

Веб-сайт как объект цифрового наследия: проблемы идентификация и систематизации

Н.В. Борисов, В.В. Захаркина, И.А. Мбого, Д.Е. Прокудин, П.П. Щербаков

Аннотация— Развитие информационно-коммуникационных технологий привело к появлению цифровых культурных форм, которые являются культурным наследием также, как и традиционные объекты культурного наследия. При этом современные тенденции развития культуры в информационном обществе порождают как оцифрованные объекты культурного наследия (вторичное цифровое наследие), так и различные цифровые объекты, не имеющие аналогов в реальном мире (первичное цифровое наследие). Динамичное развитие информационно-коммуникационных технологий поставило как актуальную проблему сохранения цифрового наследия, которую начали решать с начала XXI века как на международном, так и на национальных уровнях. Это связано, в том числе, с тем, что цифровые объекты существуют в различных форматах и, с развитием информационных технологий, могут становятся недоступными в связи с устареванием технологических решений, с прекращением поддержки определённых форматов данных, использованных при их создании и других причин. Такая ситуация ставит проблемы не только сохранения, но, прежде всего, выявления объектов цифрового наследия. Актуальными становятся вопросы классификации, описания и систематизации выявленных объектов. Решение этих проблем, в дальнейшем, позволит организовать и проводить постоянную систематическую деятельность по сохранению цифрового наследия и обеспечению эффективного доступа к нему. В данном исследовании, на основе анализа мирового опыта, предлагаются подходы к выявлению, описанию и систематизации объектов цифрового наследия. Особое внимание уделяется такому цифровому объекту как веб-сайт (информационная система), что связано с его сложной структурой и зависимостью от многих факторов. Рассматривается возможная роль созданного в Санкт-

Петербургом государственном университете «Центра по изучению, сохранению, реставрации и актуализации объектов культурного наследия» в теоретической, методической и практической деятельности в области описания и сохранения цифрового наследия.

Ключевые слова— цифровое наследие, веб-сайт, сохранение, идентификация, систематизация, описание, метаданные, Dublin Core, DSpace.

I. ВВЕДЕНИЕ

Развитие цивилизации по технократическому пути и переход к информационному обществу, базирующемуся на всестороннем использовании информационно-коммуникационных технологий, привели к развитию цифровых методов и технологий обеспечения сохранения культурного наследия. Как основной подход начинает использоваться создание электронных (цифровых) копий объектов культурного наследия – оцифровываются печатные издания и старые фотографии, создаются трёхмерные цифровые изображения музейных экспонатов и т.п.

В сотрудничестве историков, искусствоведов, археологов и специалистов информационных технологий рождаются цифровые модели и реконструкции объектов материального культурного наследия. При этом используется широкий спектр различных технологий:

- виртуальная и дополненная реальность (VR/AR);
- трёхмерное моделирование (3D-modeling);
- видео360 (video360);
- клиент-серверные технологии доступа к цифровому культурному наследию
- и др.

Зачастую создаются целые программно-аппаратные комплексы или информационные системы, обеспечивающие не только хранение созданных цифровых культурных объектов, но и доступ к ним по сети Интернет.

То есть, можно констатировать, что на сегодняшний день одним из основных методов сохранения культурного наследия является перевод его в цифровую форму. Такой подход позволяет не только сохранять объекты культурного наследия, но и увеличивает их доступность. Это происходит за счёт создания: цифровых коллекций произведений искусства и музейных экспонатов; виртуальных музеев и экскурсий; трёхмерных моделей. При этом создаются такие решения, которые обеспечивают доступность культурного наследия через сеть Интернет любым пользователем.

Статья получена 26 ноября 2023.

Борисов Николай Валентинович, доктор физико-математических наук, профессор, Санкт-Петербургский Государственный Университет, заведующий Кафедрой информационных систем в искусстве и гуманитарных науках факультета искусств, ORCID 0000-0002-1671-5525 (n.borisov@spbu.ru)

Захаркина Валентина Валентиновна, кандидат физико-математических наук, Санкт-Петербургский Государственный Университет, доцент, ORCID 0000-0002-4950-2410 (v.zaharkina@spbu.ru)

Мбого Ирина Анатольевна, Санкт-Петербургский Государственный Университет, старший преподаватель, ORCID 0000-0002-5942-7143 (i.mbogo@spbu.ru)

Прокудин Дмитрий Евгеньевич, докт. филос. наук, доцент, Санкт-Петербургский государственный университет, ORCID 0000-0002-9464-8371 (e-mail: d.prokudin@spbu.ru).

