

To: VIAF Council

From: OCLC

Re: 2014 Annual Report to VIAF Council

Date: 13 August 2014 (revised 28 August 2014)

The VIAF™ (Virtual International Authority File) combines multiple name authority files into a single name authority service. It has become a significant resource for library authority work and a hub of import for authoritative linked data on the Web.

Since VIAF's transitionⁱ to become an OCLC service in early 2012, the number of agencies participating has grown from 19 agencies in 22 countries to 34 agencies in 29 countries (as of July 2014). Twenty-four (24) of the VIAF Contributors are national libraries, and an additional 11 national libraries provide data to VIAF through federal library agencies, consortia or other arrangements, bringing the total of national libraries represented in VIAF to 35 national libraries from 30 countries. (Note: Two additional national libraries from two additional countries have contracts pending – see below.)

Table of Contents

VIAF Contributors and Data Sources.....	2
VIAF Data Inputs and Outputs	4
VIAF Source authority records by subtype	4
VIAF Source data change over time	5
VIAF Data Sources Statistics Table.....	6
VIAF Utilization.....	9
VIAF Service Changes and Enhancements	12
VIAF Outreach	15
OCLC Staff-authored/co-authored publications:.....	15
OCLC Presentations:.....	15
Selected presentations and publications by others:.....	16
VIAF Council	16

VIAF Contributors and Data Sources

Thirty-four (34) agencies (including OCLC) have renewed or become new VIAF Contributors since 2012:

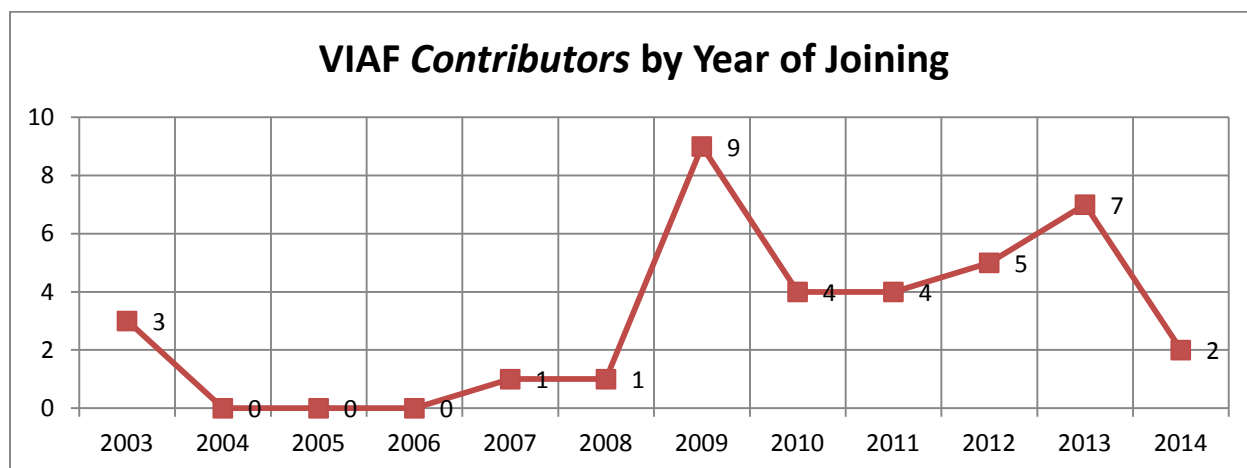


Figure 1: Agencies added to VIAF (note: 2014 data is to July)* (source: OCLC)

*This count includes the host agency, OCLC, and the following (current as of July 31, 2014):

Agency/agencies that have not entered into an updated contract:

- Országos Széchényi Könyvtár (National Széchényi Library)
- Российской государственной библиотеки (Russian State Library)

Agency/agencies with contracts pending:

- Perpustakaan Negara Malaysia (National Library of Malaysia)
- Biblioteca Nacional de Chile (National Library of Chile)

The following agencies have joined VIAF as Contributors since July 2013:

- **قبة تكملا** (Lebanese National Library)
- **Biblioteca de Catalunya** (National Library of Catalonia)
- **IZUM (Institut informacijskih znanosti)** (IZUM (Institute of Information Science))
- **Nacionalna i sveučilišna knjižnica u Zagrebu** (National and University Library in Zagreb)
- **Latvijas Nacionālā bibliotēka** (National Library of Latvia)
- **Biblioteka Narodowa** (National Library of Poland)
- **Bibliothèque nationale de Luxembourg** (National Library of Luxembourg)
- **Leabharlann Náisiúnta na hÉireann** (National Library of Ireland)

Additionally, national libraries from 5 countries have applied to VIAF for admission, and their applications are being processed. (Note: these applicants are *not* included in the counts on Figure 1.)

Other Data Providers: VIAF also imports and incorporates data from several sources beyond authority and bibliographic data provided by VIAF Contributors.

