz za2200037n 45 0
001	BNE XX896369 (VIAF cluster) Authority/Source Record
005	200107271056
008	911115[n] ajnnaabn [n aaa
003	BNE
035	1a (BNE XX896369
010	1a XX896369
016	1a BNE19913543421
040	1a SpMaBN 1b spa 1c SpMaBN 1e rdc 1f embne
100	1 0 1a Tan, Amy 1d 1952-
670	1a El Club de la Buena Estrella, 1990
670	1a CDMARC names, 1991 1b (Tan, Amy)
670	1a La esposa del dios del fuego, 1994 1b anteport. (n. 1952 en Oakland, California)
901	1a 978840808697 19 1



- It's important to note that the processed records are not an exact copy of the authority record sent to us from VIAF participants. Some fields may have been omitted or transformed to be integrated in the VIAF database. The original authority record from the contributor can be viewed if that contributor has provided us a link to their authority file. The image above provides an example of the National Library of Spain who includes authority/source record links to see the original authority when clicking on the blue hyperlink with the preferred form name. You can also view the actual authority record when clicking on the flag icon. You will note that the flags which are hyperlink have been underlined. In the second screenshot National Library of Spain, the GND, and the National Diet Library of Japan have links to their authority source records.
- The process records also contain several 9xx fields which are related to the algorithms that were run against the authority. They can be disregarded when looking at processed records.



Now that we the basics of what VIAF is and what it contains let's move on to one of its prized features. Clustering! The image in front of you is the actual VIAF cluster for the Mexican artist Frida Kahlo. As you can see this cluster is big showcasing how notable this person has been. Each icon you see represents an authority file that includes her. You can see the National Library of Lithuania who is currently in test have a hazard symbol on top of their flag icon showcasing they are in test. We also see some of the others organizations I listed as examples in the "Who contributes" slide such as National Library of Brazil, Union List of Artist Names , and ISNI.

Clustering

- Goals for matching
- Pair-wise matching
- Weekly matching and clustering builds

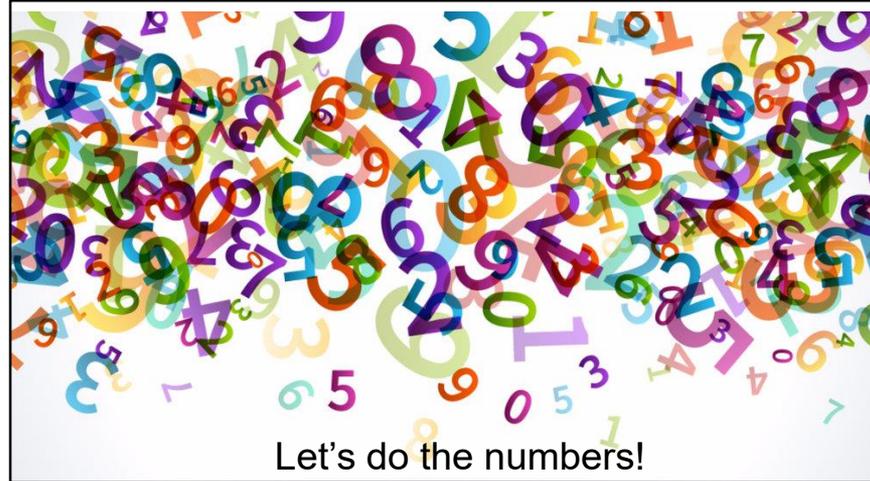


oclc

- VIAF's goal is that for any two source records in a cluster, that there is less than 1% chance that they describe two different entities, that is that the links is correct more than 99% of the time. Reaching this level of certainty means that some records will not be pulled into clusters because of limited metadata, rather than any indication that they do not match. An example of this limited metadata is when there is no title or date associated with a person in VIAF. With VIAF getting larger there is more confidence that a unique name, even if related information is missing, has a high change of being a useful match
- Pair-wise matching looks at different elements in the metadata to create links. With personal names, surnames are the primary way of bringing records together for matching. Once brought together evaluation is done to determine if the names are compatible such as dates, name/title similarity, name closeness, title closeness. These matching links are ranked and this ranking is used in determining if the entities are a match. Another matching point is when the explicit reference to another authority record is mentioned (e.g., 024 fields that have identifiers to other sources like ISNI) these types of links are considered suggested. VIAF does have a

manual override in case the matching algorithms cannot get something right. The Extended Authorities file (called xA) is used to manage matching in VIAF and causes forced links. Like all the other VIAF contributors the xA authority file is also harvested to help with the maintenance of clusters.