Щербаков Павел Петрович, кандидат физико-математических наук, старший научный сотрудник, Санкт-Петербургский Государственный Университет, доцент, ORCID 0000-0003-1158-7460, (p.scherbakov@spbu.ru)

Статья подготовлена по итогам выступления на Международной объединённой конференции «Интернет и современное общество» (IMS-2023).

Эти процессы, связанные с динамичным развитием информационно-коммуникационных технологий, привели в человеческом обществе к осознанию наличия проблемы сохранения культурного наследия. Прежде всего это связано с тем, что многочисленные цифровые объекты и технологические разработки подвержены различного рода угрозам. Поэтому цифровые объекты и технологические решения приобретают статус цифрового наследия. Такое осознание привело к принятию на международном уровне в 2003 году Хартии о сохранении цифрового наследия [1]. Так, в статье 3 «Угроза утраты» отмечается, что к факторам, которые способствуют утрате цифрового наследия, *«относится устаревание оборудования и программ, обеспечивающих доступ к цифровым материалам, неопределенность в вопросах ресурсного обеспечения, ответственности и методик обеспечения сохранности и сохранения, отсутствие соответствующих законодательных актов».*

В соответствии с Хартией о сохранении цифрового наследия можно выделить несколько основополагающих направлений деятельности в этой области:

- обеспечение преемственности цифрового наследия (статья 5). Сохранение цифрового наследия является процессом непрерывным и требует «принятия мер на протяжении всего «периода жизни» цифровой информации — от создания до получения доступа. Процесс долгосрочного сохранения цифрового наследия начинается с разработки надежных систем и процедур, способных обеспечить аутентичное и устойчивое воспроизведение цифровых объектов»;
- обеспечение преемственности цифрового наследия должно решать проблему повторного использования объектов цифрового наследия, то есть обеспечивать их доступность на уровне восприятия. Решение этой проблемы заключается не только в необходимости, наряду с сохранением визуальных цифровых объектов, созданных по технологиям компьютерной графики, компьютерной анимации, 3D, VR, AR и т.п., но и обеспечивать сохранение и доступ к исходным материалам (программный код, наборы данных, исходные файлы в соответствующих форматах) для возможности их повторного использования, модификации и адаптации к развивающимся технологиям. Данную проблему не решают технологии, основанные на резервном копировании, например, сайтов, так как это не позволяет, в случае утери, воспроизвести сайт, основанный технологиях баз данных или на использовании технологий отображения различных мультимедийных объектов (например, видео360, панорамное фото с интерактивными элементами, 3D-объекты или объекты VR/AR). Другим немаловажным аспектом данной проблемы является динамичное развитие информационно-коммуникационных технологий, что ведёт к устареванию и прекращению поддержки и развития устаревшего программного обеспечения и форматов представления цифровых данных. Например, как нами отмечалось ранее, современные

браузеры не поддерживают воспроизведение видео и анимации в форматах FLV и SWF [2], которые использовались ранее в проектах по созданию цифрового культурного наследия;

- определение объектов цифрового наследия (статья 7). Действительно, тотальная цифровизация всех без исключения видов человеческой деятельности порождает бесчисленные цифровые объекты и информационные системы. Далеко не все из них представляют собой ценность на культурном уровне. Поэтому необходимо производить отбор тех объектов, которые подлежат сохранению как цифровое наследие. При всём многообразии цифровых форм представления *«главными критериями при определении того, какие цифровые материалы следует сохранять, должны быть их значимость и их непреходящая культурная, научная, документально подтвержденная или иная ценность».* При этом не должно быть необоснованных решений, касающихся отбора. Они должны приниматься *«подотчетным образом и на основе определенных принципов, политики, процедуры и стандартов».*

