

# ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ И СИСТЕМЫ

УДК 026.06+004.63

**М. В. Стегаева**

*Президентская библиотека им. Б. Н. Ельцина*

## **Виртуальный международный авторитетный файл как перспективное направление развития авторитетных файлов**

Рассмотрена история создания виртуального международного авторитетного файла (VIAF) – совместного международного проекта национальных библиотек и организаций, которые формируют авторитетные файлы имён. Представлены организационная структура VIAF, его цели и задачи, состав участников. Особое внимание уделено архитектуре VIAF и его компонентам. Охарактеризованы поисковые возможности VIAF. Сделан вывод: проект VIAF является частью конструкции Семантической сети (Semantic web) и примером представления библиотечных связанных данных на международном уровне.

Долгосрочная цель проекта VIAF – объединение авторитетных записей имён из авторитетных файлов национальных библиотек и других библиографирующих агентств в единую глобальную службу авторитетного контроля.

**Ключевые слова:** VIAF, виртуальный международный авторитетный файл, авторитетные данные, национальные библиотеки.

---

UDC 026.06+004.63

**Maria Stegaeva**

*B. Yeltsin Presidential Library, St. Petersburg, Russia*

## **The Virtual International Authority File as a perspective vector for further development**

The Virtual International Authority File (VIAF) is a joint international project of national libraries and organizations generating name authority files. The origins of VIAF project, VIAF organizational structure, its goals and tasks are described, project members are introduced. The VIAF architecture and the components are examined in particular, search features are characterized. The author concludes that the VIAF project becomes a component of the Semantic web structure and a model for presenting library linked data internationally. The VIAF long-term goal is to integrate authority name records of the authority files being generated by national libraries and bibliographic agencies into a single authority control service.

**Keywords:** VIAF, Virtual International Authority File, Authority data.

---

The problem of international authority file (AF) is discussed in the framework of IFLA since the early 1970s. Most large libraries had authority files of names of persons, names of organizations, geographic names, and other objects to maintain their uniformity in bibliographic records. These lists are compiled worldwide and supported in various ways in accordance with different national rules. The main objective of creation the Virtual International Authority File (VIAF) is to associate AF available in different languages into a single cluster, which will make use of reputable data cheaper and accessible, not only for the library community, but also for other communities (archives, museums, scientific, and others). VIAF project started in 1998 as a joint initiative of the Library of Congress, the German National Library and OCLC to create a connection between the national AF being maintained in these facilities – Library of Congress Name Authority File and Die Personennamendatei (authority file names of the German National Library, PND) – into a single system. The project head was Mrs. Barbara B. Tillett, Library of Congress. The first results were presented in the report at the 72th IFLA General Conference in Seoul in 2006. To automate the selection of names in the VIAF project the methodology of calculation of "strong", "medium" and "weak" match the metadata features of personal names. By "strong" signs included: name, title, the ISBN, co-authors, dates of life; to the "average" – publishing house, the role of the person, date of birth; to the "weak" – the language, country of publication. Methods for establishing matching pairs of names and identifying conditions for their automatic coupling have focused on the OAI-PMH protocol; 70% of the authority records of names that are common in both files were automatically linked, and the error rate was less than 1%. After considering various proposals, it was decided to transfer the operation of the VIAF keeping to the OCLC, which actually took place in 2012. The VIAF data are available under the – Open Data Commons Attribution License (ODC-BY). According to data from the OCLC 2014 report, the Russian state library input is 997 entries.

---

Виртуальный международный авторитетный файл (*Virtual International Authority File, VIAF*) – совместный международный проект национальных библиотек и организаций, формирующих авторитетные файлы имён.

Основная цель проекта – обеспечить библиотечному сообществу и информационным агентствам широкий и удобный доступ к основным мировым авторитетным файлам имён.

