

# ПРОЕКТИРОВАНИЕ ЕДИНОГО ХРАНИЛИЩА МУЛЬТИМЕДИЙНЫХ ДАННЫХ ИЗ ПОЛЕВЫХ ЭТНОГРАФИЧЕСКИХ ЭКСПЕДИЦИЙ

Д. С. Левшун<sup>1\*</sup>

<sup>1</sup> СПИИРАН, Санкт-Петербург, 199178, Российская Федерация

\* Адрес для переписки: levshun@comsec.spb.ru

## Аннотация

Неотъемлемой частью современных полевых этнографических экспедиций стали массивы разнородных цифровых данных: аудио- и видео записи, фотографии и текстовые документы, позволяющие зафиксировать в цифровом формате интервью, ритуалы, элементы быта и фольклора изучаемого этноса. При этом объем накапливаемых качественных данных настолько велик, что все более актуальным становится решение проблем их хранения, обработки, поиска и архивации, а также обеспечения процесса обмена подобными данными между исследователями. **Предмет исследования.** Данная работа посвящена проектированию единого хранилища мультимедийных данных из полевых этнографических экспедиций. **Метод.** Рассмотрены основные задачи, на решение которых направлены исследования данной тематики. Рассмотрены современные системы хранения мультимедийных данных. **Основные результаты.** Сформированы требования к современной системе мультимедийных данных. Осуществлен сравнительный анализ основных подходов к построению баз данных. Обосновано построение единого хранилища на основе системы управления содержимым. Представлена архитектура полученного решения. **Практическая значимость.** Разработанное единое хранилище мультимедийных данных из полевых этнографических экспедиций представляет собой одно из возможных решений проблем хранения, обработки, поиска и архивации цифровых данных из полевых этнографических экспедиций, а также обеспечивает процесс обмена подобными данными между исследователями.

## Ключевые слова

хранилище мультимедийных данных, проектирование хранилищ данных, этнографические экспедиции, полевые экспедиции.

## Информация о статье

УДК 004.55

Язык статьи – русский.

Поступила в редакцию 30.06.17, принята к печати 01.09.17.

**Ссылка для цитирования:** Левшун Д. С. Проектирование единого хранилища мультимедийных данных из полевых этнографических экспедиций // Информационные технологии и телекоммуникации. 2017. Том 5. № 3. С. 84–92.

# DESIGN OF UNIFIED REPOSITORY FOR MULTIMEDIA DATA FROM FIELD ETHNOGRAPHIC EXPEDITIONS

D. Levshun<sup>1\*</sup>

<sup>1</sup> SPIIRAS, St. Petersburg, 199178, Russian Federation

\* Corresponding author: levshun@comsec.spb.ru

**Abstract**—An integral part of the modern field of ethnographic expeditions is arrays of heterogeneous digital data: audio and video recordings, photos and text documents that allow to save interviews, rituals, elements of everyday life and folklore of the ethnos studied in digital format. At the same time, the volume of accumulated data is so great that it becomes increasingly relevant to solve the problems of their storage, processing, retrieval and archiving, and to provide a process for exchanging such data among researchers. **Research subject.** This paper is devoted to the design of unified repository for multimedia data from field ethnographic expeditions. **Method.** The paper contains the analysis of the main challenges of this research field. Also it presents the modern repositories for multimedia data and requirements for them. **Core results.** A comparative analysis of the repositories and general approaches to the repositories construction was carried out and described. Based on the requirements and existing solutions the new repository architecture was proposed. **Practical relevance.** Developed unified repository for multimedia data from field ethnographic expeditions represents one of the possible solutions to the problems of storage, processing, retrieval and archiving of digital data from field ethnographic expeditions, and also provides a process for the exchange of such data between researchers.

**Keywords**—multimedia data repository, design of data repository, ethnographic expeditions, field expeditions.

## Article info

Article in Russian.

Received 30.06.17, accepted 01.09.17.

**For citation:** Levshun D.: Design of Unified Repository for Multimedia Data from Field Ethnographic Expeditions // Telecom IT. 2017. Vol. 5. Iss. 3. pp. 84–92 (in Russian).