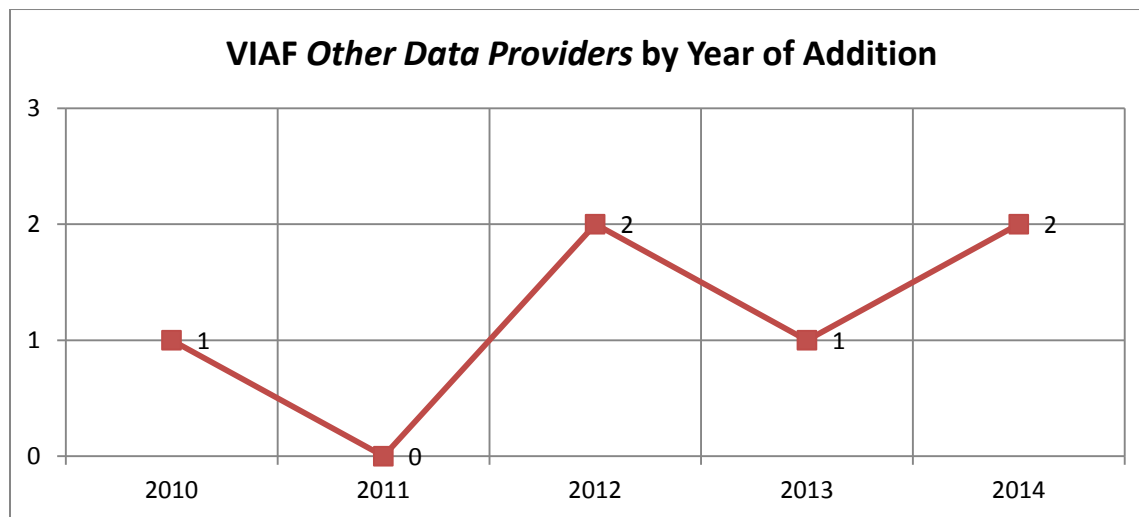


Figure 2: Other Data Providers/Sources added to VIAF (note: 2014 data is to July) (source: OCLC)

File	Initial Year of Use	VIAF Symbol
<i>xA (OCLC Research Extended Authority File)*</i>	2010	xA
<i>International Standard Name Identifier (ISNI)</i>	2012	ISNI
<i>Wikipedia</i>	2012	WKP
<i>Perseus Digital Library</i>	2013	PERSEUS
<i>Syriac Reference Portal</i>	2014	SRP
<i>xR (OCLC Research Authority File)*</i>	2014	xR

Table 1: Other Data Providers/Sources included in VIAF (June 2014) (source: OCLC)

*xA and xR are files developed and hosted by OCLC.

- xA provides a mechanism to allow the override of automatic clustering in cases where clustering is problematic based on the data available
- xR has been developed to support the addition of multilingual enrichment from WorldCat^{®1}

Other Data Providers include Wikipedia and ISNI which continue to enrich VIAF and vice-versa. Adjustments to the data flows among the parties continued during the FY14 period.

2013 also was the initial year for work -- as part of an OCLC Research effort, *Scholars' Contributions to VIAF²* -- to move from discussion and testing to routine inclusion of data from scholarly efforts in VIAF. Two projects, the **Perseus Digital Library**, and the **Syriac Reference Portal**, have helped establish both the degree of difficulty involved, and the measure of value that can arise, from linking authority data stewarded by libraries and other cultural heritage organizations with kindred specialist work products developed to support deep scholarship and research.

¹ For more information, see: <http://oclc.org/research/activities/multilingual-bib-structure.html>

² For more information, see: <http://www.oclc.org/research/activities/viaf-scholars.html>

VIAF Data Inputs and Outputs

The source data upon which VIAF is built has grown to ~45 million authority records and 105.5 million bibliographic records.

VIAF Source authority records by subtype

Personal: 35.1 million

Corporate: 5.4 million (includes conferences)

Geographic: 0.4 million (mostly jurisdictional)

Uniform titles: 3.9 million

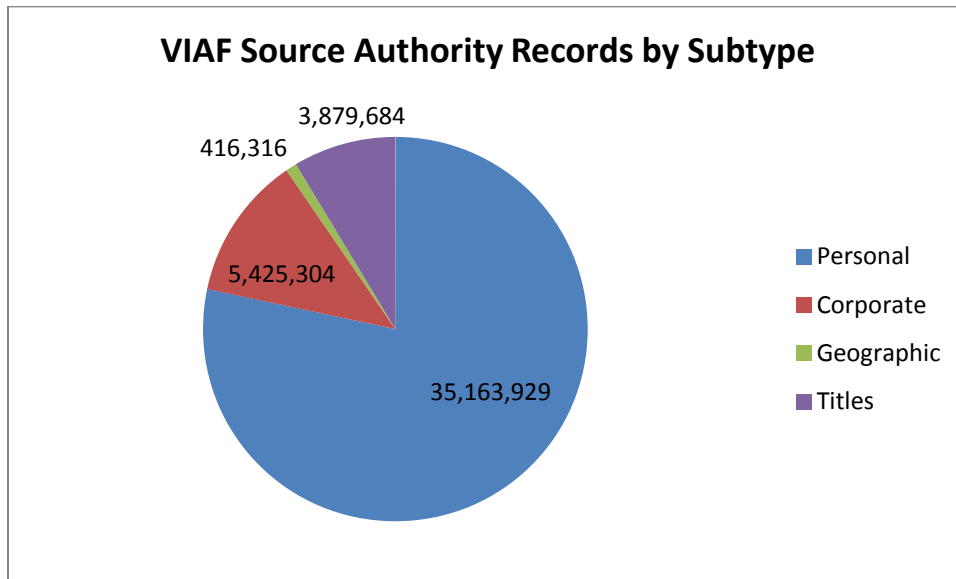


Figure 3: Source authority records by Subtype (June 2014) (source: OCLC)