Проблема международного авторитетного файла (далее – АФ) дискутируется в рамках деятельности ИФЛА с начала 1970-х гг. Большинство крупных библиотек составляли авторитетные списки имён лиц, наименований организаций, географических названий и других объектов для поддержания их единообразия в библиографических записях. Эти списки, или в настоящее время – АФ, ведутся по всему миру и поддерживаются различными способами и в соответствии с различными правилами и методиками, что становится особенно очевидным, когда библиографические данные связываются в единые информационные массивы в рамках корпоративных международных проектов, таких как *WorldCat (OCLC)* и Европейская библиотека (*The European Library, TEL*).

Основная задача VIAF – связать имеющиеся АФ на различных языках в единый кластер, что сделает использование авторитетных данных более дешёвым и доступным не только для библиотечного сообщества, но и для других сообществ (архивных, музейных, научных и др.).

Проект VIAF стартовал в апреле 1998 г. и первоначально развивался как совместный проект Библиотеки Конгресса США, Немецкой национальной библиотеки и OCLC по созданию виртуального АФ путём связи национальных АФ, ведущихся в этих учреждениях, – *Library of Congress Name Authority File* (Авторитетный файл имён Библиотеки Конгресса США, *LCNAF*) и *Die Personennamendatei* (Авторитетный файл имён Немецкой национальной библиотеки, *PND*) – в единую систему [1]. Возглавила проект Барбара Тиллетт (*Barbara B. Tillett*, Библиотека Конгресса США).

В 2002 г. была разработана Концепция создания Виртуального международного авторитетного файла [2]. Предполагалось, что файлы будут связываться таким образом, чтобы они могли отражать эквиваленты, т.е. создавать пары или группы записей, которые относятся к одному и тому же объекту – лицу, вне зависимости от того, что сформированы по разным правилам каталогизации или же на разных языках/различной графической основе [3. С. 207].

Первые результаты совместных усилий по реализации VIAF были представлены в докладе участников проекта «*VIAF (Виртуальный международный авторитетный файл): авторитетные файлы имён, связывающие Немецкую библиотеку и Библиотеку Конгресса*» на 72-й Генеральной конференции ИФЛА в Сеуле в 2006 г. [4, 5]. Авторы доклада сообщали, что Немецкая национальная библиотека, Библиотека Конгресса США и OCLC совместно разрабатывают виртуальный международный АФ имён лиц, в рамках которого предстояло решить следующие задачи: установить связи между авторитетными записями библиотек – участниц проекта; предоставить свободный доступ к виртуальному международному АФ через веб-интерфейс.

Конечный результат проекта – проверка возможности автоматического установления связей между авторитетными записями различных национальных АФ и наглядная демонстрация эффективности таких связей.

На начальном этапе на базе авторитетных и библиографических файлов Библиотеки Конгресса США и Немецкой национальной библиотеки был создан виртуальный АФ, который включал свыше 6 млн имён и более 500 тыс. связей. Далее были определены категории объектов пересечений, подобраны пары совпадений. В задачу проекта входила и разработка алгоритмов автоматического подбора пар имён на основе сведений из авторитетных и библиографических записей.

Для автоматизации подбора имён в проекте VIAF использовалась методика подсчёта «сильных», «средних» и «слабых» признаков совпадения метаданных о лицах. К «сильным» признакам были отнесены: имя, заглавие, ISBN, соавторы, даты жизни; к «средним» – издательство, роль лица, дата рождения; к «слабым» – язык, страна публикации. Методы установления совпадений пар имён и выявления условий для их автоматического связывания были ориентированы на применение протокола для сбора метаданных OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*, Протокол сбора метаданных Инициативы Открытых Архивов) [6].

Разработчики также предложили несколько моделей для практической реализации виртуального АФ [2. С. 7, 8].

*Модель 1* – распределённый АФ (рис. 1). Эта модель предполагает наличие отдельных АФ, связанных между собой. Поиск производится в независимых АФ различных участников с использованием протокола, подобного Z-39.50.