- Clusters can change after they are re-calculated each processing cycle (which is currently weekly). While the vast majority of the clusters are stable, there are splits or merges that occur based on changes in the records or VIAF algorithms. Recognizing that the clusters are fluid and that no individual member record can be permanently assigned to a single VIAF identifier, we do try to have these identifiers to be persistent as possible to allow for linking.



I thought it would be worthwhile to share some stats to give you a perspective on how much rich data VIAF has.

Stats

- 55 VIAF sources
 - Authorities: 73,846,672
 - Bibliographic: 163,730,371
- 33,877,285 total clusters
 - Personal 23,239,186
 - Corporate 5,677,664
 - Geographic 606,809
 - Uniform Title Expression 1,495,187
 - Uniform Title Work 2,858,439



Most represented languages in VIAF

Language*	Approximate number of authorities
English	21 million
German	9 million
French	6 million
Polish	3.1 million
Japanese	2.8 million
Dutch	2.7 million
Norwegian	1.9 million
Hebrew	1 million
Czech	876,000
Spanish	644,000

*Values reported in 040 \$b (MARC) or 100 \$a/9-11 (UNIMARC)



Just like many of the other OCLC sites that you use with cataloging such as Connexion or Record Manager that have mechanisms in place to report issues or errors, VIAF also has ways you can communicate with us to report problems.

Reporting issues in VIAF

Report to bibchange@oclc.org or use the “Send us a comment” on the VIAF website

- Combine clusters
- Pull apart a cluster
- VIAF webpage containing incorrect information about the entity

Report to oclcviaf@oclc.org

- VIAF website issues (e.g., display, broken links, coding)
- General issues in VIAF that do not relate to clusters or information about the entity



When dealing with clusters that need to be modified whether that be to combine duplicate entries or pull apart different entities in the same cluster those should be reported directly to bibchange@oclc.org or you are welcome to using the “Send us a comment” link on the VIAF website. Both places will make sure that someone from the WorldCat Metadata Quality team looks into the request and correct as needed. Also, if you find incorrect data on an entity found in VIAF those errors can also be reported to [bibchange](mailto:bibchange@oclc.org) email to send to us as a comment from the VIAF website.

For issues with the VIAF website itself, such as display, links being broken, or the coding of the site causing an issue, please email those to oclcviaf@oclc.org you can also send to this email any general issues or concerns with VIAF that do not deal with clusters or the VIAF webpage of an entity.

Resources

- General overview of the VIAF program: <https://www.oclc.org/en/viaf.html>
- General overview of managing ambiguity in VIAF: <http://www.dlib.org/dlib/july14/hickey/07hickey.html>
- Download the VIAF dataset: <http://viaf.org/viaf/data/>
- VIAF website: <http://viaf.org/>

If you are interested in learning more about VIAF I would highly recommend checking out the resources listed here.

Questions?



Laura Ramsey
Section Manager,
WorldCat Quality



Hayley Moreno
Database
Specialist II



Robin Six
Database
Specialist II



Jay Weitz
Senior Consulting
Database
Specialist



Shanna Griffith
Database
Specialist II



Robert Bremer
Senior Consulting
Database
Specialist

Please submit questions through chat



Thank you!

Send cataloging policy questions
at anytime to:
askqc@oclc.org

Session links available at:
oclc.org/askqc

Next Virtual AskQC Office Hours:

OCLC training resources
Wednesday, April 10, 2019
1:00 PM Eastern
Register at oclc.org/askqc

Because
what is
known must
be shared.®