VIAF Source data change over time

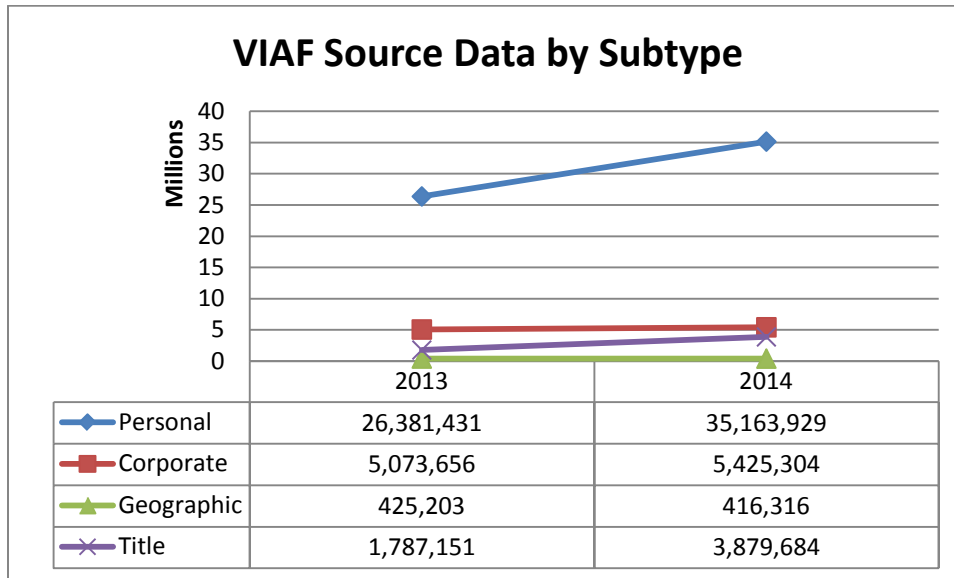


Figure 4: Source authority records by Subtype (June 2013 vs. June 2014) (source: OCLC)

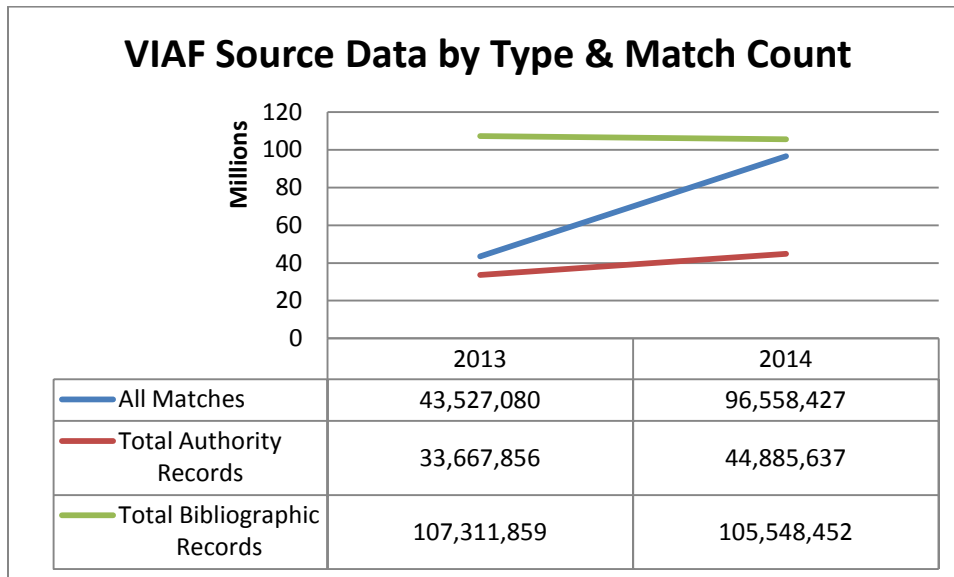


Figure 5: VIAF Source Data by Type & Match Count June 2013 vs. June 2014) (source: OCLC)

VIAF Data Sources Statistics Table

Source	Personal name records	Corporate and/or conference name records	Geographic name records	Uniform title records	Total number of authority records	% of records associated with bib data	% of Records matching at least one other source	Total number of matches	Total number of bibliographic records
BAV	278,577	36,687	5,626	8	320,898	95%	77%	1,285,365	646,216
BIBSYS	65,776	7,066	0	103	72,945	86%	62%	144,537	247,219
BNC	115,138	13,836	47	3	129,024	37%	67%	637,060	1,653,733
BNE	380,129	63,945	387	103,301	547,763	76%	64%	1,442,278	2,480,175
BNF	1,473,125	327,915	7,317	183,399	1,991,756	85%	79%	6,421,149	12,477,067
DBC	101,500	0	0	0	101,500	88%	38%	217,180	249,216
DNB	7,734,688	1,742,100	247,649	199,276	9,924,090	21%	44%	12,910,022	20,892,312
EGAXA	34,551	352	8	11	34,922	99%	63%	100,596	142,730
ICCU	169,000	28	0	0	169,028	0%	39%	272,199	41,638
ISNI	6,632,886	112,036	0	0	6,744,922	53%	100%	16,321,728	0
JPG	185,442	39,367	0	0	224,809	0%	40%	350,820	0
LAC	432,220	188,168	11,314	54,894	686,596	67%	70%	1,245,193	1,556,835
LC	6,056,206	1,642,525	134,341	1,151,307	8,984,379	51%	62%	14,799,695	14,195,071
LNB	142,483	12,878	11	67	155,439	96%	82%	782,947	422,196
LNL	8,036	0	0	0	8,036	99%	55%	23,605	13,832
NDL	817,047	191,062	7,454	3,154	1,018,717	95%	49%	1,287,354	4,724,655
NKC	584,712	131,295	0	0	716,007	50%	67%	2,163,978	2,331,675
NLA	712,685	209,410	790	88,697	1,011,582	0%	74%	2,968,410	0
NLB	8	4	396	0	408	0%	100%	635	0
NLI	493,409	61,365	0	1,505	556,279	50%	76%	1,960,443	1,358,893
NLP	655,311	89,313	47	48,470	793,144	77%	79%	2,384,206	1,837,130
NSK	430,775	31,676	32	1,559	464,042	86%	61%	852,902	877,329
NSZL	31,734	1,913	2	78	33,727	99%	93%	90,983	1,750,927
NTA	2,501,634	0	0	0	2,501,634	95%	82%	7,195,454	17,752,367
NUKAT	1,174,033	127,777	77	0	1,301,887	97%	88%	4,931,832	2,793,772