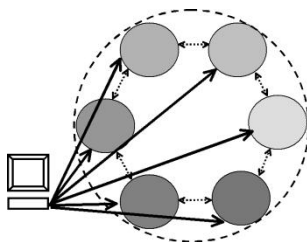


Рис. 1. Распределённая модель ведения АФ

*Модель 2* – «центральная связь» (рис. 2). Модель предполагает выбор одного АФ в качестве центрального и связь каждого АФ с ним. Поиск происходит с единственным поисковым входом в центральный АФ. В этом случае всем участникам нет необходимости устанавливать связи между собой, однако может возникнуть проблема при выборе центрального АФ.

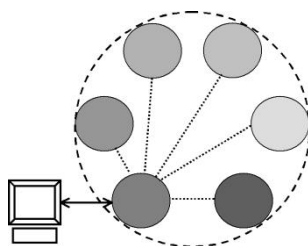


Рис. 2. Модель ведения АФ с центральной связью

*Модель 3* – централизованный АФ (рис. 3). Эта модель предполагает использование протокола, подобного протоколу OAI (*Open Archives Initiative*, Инициатива Открытых Архивов). Здесь используется сервер для хранения данных, поступающих из различных АФ. Информация на сервере постоянно обновляется по мере изменений, вносимых в отдельные АФ. За процесс ведения, обновления и поставки записей на центральный сервер отвечают сами участники, владеющие АФ. Недостатком этой модели могут быть потенциальные потери вследствие ошибок при обмене данными.

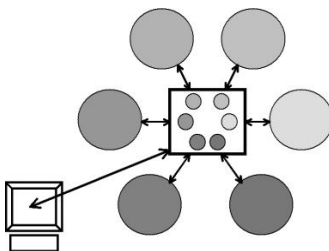


Рис. 3. Централизованная модель ведения АФ

Тестирование предложенных моделей показало, что именно модель 3 наиболее приемлема для воплощения идеи VIAF.

Успешная реализация проекта доказала возможность и целесообразность алгоритмической установки связи между именами лиц из различных национальных АФ; при этом 70% авторитетных записей имён, общих для обоих файлов, были связаны автоматически, а доля ошибок составила менее 1% [8. С. 169].

В августе 2003 г. в рамках 69-й Генеральной конференции ИФЛА, прошедшей в Берлине, Библиотека Конгресса США, Немецкая национальная библиотека и OCLC подписали соглашение о создании *Консорциума VIAF*. В октябре 2007 г. к Консорциуму присоединилась Национальная библиотека Франции. Эти четыре организации взяли на себя руководство Консорциумом и общую ответственность за функционирование VIAF, а также поставку авторитетных данных.

В 2010–2011 гг. среди участников Консорциума развернулась дискуссия о создании более подходящей долгосрочной организационной структуры, которая будет обеспечивать функционирование VIAF. После рассмотрения различных предложений было решено передать функционирование VIAF в ведение OCLC, что фактически состоялось в 2012 г.

В настоящее время OCLC реализует Программу VIAF, направленную на поддержку Виртуального международного авторитетного файла. В рамках этой Программы утверждается Совет VIAF (*VIAF Council, VIAFC*), который обеспечивает руководство политическими, практическими и управленческими вопросами VIAF.

В задачи VIAFC входит:

способствовать решению политических, практических и управленческих вопросов VIAF, а также политических вопросов, связанных с избранием новых участников VIAF;

консультировать по сервисным возможностям и форматам данных, поддерживаемых VIAF;

консультировать о способах и средствах, содействующих обеспечению эффективного использования VIAF;

назначать председателя VIAFC.

Организационная структура VIAFC – это члены, имеющие право голоса, и один представитель OCLC по связям. Каждый участник VIAF имеет право назначить голосующего члена и представить его в VIAFC. Представителя по связям назначает OCLC.

VIAFC ежегодно проводит выборы председателя (*VIAFC Chair*), который занимает свою должность в течение одного года. В конце годового собрания, которое по традиции проводится накануне начала работы Всемирного библиотечного и информационного конгресса ИФЛА, происходит смена председателя.