## Введение

Неотъемлемой частью современных полевых этнографических экспедиций стали массивы разнородных цифровых данных: аудио- и видео записи, фотографии и текстовые документы, позволяющие зафиксировать в цифровом формате интервью, ритуалы, элементы быта и фольклора изучаемого этноса [1]. При этом объем накапливаемых качественных данных настолько велик, что все более актуальным становится решение проблем их хранения, обработки, поиска и архивации, а также обеспечения процесса обмена подобными данными между исследователями.

В рамках единого хранилища мультимедийных данных, текстовые файлы хранятся совместно с графическими, аудио- и видеофайлами различного формата, в то время как основной массив текстовых данных может быть вынесен в качестве отдельного лингвистического корпуса [2]. При этом специфика дан-

ных определяет ряд дополнительных требований к реализации хранилища, а современные исследования данной тематики, как правило, направлены на решение следующих задач [3]:

- представление результатов полевых этнографических экспедиций в полном объеме и без возможности разночтения;
- публичное представление результатов полевых этнографических экспедиций в средствах массовой информации;
- предоставление удаленного доступа к результатам полевых этнографических экспедиций посредством сети Интернет.

На момент проведения исследования, автору данной работы был предоставлен доступ к полевому архиву, содержащему более сотни видео- и аудиозаписей процесса исполнения эпических сказаний, проведения ритуалов и изготовления предметов с ритуальным значением [4]. Также были предоставлены интервью со сказителями и участниками ритуалов, их родными и близкими. Кроме того, был предоставлен доступ к нескольким тысячам фотографий, отображающих различные моменты ритуалов и сказаний, местных животных и растения, элементы быта и обихода, имеющих важное семиотическое значение.

Работа с этнографами, а также реальными данными экспедиций позволила автору на основе существующих решений сформулировать ряд ключевых требований, которым должно соответствовать современное единое хранилище мультимедийных данных с учетом специфики предметной области. На основе данных требований была разработана архитектура единого хранилища, что в дальнейшем легло в основу разработки современного хранилища мультимедийных данных из полевых этнографических экспедиций.

## **1. Постановка задачи**

Активное развитие цифровой обработки и архивации результатов этнографических исследований привело к появлению многочисленных систем хранения мультимедийных данных. В качестве примера наиболее распространенных на территории Российской Федерации можно привести следующие: проект "Фольклор и народная культура России"<sup>1</sup>, проект "Звуковые Архивы: Европейская память о ГУЛАГе"<sup>2</sup>, проект "Устная история"<sup>3</sup>, проект "По ту сторону войны"<sup>4</sup> и портал "Антропогенез.ru"<sup>5</sup>. Анализ данных решений, представляющих собой как специализированные разработки, так и сайты на основе распространенных систем управления содержимым (CMS), позволил сформировать перечень основных требований к единому хранилищу мультимедийных данных из полевых этнографических экспедиций.

### **Требования к хранению мультимедийных данных:**

- 1) поддержка хранения, архивирования и публикации текстовых, аудио- и видеоматериалов, изображений;

---

<sup>1</sup> Фольклор и народная культура России. URL: <http://folk.ru>

<sup>2</sup> Звуковые Архивы: Европейская память о ГУЛАГе. URL: <http://museum.gulagmemories.eu/ru>

<sup>3</sup> Устная История. URL: <http://oralhistory.ru>

<sup>4</sup> По ту сторону войны: устная история военнопленных и Остарбайтеров. URL: <http://archive.tastorona.su>

<sup>5</sup> Антропогенез.ru. URL: <http://antropogenez.ru>

- 2) поддержка привязки мультимедийных данных к географическим координатам места их получения;
- 3) поддержка привязки мультимедийных данных к временной отметке и часовому поясу места их получения;
- 4) поддержка привязки мультимедийных данных к произвольным текстовым меткам.

#### **Требования к отображению мультимедийных данных:**

- 1) отображение маршрута полевой этнографической экспедиции и обнаруженных материалов на специальной карте;
- 2) отображение этапов полевой этнографической экспедиции и обнаруженных материалов на специальной хронологической линейке;
- 3) группировка изображений и представление их в виде галереи, наличие видео- и аудиоплеера.