PERSEUS	1,228	0	0	0	1,228	0%	89%	4,549	0
PTBNP	297,269	72,936	0	0	370,205	96%	69%	801,734	949,424
RERO	156,054	56,145	442	22	212,664	93%	79%	888,297	1,677,804
RSL	997	0	0	0	997	0%	44%	3,058	0
SELIBR	156,773	17,954	175	6,548	181,450	57%	69%	681,337	2,142,426
SRP	209	0	0	0	209	0%	62%	1,421	0
SUDOC	2,161,368	241,447	0	23,391	2,426,228	13%	69%	5,623,984	11,284,718
SWNL	36,955	5,327	201	1	42,484	0%	69%	57,696	0
VLACC	4,820	771	0	0	5,591	92%	86%	38,276	1,049,092
WKP	1,135,025	0	0	0	1,135,025	0%	37%	2,119,433	0
XA	238	6	0	0	244	0%	99%	2,266	0
XR	1,888	0	0	2,013,890	2,015,778	0%	100%	5,545,805	0
	35,163,929	5,425,304	416,316	3,879,684	44,885,637			96,558,427	105,548,452

Table 2: VIAF Data Sources Statistics (note: data current as of end of June 2014) (source: OCLC)

VIAF processes the source data to yield at least 43 million clusters.

Clusters by type:

Expression	287,211 (current as of June 2013)
Geographic	416,316
Work	1,685,745 (current as of June 2013)
Corporate	5,425,304
Personal	35,163,929

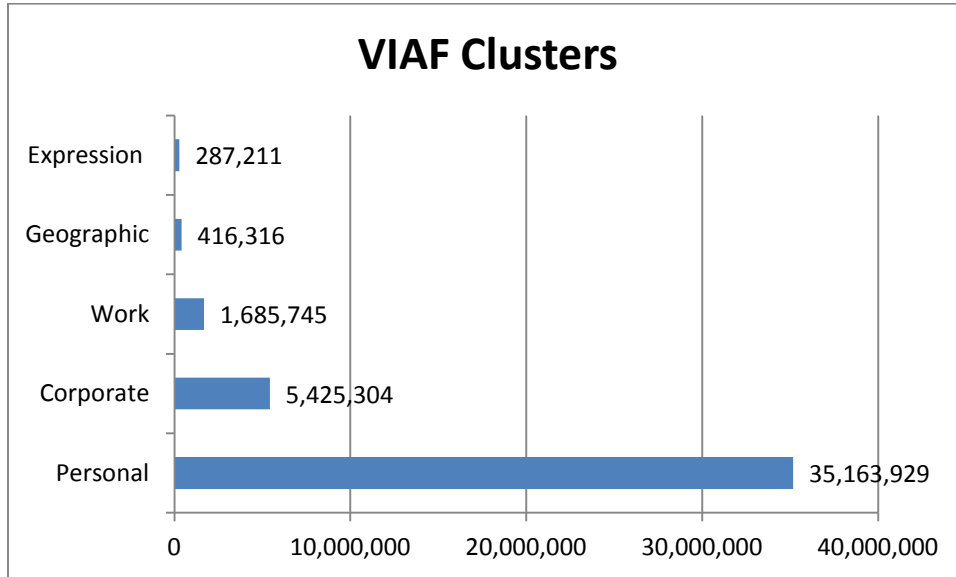


Figure 6: VIAF Clusters by Type (July 2014 [except as noted]) (source: OCLC)