С переходом VIAF в ведение OCLC число его участников возросло с 19 учреждений из 22 стран (на начало 2012 г.) до 52 – из 37 стран (на ноябрь 2016 г.). 42 участника VIAF – это национальные библиотеки, ещё 10 библиотек предоставляют свои авторитетные данные в VIAF через федераль-

ные библиотечные агентства, консорциумы или в соответствии с другими соглашениями.

Для организаций, желающих стать участниками VIAF, на сайте OCLC размещено онлайн-приложение, в котором необходимо указать сведения об организации, контактных лицах и другую информацию.

После заполнения заявки, организации направляют в OCLC тестовые файлы с авторитетными и библиографическими записями. OCLC проверяет записи на соответствие их данных уже представленным ранее в VIAF другими участниками. Если тестирование завершается успешно, организация заключает с OCLC соглашение и предоставляет в OCLC полные АФ вместе с библиографическими записями. Организации также берут на себя ответственность регулярно поставлять в OCLC обновления.

Российское участие в проекте VIAF отражено АФ имён Российской государственной библиотеки, при этом, как отмечено на официальном сайте VIAF, РГБ, поставив ранее свои данные для VIAF, в настоящее время не обеспечивает обновление файла и не возобновила договор [9]. По данным отчёта OCLC за 2014 г., РГБ за весь период поставила в VIAF 997 записей [10]. С 2015 г. Российская национальная библиотека загружает записи в VIAF в режиме тестирования [9, 11].

Архитектурно VIAF представляет собой модель, разработанную для обеспечения интероперабельности различных АФ, т.е. способности к взаимодействию и функционированию на практическом уровне АФ различных учреждений из различных стран. Сервис VIAF не требует аутентификации пользователей, данные доступны по лицензии, регламентирующей открытые данные, – *Open Data Commons Attribution License* (ODC-BY – Лицензия, регламентирующая открытые данные, опубликованная сетью *Open Knowledge Foundation*). Разработан собственный интерфейс прикладного программирования – *API VIAF*, спецификации которого находятся в открытом доступе. Хостинг VIAF, а также поддержка программного обеспечения – в ведении OCLC [12].

Участвующие в проекте учреждения предоставляют авторитетные и библиографические данные в соответствии с установленными правилами и практикой функционирования VIAF. В этих данных выявляются однородные авторитетные данные для какого-либо объекта, происходят их сопоставление между собой и связывание в единую «*суперавторитетную запись*», или «*облако записей*» [13. С. 49].

Каждая запись, представленная в VIAF, включает: уникальный номер данных (*unique data number*), поля ссылок «см.» и «см. также» из оригина-

нальной записи и ссылки/связи с оригинальными авторитетными записями. Уникальный номер данных используется также для связи с библиографическими записями.

Записи в VIAF доступны либо как кластер записей (*cluster records*), либо как записи источников (*source records, authority source*).

*Кластер записей* объединяет различные АФ международных участников в одну авторитетную запись (*authority cluster*) для какого-либо конкретного объекта – лицо, организация, произведение и др.

С помощью ссылок, инструментом реализации которых выступает унифицированный идентификатор ресурса (URI), прописывается путь к национальным (региональным, локальным и др.) АФ – месту хранения/размещения информации об объекте. В результате несколько идентификаторов связываются в единую систему классификации и поиска информации.

URI направляет пользователя к конкретному (уникальному) имени объекта, принятому в каком-либо библиографирующем учреждении, предоставляя также возможность не только видеть запрашиваемую форму, но и просматривать другие принятые формы, установленные другими библиографирующими учреждениями для данного объекта. Таким образом, пользователь получает доступ к разрозненным авторитетным данным из единой точки, объединяющей посредством ссылок/связей различные имена конкретного объекта.

*Authority cluster* – конечная точка, которая обеспечивает идентификацию имени объекта, доступ к представлению данных и удобную навигацию за счёт инструментов, обеспечивающих поиск и просмотр авторитетных данных.

*Записи источников* – это авторитетные записи в различных MARC-форматах (MARC21, UNIMARC), которые были проиндексированы с целью выявления однородных сведений.