#### **Требования к поиску мультимедийных данных:**

- 1) поиск материалов по обнаружившему их участнику этнографической экспедиции;
- 2) поиск материалов по названию экспедиции;
- 3) поиск материалов по дате их обнаружения;
- 4) поиск материалов по произвольному запросу.

**Требования к приватности отображаемых мультимедийных данных:** возможность ограничения доступа к мультимедийным данным на основе типа учетной записи пользователя.

При этом важно отметить необходимость наличия удобного пользовательского интерфейса, предоставляющего функционал взаимодействия с мультимедийными данными (добавить, изменить, удалить, ограничить доступ), т. к. подавляющее большинство пользователей подобных систем не являются техническими специалистами [5].

## **2. Анализ подходов к разработке архитектуры хранилища мультимедийных данных**

С точки зрения разработки архитектуры хранилища мультимедийных данных существует два основных подхода: онтологический и реляционный [6]. Рассмотрим каждый из них более подробно.

**При онтологическом подходе** к построению баз данных выделяют три основных объекта представления данных: класс, атрибут и экземпляр. Класс — это абстрактное представление данных, соответствующее некоторому понятию. Например, в качестве классов данных для построения единого хранилища мультимедийных данных из полевых этнографических экспедиций могут быть использованы такие понятия как: экспедиция, поселение, народ, язык, этнограф и т. д. Экземпляр — это конкретная реализация класса. Например, к классу народ можно отнести экземпляр ифугао, а к классу язык можно отнести экземпляр худхуд. При этом взаимосвязь между соответствующими классами и экземплярами достигается благодаря использованию атрибутов.

Грамотно построенная онтология представляет собой сложную структуру данных, в которой описаны все необходимые классы данных в соответствии, которым при помощи атрибутов ставятся различные экземпляры. Запрос к любому из экземпляров данных позволяет извлечь полную информацию о взаимосвязи

данного экземпляра со всеми его классами, что является важным преимуществом онтологического подхода. Например, запрос по фамилии конкретного этнографа позволит извлечь данные как о возрасте, образовании и месте рождения исследователя, так и обо всех экспедициях, в которых данный этнограф принял участие, а также обо всех мультимедийных материалах, которые были получены с его помощью.

К сожалению, сложная и разветвленная структура подробной онтологии может существенно снизить скорость работы с данными в базе. При этом описание большого массива мультимедийных данных — задача, выполняемая только коллективными усилиями, а представления об описании мультимедийных данных у разных людей могут отличаться, что может повлечь за собой появление нескольких онтологий, слабо связанных между собой.

**При реляционном подходе** к построению баз данных задается более жесткая и менее адаптивная структура данных. Основными объектом представления данных является таблица, содержащая информацию о конкретном понятии, а реляционная база данных представляет собой набор взаимосвязанных таблиц. Каждая таблица представляет собой набор столбцов, соответствующих атрибутам понятия, и строк или кортежей, соответствующих описанию экземпляра понятия. Для отождествления однозначным образом отдельных экземпляров понятия в таблице используются первичные ключи. Связь между таблицами устанавливается при помощи внешних ключей. Таким образом, извлечение данных об конкретном экземпляре предоставит информацию об этом экземпляре только в рамках определенного понятия. Если же необходимо извлечь полную информацию об экземпляре, то к реляционной базе данных строится сложный запрос, объединяющий и, если необходимо, преобразующий данные нескольких таблиц.

В рамках такой логической модели данных представление и хранение неструктурированных данных, а именно к этому виду данных относятся мультимедийные данные из полевых этнографических экспедиций, — сложная задача. С другой стороны, сложный математический аппарат, которые лежит в основе реляционного подхода к построению баз данных, предоставляет оптимальные характеристики простоты, устойчивости, гибкости, производительности, масштабируемости и совместимости при управлении данными. Оптимальные характеристики не означают, что показатели по каждому отдельному показателю у реляционного подхода также выше, чем у онтологического. К тому же, обеспечение оптимальности данных характеристик повлекло за собой сложность внутренней реализации реляционных баз данных. Основываясь на алгоритмах оценки стоимости, реляционные системы управления базами данных оценивают каждый запрос и составляют план его выполнения. Сложность реляционных баз данных накладывает некоторые ограничения на их масштабируемость. Это означает, что существует некоторый предел масштабируемости, за пределами которого преимущества реляционного подхода уже не так очевидны. Отметим, что для задачи построения единого хранилища мультимедийных данных из полевых этнографических экспедиций масштабируемости реляционных баз данных более чем достаточно.