VIAF Utilization

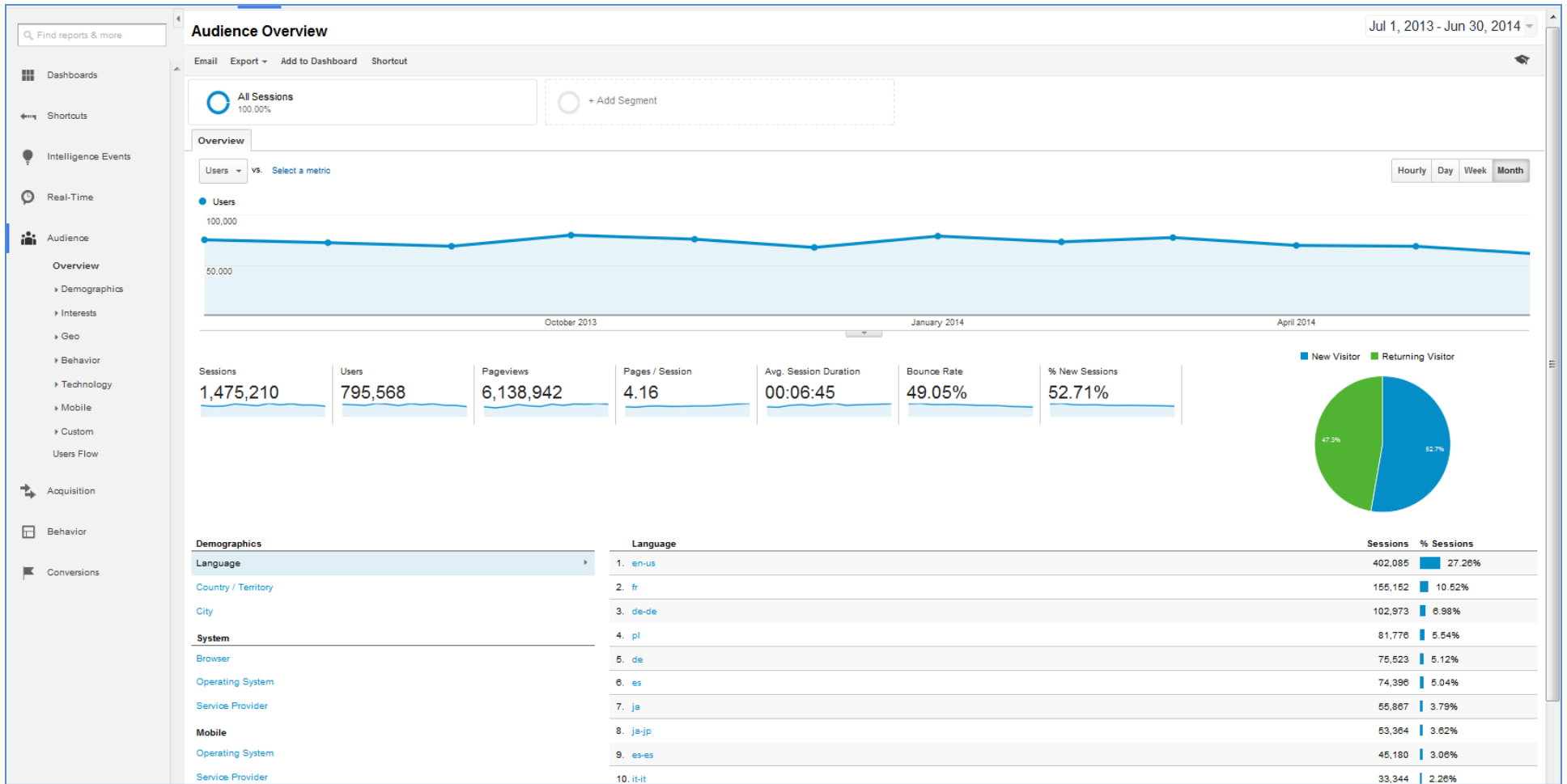


Figure 7: <http://viaf.org> metrics (1 July 2013 to 30 June, 2014) (source: Google Analytics)