Весь массив авторитетных данных VIAF находится в открытом доступе и может быть использован для обмена между библиографирующими учреждениями в электронном виде.

Как уже отмечалось, практическая реализация VIAF построена на использовании протокола OAI с центральным сервером для сбора данных из национальных (региональных, локальных и т.п.) АФ. Система представляет собой виртуальную надстройку, так как оригинальная авторитетная запись остаётся в АФ учреждения – участника проекта VIAF и лишь небольшая часть данных собирается на центральном сервере [14].

Единой точкой входа для пользователей VIAF является официальный сайт VIAF (<http://viaf.org>).



Пользователю предоставлены возможности:

проводить поиск авторитетных данных по ключевому слову, локальному имени, принятой форме имени, заглавию, источнику или контрольному номеру записи источника;

получить – в результате поиска – авторитетные записи и отношения (связи) между авторитетными записями из различных источников;

получить сводную информацию из авторитетных записей, а также отношения (связи) между записями из различных источников;

получить авторитетные записи в различных форматах из нескольких национальных библиотек.

Используя интерфейс сайта VIAF, можно проводить поиск на языке как оригинала (т.е. языке объекта), так и пользователя. Результат поиска достигается посредством слияния совпадающих пар (или групп записей) для одинаковых имён конкретного объекта. При поиске используется механизм внутренней индексации, что позволяет избежать перегрузки сайта. Алгоритмы поиска и сбора информации обеспечивают создание виртуальной расширенной авторитетной записи, в состав которой входят все выявленные формы имён конкретного объекта [14]. Поиск возможен как во всём VIAF, так и в отдельных АФ участников проекта.

Информация для пользователя выводится в текстовом виде (с гиперссылками на источники) и в виде интерактивной графической схемы. Логическое представление информации о найденном объекте соответствует структуре авторитетной записи и включает:

список принятых точек доступа (с учётом совпадающих пар);

принятые формы имён (из всех авторитетных записей);

вариантные формы имён (из всех авторитетных записей);

связанные имена (из всех авторитетных записей);

источники (библиографические записи) и другую информацию.

При выводе данных на дисплей сохраняются также метки исходных форматов представления авторитетных данных – MARC21 и UNIMARC, возможен просмотр в форматах XML, RDF и др.

Проект VIAF – пример реализации представления библиотечных связанных данных на международном уровне. Связанные АФ используются в качестве многоязычных, многонациональных интерфейсов, поддерживающих доступ к библиографическим данным, а также функционируют как консолидированный справочник форм имён объекта, созданных различными библиографирующими учреждениями и доступных для поиска и отображения.

Создатели VIAF отмечают, что он является частью конструкции Семан-

тической сети (*Semantic web*), позволяющей представлять имена объектов в предпочтительном для пользователей языковом оформлении [13. С. 49].

Проект создания VIAF способствует расширению универсального библиографического учёта, позволяя: 1) сосуществовать национальным (региональным, локальным и т.п.) различиям в принятых формах имён объектов и 2) поддерживать интересы пользователей авторитетных и библиографических данных, принадлежащих к различным языковым сообществам.

Долгосрочная цель проекта VIAF – объединение авторитетных записей имён из АФ национальных библиотек и других библиографирующих агентств в единую глобальную службу авторитетного контроля.

### СПИСОК ИСТОЧНИКОВ

1. **Hengel C.** Normdaten und Metadaten. Die Idee eines Internationalen Authority File // Z. für Bibliothekswesen u. Bibliogr. – 2003. – Jg. 50. – № 4. – С. 210–214.

2. **Tillett B. B.** The Virtual International Authority File. [2002]. – Режим доступа: <http://www.iccu.sbn.it/upload/documenti/TillettAF.ppt>.

3. **Селиванова Ю. Г.** Современные принципы и средства библиографического контроля / Селиванова Ю. Г., Масхулия Т. Л. // Тр. Междунар. библиогр. конгр. Ч. 2. – Санкт-Петербург, 2012. – С. 205–211.