Подытоживая вышесказанное, отметим, что в применимости к построению единого хранилища мультимедийных данных из полевых этнографических экспедиций нет однозначно лучшего подхода. Как онтологические, так и реляционные

базы данных обладают помимо важных достоинств и серьёзными недостатками [7]. Главным преимуществом реляционного подхода к построению баз данных является наличие CMS, позволяющих обеспечить поддержку единого хранилища мультимедийных данных за счет предоставления средств добавления и архивации данных, контроля их целостности, а также соблюдения режима доступа к ним.

### 3. Анализ современных CMS

Проанализируем возможность построения единого хранилища мультимедийных данных из полевых этнографических экспедиций на основе CMS общего назначения Drupal и TYPO3, а также специализированной CMS Omeka.

CMS Drupal — это открытая система управления содержимым, которая также может быть использована в качестве каркаса для веб-приложений (CMF). CMS Drupal написана на языке PHP, в качестве хранилища данных используются такие системы управления базами данных (СУБД) как MySQL, PostgreSQL и др. Важной особенностью Drupal является гибкая схема организации структуры хранилища данных на основе таксономии. Таксономия — это механизм, позволяющий задавать произвольные тематические категории для содержимого хранилища, ассоциируя их с модулями ввода и вывода информации. При этом категории могут быть представлены в виде плоских или иерархических списков, а также более сложных структур, в которых элемент может иметь более одного родителя. Другой важной особенностью Drupal является расширение CCK (*Content Construction Kit*). CCK позволяет дополнять объекты полями данных различных типов, в том числе и отображения мультимедийных данных. Drupal позволяет работать с текстовыми, аудио- и видеоматериалами, а также изображениями при помощи различных расширений. Также Drupal поддерживает взаимодействие с различными картографическими сервисами, в том числе с Картами Google и Яндекс Картами. Drupal поддерживает хронологические метки, позволяя расположить мультимедийные данные на хронологической линейке. Доступ к мультимедийным данным может быть разграничен относительно роли пользователя в системе или авторства.

CMS TYPO3 — это открытая система управления содержимым, которая, как и Drupal может быть использована в качестве CMF. CMS TYPO3 написана на языке PHP, в качестве хранилища данных используются такие СУБД как MySQL, Oracle Database, PostgreSQL и др. Взаимодействие с различными СУБД осуществляется при помощи специального расширения TYPO3 DBAL. DBAL или Database Abstraction Layer представляет собой универсальный интерфейс взаимодействия CMS с различными СУБД. TYPO3 позволяет работать с текстовыми, аудио- и видеоматериалами, а также изображениями при помощи различных расширений. Также TYPO3 поддерживает взаимодействие с различными картографическими сервисами, в том числе с Картами Google и Яндекс Картами. TYPO3 позволяет расположить мультимедийные данные на хронологической линейке при помощи специальных расширений. Доступ к мультимедийным данным может быть разграничен относительно роли пользователя в системе.

CMS Omeka — это открытая система управления содержимым. CMS Omeka написана на языке PHP, а в качестве хранилища данных используется MySQL СУБД. CMS Omeka активно поддерживается библиотекарями и архивариусами, учеными, сотрудниками исторических организаций, различными общественными



организациями, преподавателями и студентами, а также различными компаниями и энтузиастами. При помощи специального расширения, CMS Omeka позволяет настроить взаимодействие с картографическим сервисом Google на основе географических меток мультимедийных данных, а другое расширение позволяет CMS Omeka на основе временных меток мультимедийных данных позволяет сформировать хронологическую линейку и настроить ее отображение. Также поддерживается ролевое разграничение доступа к мультимедийным данным.

Таким образом, каждая из рассмотренных CMS соответствуют требованиям к единому хранилищу мультимедийных данных (см. раздел 1). Однако, соответствие каждой CMS определенному требованию не означает, что функционал, предоставляемый каждой системой относительно отдельного требования также одинаков. Наиболее подходящей CMS для реализации единого хранилища с точки зрения предоставления некоторых специфических функций, а именно функций работы с картой и хронологической линейкой полевых этнографических экспедиций, является CMS Omeka.