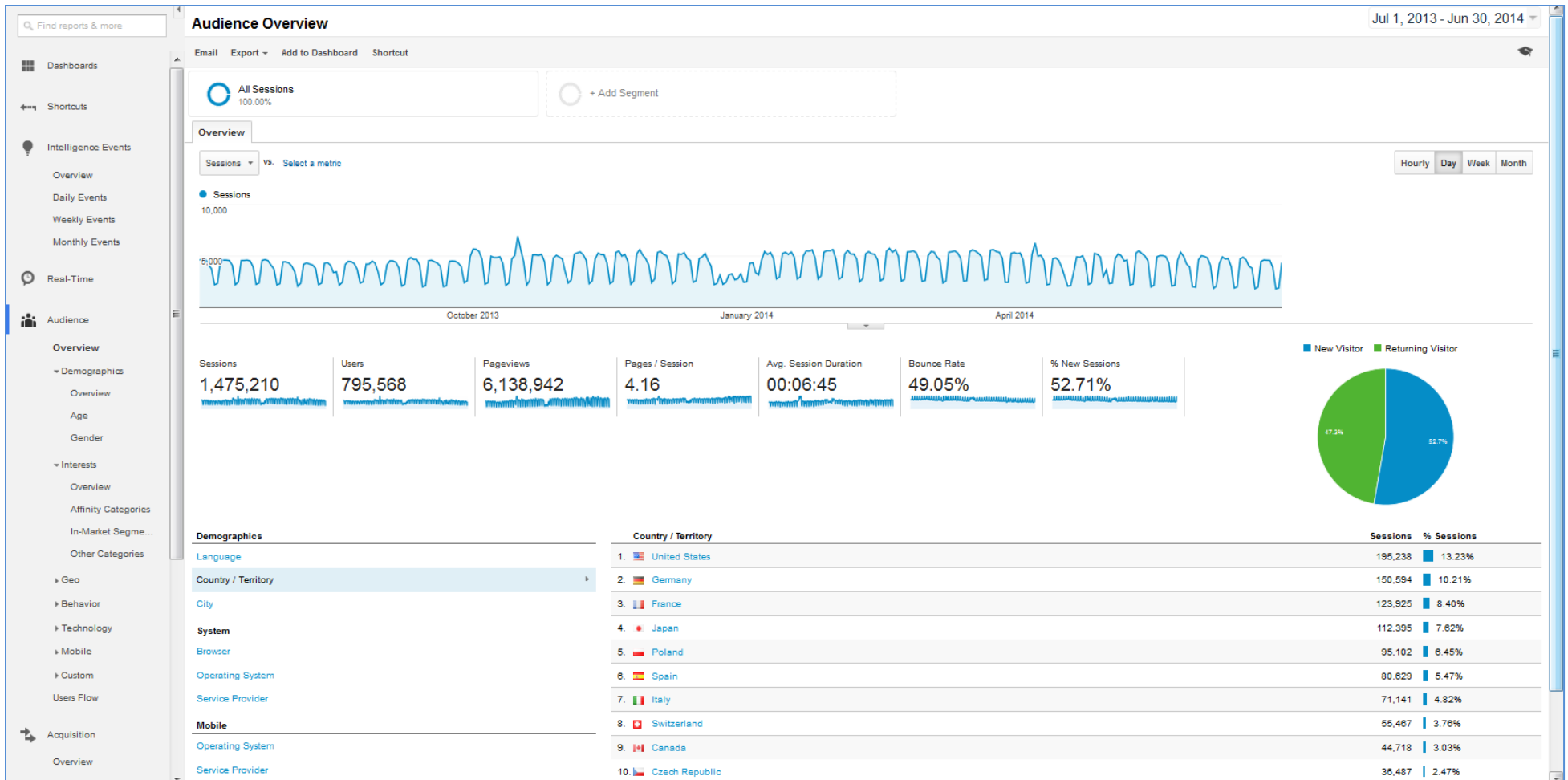


Figure 2: Major traffic to viaf.org by country of origin (1 July 2013 – 30 June, 2014) (source: Google Analytics)

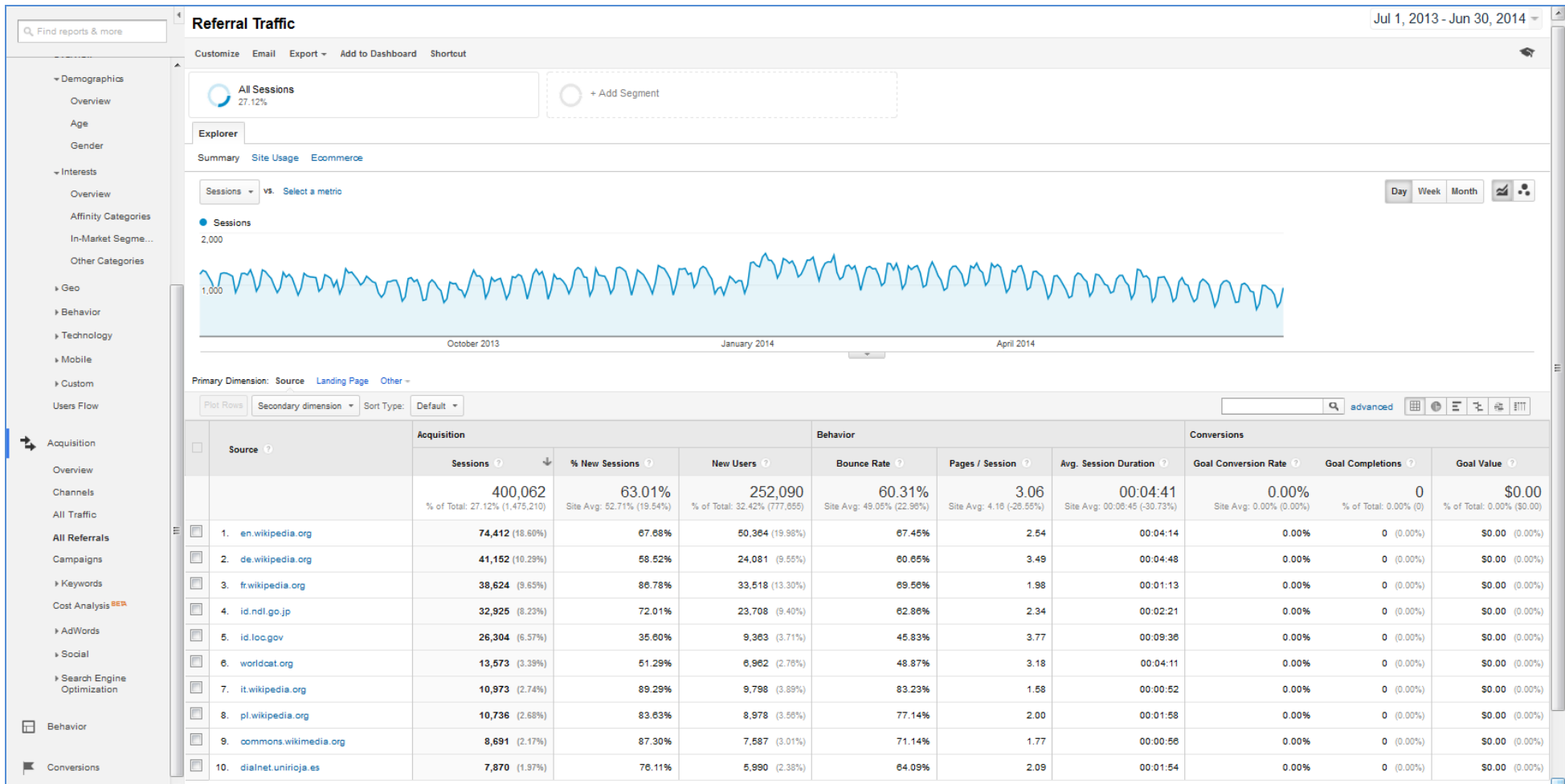


Figure 3: Referrals [subset of All Traffic] to viaf.org (1 July 2013 – 30 June, 2014) (source: Google Analytics)

VIAF Service Changes and Enhancements

OCLC continued to make enhancements to VIAF.