*Selivanova Yu. G. Sovremennye printsipy i sredstva bibliograficheskogo kontrolya / Selivanova Yu. G., Mashuliya T. L. // Tr. Mezhdunar. bibliogr. kongr. Ch. 2. – Sankt-Peterburg, 2012. – S. 205–211.*

4. **VIAF** (Virtual International Authority File): linking die Deutsche Bibliothek and Library of Congress name authority files / Rick Bennett, Christina Hengel-Dittrich, Edward T. O’Neill, Barbara B. Tillett. – IFLA, 2006. – Режим доступа: <http://www.ifla.org/IV/ifla72/papers/123-Bennett-en.pdf>.

5. **VIAF** (Виртуальный международный авторитетный файл) : авторитетные файлы имён, связывающие Немецкую библиотеку и Библиотеку Конгресса США / Р. Беннетт, К. Хенгель-Диттрих, Э. Т. О’Нейл, Б. Б. Тиллетт // Новости Междунар. федерации библиотеч. ассоц. и учреждений. – 2007. – № 6 (69). – С. 35–43.

*VIAF (Virtualnyy mezhdunarodnyy avtoritetnyy fayl) : avtoritetnye fayly imen, svyazyvayushchie Nemetskuyu biblioteku i Biblioteku Kongressa SSHA / R. Bennett, K. Hengel-Dittrich, E. T. O’Neyl, B. B. Tillett // Novosti Mezhdunar. federatsii bibl. assots. i uchrezhdeniy. – 2007. – № 6 (69). – S. 35–43.*

6. **Стрелкова Е. В.** Развитие системы авторитетных файлов и оцифровка библиотечных фондов. [2009]. – Режим доступа: <http://fulltext.pl.spb.ru/unor/metod/bibl/2009/StrelkovaE.pdf>.

*Strelkova E. V. Razvitiye sistemy avtoritetnykh faylov i otsifrovka bibliotечnykh fondov. [2009].*

7. **Tillett B. B.** VIAF: Virtual International Authority File. – 2006. – Режим доступа: <http://www.tnla.org/associations/5700/files/TL562behrend4.pdf>.

8. **Каталогизация:** современные технологии. Тенденции и перспективы развития / Селиванова Ю. Г., Масхулия Т. Л., Жлобинская О. Н., Стегаева М. В. – Москва, 2007. – 215 с.

*Katalogizatsiya: sovremennye tehnologii. Tendentsii i perspektivy razvitiya / Selivanova Yu. G., Mashuliya T. L., Zhlobinskaya O. N., Stegaeva M. V. – Moskva, 2007. – 215 s.*

9. **VIAF®** : Virtual International Authority File. OCLC, 2014. – Режим доступа: <http://www.oclc.org/viaf.en.html>.

10. **OCLC 2014 Annual Report to VIAF Council.** OCLC, cop. 2014. – Режим доступа: <http://www.oclc.org/content/dam/oclc/viaf/OCLC-2014-VIAF-Annual-Report-to-VIAF-Council.pdf>.

11. **OCLC Research: 2015 Activity Report.** OCLC, cop. 2016. – Режим доступа: <http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-2015-activity-report.pdf>.

12. **VIAF** (Virtual International Authority File). OCLC, 2015. – Режим доступа: <https://www.oclc.org/developer/develop/web-services/viaf.en.html>.

13. **Земсков А. И.** Новые формы обработки электронных документов // Науч. и техн. б-ки. – 2013. – № 8. – С. 44–51.

*Zemskov A. I. Nove formy obrabotki elektronnykh dokumentov // Nauch. i tehn. b-ki. – 2013. – № 8. – S. 44–51.*

14. **VIAF** // Википедия. [2012]. – Режим доступа: <https://ru.wikipedia.org/wiki/VIAF>.

*VIAF // Vikipediya. [2012].*

---

*Maria Stegaeva, Deputy Department Head, B. Yeltsin Presidential Library;  
stegaeva@prlib.ru  
3, Senatskaya pl., 190000 St. Petersburg, Russia*