#### **4. Единое хранилище мультимедийных данных**

Для реализации и заполнения единого хранилища мультимедийных данных из полевых этнографических экспедиций были использованы: CMS Omeka v2.5.1, расширение Geolocation v2.2.5, расширение Neatline v2.6.0, расширение NeatlineSimile v2.0.4, расширение NeatlineText v1.1.0, расширение NeatlineTime v2.1.0, расширение NeatlineWaypoints v2.0.2, расширение HTML5 Media v2.6. Рассмотрим каждое из расширений более подробно.

Расширение Geolocation позволяет добавлять географическую метку к описанию мультимедийных данных. При этом расширение автоматически формирует географическую карту, содержащую отмеченные мультимедийные данные, а также отдельную веб-страницу для отображения карты (географическая карта строится при помощи картографического сервиса Карты Google).

Расширение Neatline представляет собой фреймворк, который позволяет организовывать интерактивное представление массива мультимедийных данных. Так, на основе NeatlineSimile и NeatlineTime открывается возможность расположения данных в хронологическом порядке посредством специальных временных меток. На основе NeatlineWaypoints предоставляется возможность нанести на карту региона экспедиции (прикрепляется в виде фонового изображения) точки маршрута, а также направления следования между ними. NeatlineText позволяет добавить интерактивное текстовое описание объектам фреймворка Neatline.

Расширение HTML5 Media позволяет проигрывать аудио- и видеоматериалы при использовании устройств и браузеров, поддерживающих HTML5. Данное расширение предоставляет возможность отображения видео- и аудиозаписей различных форматов, однако наиболее качественная поддержка гарантируется для форматов H.264 и MP3 соответственно.

Таким образом, реализация единого хранилища мультимедийных данных из полевых этнографических экспедиций на основе CMS Omeka может быть сведена в едином представлении (см. табл.).

## Реализация единого хранилища на основе CMS Omeka

| Требование   | Элемент  |
|--|--|
| Поддержка хранения, архивирования и публикации текстовых, аудио- и видеоматериалов, изображений                        | Omeka v2.5.1   |
| Поддержка привязки мультимедийных данных к географическим координатам места их получения                               | Geolocation v2.2.5   |
| Поддержка привязки мультимедийных данных к временной отметке и часовому поясу места их получения                       | Neatline v2.6.0, NeatlineSimile v2.0.4, NeatlineTime v2.1.0                      |
| Поддержка привязки мультимедийных данных к произвольным текстовым меткам   | Omeka v2.5.1   |
| Отображение маршрута полевой этнографической экспедиции и обнаруженных материалов на специальной карте                 | Geolocation v2.2.5, Neatline v2.6.0, NeatlineWaypoints v2.0.2                    |
| Отображение этапов полевой этнографической экспедиции и обнаруженных материалов на специальной хронологической линейке | Neatline v2.6.0, NeatlineSimile v2.0.4, NeatlineTime v2.1.0, NeatlineText v1.1.0 |
| Группировка изображений и представление их в виде галереи, наличие видео- и аудиоплеера                                | Omeka v2.5.1, HTML5 Media v2.6   |
| Поиск материалов по обнаружившему их участнику этнографической экспедиции  | Omeka v2.5.1   |
| Поиск материалов по названию экспедиции, дате их обнаружения, произвольному запросу                                    | Omeka v2.5.1, NeatlineTime v2.1.0  |
| Возможность ограничения доступа к мультимедийным данным на основе типа учетной записи пользователя                     | Omeka v2.5.1   |

### Заключение

В качестве основы для проектирования единого хранилища мультимедийных данных из полевых этнографических экспедиций была использована CMS Omeka. Применение CMS позволяет обеспечить хранение данных, контроль целостности данных, соблюдение режима доступа к данным, а также предоставление данных в виде, удобном для навигации и поиска в автоматическом режиме. При этом архитектура единого хранилища мультимедийных данных из полевых этнографических экспедиций, а также последующая поддержка данной архитектуры осуществляется автоматически при помощи средств CMS.