Details presented below are drawn from monthly update messages issued by the VIAF Team to the VIAF Contributors

Ingest, additions of data sources

- Addition of BNC, Biblioteca de Catalunya (Library of Catalonia) (July 2013)
- Wikipedia is now being updated monthly (July 2013)
- Addition of the National Library of Latvia (August, 2013)
- Added the Lebanese National Library (September, 2013)
- Added the National and University Library in Zagreb, Croatia (December, 2013)
- First installment of SRP (Syriac Reference Portal) records are loaded (March, 2014)
- Updated file from the National Library of Portugal (PTBNP) (March, 2014)
- Expanded file from ICCU (Italy) (March, 2014)
- VIAF noticed that some of OCLC LC harvests had failed, so VIAF reran them to get back up to date (April, 2014)
- New data from the DBC in Denmark (April, 2014)
- All the ISNI records (even those with little new information) are loaded as standard records (May, 2014)

Interface and other user experience-related changes

- Added an Hispánica view to viaf.org (July 2013)
- The preferred headings are now links only if they link out of VIAF (the little VIAF icon in front of them remains linked to the processed versions of the records) (September, 2013)
- Fixed a long-standing bug in the display of commas within personal names for ICCU records (September, 2013)
- \$g's in MARC-21 names are now displayed in parentheses (September, 2013)
- VIAF have updated OCLC autosuggest software which should make it easier to keep the autosuggest in sync with the database (September, 2013)
- Refreshed viaf.org's homepage: Added a map display, linked to descriptions of each file (per the recommendations and initial work of the VIAF Council Working Group on Advocacy) (November, 2013)
- Added a Polish language interface (displays if a user's interface language to be set to Polish) to viaf.org (December, 2013)

- Fixed a problem in the UNIMARC documentation page (<http://viaf.org/viaf/unimarctags.xml>) display (December, 2013)
- Added a 'sparse' note after the main heading display in the browser display & \$9 in MARC display (January, 2014)
- Only the first 100 field of each ISNI is displayed in the preferred headings section (May, 2014)

Machine services changes

- Fixed how VIAF respond to HTTP HEAD requests (August, 2013)
- Fixed a problem in the MARC-21 cluster dump (December, 2013)

Processing date information

- Better checks on dates when trying to find duplicate records within a source (September, 2013)
- Modified how dates are handled -- Now if only one of the dates is approximate VIAF will do exact matching on the other. (November, 2013)
- Improved OCLC date handling of UNIMARC dates of the form '19..' (December, 2013)
- Improved date handling of Getty Union List of Artist Names (January, 2014)
- Now recognize many dates based on Japanese eras and reigns (February, 2014)

xR/Multi-lingual bibliographic structure project

- Much of VIAF time this month was spent experimenting with uniform title records derived from WorldCat. VIAF have several hundred thousand sample authority records VIAF loaded in OCLC test system under the 'xR' source abbreviation. These represent the easiest to generate uniform title records for works and expressions published in multiple languages, and VIAF think it will be a great benefit to the usefulness of uniform titles in VIAF. It adds new work and expression-level records, along with information about translators, original languages and variant forms. (February, 2014)
- All of this is intertwined with OCLC FRBR processing as VIAF control names to VIAF IDs when possible, and use OCLC FRBR algorithms to group records to produce the xR uniform title records. Some of this should be showing up in WorldCat.org over the next few months as better work clusters, and OCLC experience using VIAF to control headings from WorldCat has prompted some of the recent changes in VIAF clustering. (February, 2014)
- VIAF refreshed the xR work/expression records. While these are all machine generated, VIAF now have a way to inform the process, helping with problem cases and giving us more control over the resulting records. VIAF is starting to use the xR records in OCLC FRBR work-set processing for WorldCat (April, 2014)
- Much of OCLC effort has been towards improving and extending the xR uniform title records VIAF is adding to VIAF (along with improving OCLC techniques for matching VIAF to WorldCat). You will now find approximately two million work/expression records derived from WorldCat and loaded into VIAF under the 'xR' source. (June, 2014)

- VIAF is starting to create xR work records for ‘orphan’ expression records (expressions with no corresponding work record) (May, 2014)

ISNI (International Standard Name Identifier)

- Modified the treatment of ISNI (including suppressing redundant information -- ISNI records may originate with VIAF clusters) (September, 2013)
- Avoid some mistaken merges of ISNI nodes (February, 2014)
- Improved given-name synonym file which should help us avoid some ‘unusual’ matches (March, 2014)
- Now recognize MARC-21 024 ISNI and VIAF links, as well as UNIMARC 010 ISNI links (April, 2014)
- Subfields \$g (MARC-21) and \$n (UNIMARC) are now treated as work subfields, not expression (May, 2014)

Parsing, clustering, merging, and other logic enhancements

- More refinements to handling of records marked as undifferentiated (July 2013)
- Better cluster stability going forward (July 2013)
- Not permitting the LC/NACO undifferentiated records to cluster with other records since the expectation is that they represent a mixed identity (August, 2013)
- Better title normalization on the titles associated with each entity (September, 2013)
- Better clustering for names with multiple links to multiple sources files (September, 2013)
Changed how the preferred forms in the full cluster display are handled - the names are all linked to internal processed records, rather than have some go to the source's site (November, 2013)
- Work on the clustering in VIAF, trying to bring together clusters that look like they should be merged (January, 2014)
- ‘Sparse’ (formerly called ‘lonely’) records are incorporated into other clusters if there is a clear cluster it should merge with. Sparse records are those for which VIAF has very little information other than a name and source identifier. Many of these are marked as undifferentiated, but others are records without informative notes that VIAF is unable to match to bibliographic records for more information. This affects more than 600,000 records that formed singleton clusters by themselves. (January, 2014)
- No longer attempt to match undifferentiated records with bibliographic information. (January, 2014)
- Another matching change VIAF made was to look for pairs of clusters in which the majority of the records are linked to records in the other across clusters (January, 2014)
- Changed how VIAF is treating Wikipedia parenthetical data in the names, avoiding some false matches. (January, 2014)
- Tightened up the rules for what makes records ‘sparse’ and available for merging on name only (February, 2014)

- Found a bug that was ignoring some sparse records, so more are now merged into clusters with other records (February, 2014)
- VIAF is more careful about treatment of matching undifferentiated records within a source (February, 2014)
- Made a change in how we select the VIAF ID when merging clusters (going forward we will use the lowest VIAF ID when merging equal-sized clusters). (April, 2014)
- Better job recognizing citations in 670 fields, giving us more title matches and as a result, bringing more records together. (June, 2014)

VIAF Outreach

OCLC Staff-authored/co-authored publications:

Hickey, Thomas B. and Jenny A. Toves. 2014. "Managing Ambiguity in VIAF." *D-Lib Magazine* (July/August 2014) v. 20, no. 7/8 [[doi:10.1045/july2014-hickey](https://doi.org/10.1045/july2014-hickey)]

Klein, Maximilian and Alex Kyrios. 2013. "VIAFbot and the Integration of Library Data on Wikipedia." *Code4Lib Journal*. Issue 22, 2013-10-14[<http://journal.code4lib.org/articles/8964>]

Marcia L. Zeng, Edward T. O'Neill, Jeff Mixer, Maja Žumer, and Xia Lin. 2013. Maximizing the usage of value vocabularies in the linked data ecosystem. In *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries* (ASIST '13), Andrew Grove (Ed.). American Society for Information Science, Silver Springs, MD, USA, , Article 57 , 2 pages. [<http://dl.acm.org/citation.cfm?id=2655837>]

OCLC Presentations:

OCLC staff members have presented information about VIAF in a variety of settings including:

VIAF API Workshop

OCLC Developer Network Webinar, 17 July 2014

<http://oclc.org/developer/events/api-workshops/viaf-api-workshop.en.html>

Thom Hickey

Cooperative Authority Control: Virtual International Authority File (VIAF)

[NISO/DCMI webinar](#), 4 December 2013

[Download the presentation](#) (.pptx: 6.8MB/71 slides)

[View on SlideShare](#)

Maximillian Klein

Authority Addicts: The New Frontier of Authority Control on Wikidata

[Wikimania 2013 International Wikimedia Conference](#), 7-11 August 2013, Hong Kong

[Download the presentation](#) (.pptx: 1MB/17 slides)

[View on SlideShare](#)

[Various OCLC staff]

Presentations at VIAF Council meeting, 16 August 2013, Singapore

See: <http://www.oclc.org/events/2013/viaf-at-ifla.en.html>

Selected presentations and publications by others:

ANGJELI, Anila and MAC EWAN, Andrew and BOULET, Vincent (2014) *ISNI and VIAF – Transforming ways of trustfully consolidating identities*. Paper presented at: IFLA WLIC 2014 – Lyon - Libraries, Citizens, Societies: Confluence for Knowledge in Session 86 - Cataloguing with Bibliography, Classification & Indexing and UNIMARC Strategic Programme. In: IFLA WLIC 2014, 16-22 August 2014, Lyon, France. (<http://library.ifla.org/985/>) [Presentation: 18 August 2014, Lyon, France]

VIAF Council

As part of the transition of VIAF to an OCLC service in 2012, the VIAF Council (VIAFC) was formed to advise OCLC about VIAF. All VIAF Contributors are invited to appoint an official member to the VIAFC.

In the 2013-2014 Council year³, work has continued on Privacy and Advocacy. The Council has been in close consultation with OCLC for the purpose of developing a document that would set forth eligibility guidelines for guiding OCLC in evaluating VIAF Contributor candidates.

VIAF Council will hold its 2014 Annual Meeting on 15 August, 2014 in conjunction with the IFLA WLIC 2014 meeting in Lyon, France.

Chairs of VIAF Council:

Beacher Wiggins (Library of Congress) (2016)

Vincent Boulet (Bibliothèque nationale de France) (2015)

Brigitte Wiechmann (Deutsche Nationalbibliothek) (2014)

Vincent Boulet (Bibliothèque nationale de France) (2013)

Barbara Tillett (Library of Congress) (2012)

ⁱ VIAF transitioned from a joint, experimental activity of the United States Library of Congress (LC), the German National Library (Deutsche Nationalbibliothek, or DNB) National Library of France (Bibliothèque nationale de France, or BnF) and OCLC to become an OCLC service in early 2012. For more information see the OCLC news announcement: <http://www.oclc.org/news/releases/2012/201224.en.html>

³ VIAF Council news may be found on this web site: <http://www.oclc.org/viaf/news.en.html>