Отметим, что CMS Omeka поддерживает взаимодействие только с реляционными СУБД. Реляционный подход к формированию единого хранилища мультимедийных данных задает более жесткую и менее адаптивную структуру данных, чем при онтологическом подходе. Сложный математический аппарат, лежащий в основе реляционного подхода, позволяет получить более высокую скорость работы, а единые хранилища такого типа лучше приспособлены для коллективного заполнения. При этом основной задачей человека с правами администратора при работе с единым хранилищем станет внимательное описание добавляемых мультимедийных объектов в соответствии с архитектурой хранилища данных.

В рамках дальнейших исследований планируется привлечение технических специалистов, а также этнографов, антропологов и фольклористов для проведения экспертной оценки полученного решения. На основе обратной связи будет осуществлена доработка и актуализация единого хранилища мультимедийных



данных из полевых этнографических экспедиций с целью его более полного соответствия предъявляемым требованиям.

### Литература

1. Чечулин А. А. Основные проблемы при формировании электронного каталога полевых данных // Маклаевские чтения. 2014.
2. Касаткина А. К., Чечулин А. А. Мультимедийная база данных по фольклору Ифугао (Филиппины): постановка задач и выбор решений // Всероссийская научно-практическая конференция «Малочисленные этносы в пространстве доминирующего общества: практика прикладных исследований и эффективные инструменты этнической политики». 2014. С. 312–317.
3. Левшун Д. С., Чечулин А. А. Постановка задачи построения единого хранилища мультимедийных данных из полевых этнографических экспедиций // Технические науки – от теории к практике. 2015. № 51. С. 25–30.
4. Станюкович М. В. Pilipinas Muna! Филиппины прежде всего! // Маклаевский сборник. 2011. С. 500–510.
5. Афонин А. А., Крейнес М. Г. Поиск образовательных информационных ресурсов: принципы, архитектура, реализация // Интернет-порталы: содержание и технологии. 2003. С. 584–634.
6. Левшун Д. С., Чечулин А. А. Сравнение подходов к построению баз данных для организации поиска материалов в едином хранилище мультимедийных данных из полевых этнографических экспедиций // Юбилейная XV Санкт-Петербургская Международная конференция «Региональная информатика». 2016. С. 330–331.
7. Rehberger D. Getting Oral History Online: Collections Management Applications // Oral History Review. 2013. № 1 (40). С. 83–94.

### References

1. Chechulin A. The main problems in the formation of the electronic catalog of field data // Maclay Readings. 2014.
2. Kasatkina A., Chechulin A. Multimedia database on the folklore of Ifugao (Philippines): Goal Setting and Selecting Solutions // All-Russian scientific practical conference «Ethnic Minorities in the Space of the Dominant Society: The Practice of Applied Research and Effective Instruments of Ethnic Politics». 2014. pp. 312–317.
3. Levshun S., Chechulin A. Problems of Construction a Single Repository of Multimedia Data from Field Ethnographic Expeditions // Tekhnicheskie nauki – ot teorii k praktike. 2015. No. 51. pp. 25–30.
4. Stanyukovich M. Pilipinas Muna! The Philippines is a Priority// Maclay Publications. 2011. pp. 500–510.
5. Afonin A., Kreines M. Search for Educational Information Resources: Principles, Architecture, Implementation // Web Portal: Content and Technology. 2003. pp. 584–634.
6. Levshun D., Chechulin A. Comparison of Approaches for the Construction of a Database for the Organization of Research Materials in a Unified Repository for Multimedia Data from Field Ethnographic Expeditions // Anniversary XV St. Petersburg International Conference “Regional Informatics”. 2016. pp. 330–331.
7. Rehberger D. Getting Oral History Online: Collections Management Applications // Oral History Review. 2013. № 1 (40). pp. 83–94.

**Левшун Дмитрий Сергеевич**

– младший научный сотрудник, СПИИРАН,  
Санкт-Петербург, 199178,  
Российская Федерация,  
levshun@comsec.spb.ru

**Levshun Dmitry**

– Research Assistant, SPIIRAS, St. Petersburg,  
199178, Russian Federation,  
levshun@comsec.spb.ru