II. ВЫЯВЛЕНИЕ И ИДЕНТИФИКАЦИЯ ОБЪЕКТОВ ЦИФРОВОГО НАСЛЕДИЯ

Одним из основополагающих видов деятельности в области сохранения цифрового наследия является определение объектов, которые следует учитывать в качестве цифрового наследия. Анализ состояния нормативной базы и научных публикаций позволяет констатировать, что в России сохранение цифрового наследия на государственном уровне не организовано. Так на уровне государства не была ратифицирована «Хартия о сохранении цифрового наследия», принятая ЮНЕСКО ещё в 2003 году. В федеральном проекте «Цифровая культура», входящем в структуру национального проекта «Культура», паспорт которого утвержден президиумом Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам (протокол от 24 декабря 2018 года № 16, реализуется в соответствии с указами Президента Российской Федерации от 7 мая 2018 года № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года» и от 21 июля 2020 года № 474 «О национальных целях развития Российской Федерации на период до 2030 года»), к направлению по сохранению культурного наследия в цифровой форме отнесены следующие задачи:

- оцифровка и включение в Национальную электронную библиотеку книжных памятников;
- оцифровка фильмовых материалов на цифровых носителях Госфильмофонда России.

Необходимо отметить, что на общенациональном уровне целенаправленная работа по выявлению объектов цифрового наследия не ведется, поэтому широкомасштабная деятельность в направлении выявления объектов цифрового наследия носит скорее инициативный характер.

Наиболее естественно эта задача может решаться в

рамках деятельности библиотек и музеев, так как в их обязанность давно включена работа по оцифровке документов и изданий, которые учтены и каталогизированы. Деятельность по оцифровке печатных изданий проводится отечественными библиотеками уже более двух десятилетий. На отечественном уровне эта деятельность рассматривается в контексте мировых тенденций и европейских инициатив [3]. При этом руководство библиотек понимает необходимость сохранения культурного наследия в цифровой форме и библиотеки на регулярной основе занимаются этой деятельностью [4, 5, 6]. Однако, отсутствие координации в этой области приводит к тому, что отсутствуют единые методики, стандарты и технологии. Так, например, подавляющее большинство библиотек за формат цифрового представления цифровых копий печатных материалов принимают PDF (например, Российская государственная библиотека, Российская национальная библиотека, Научная педагогическая библиотека им. К.Д. Ушинского и др.), но наряду с ним встречаются и другие форматы. Созданный и развиваемый Библиотекой Российской академии наук проект «Научное наследие России» (<http://www.e-heritage.ru>) призван решать задачу «сохранения научного наследия и создание условий его эффективного освоения», что «рассматривается как составная часть стратегии развития отечественной науки и, шире, развития отечественной культуры» [7]. Представленные в этой электронной библиотеке публикации оцифрованы в формате PNG, то есть каждая страница хранится в отдельном графическом файле. Это затрудняет повторное использование оцифрованных публикаций - необходимо сначала скачать каждую страницу, а затем преобразовать их в один документ формата PDF. Для поиска по тексту необходимо будет ещё и распознать его.

Музеи также занимаются оцифровкой своих экспонатов. На государственном уровне созданы, наполняются и доступны государственные каталоги национальных музейных фондов, которые аккумулируют информацию и фотографии музейных экспонатов государственных музеев [8, 9]. Также и сами музеи на своих сайтах предоставляют доступ к электронным каталогам своих коллекций, например, Государственный Эрмитаж (<https://collections.hermitage.ru>) и Русский музей (<https://www.rusmuseum.ru/collections/>). Есть и примеры представления культурного наследия в цифровой форме как на мировом, так и региональном уровнях:

- Мировая цифровая библиотека, созданная совместно ЮНЕСКО и Библиотекой конгресса США (<http://www.wdl.org/ru>);
- Европейская цифровая библиотека, созданная и поддерживаемая на уровне Европейского союза (<http://europeana.eu>).

Как видим, музеи оцифровывают свои коллекции, а проекты и фонды разного уровня поддерживают агрегацию наиболее значимых объектов культурного наследия в цифровой форме. При этом, иногда, качество представленных цифровых копий оставляет желать

лучшего - в основном это фотографии и зачастую в низком качестве или разрешении. Если для плоских объектов (таких как рукописные и печатные документы и издания, картины, ковры и т.п.) оцифрованное фото приемлемо, то для объёмных экспонатов такое представление не позволяет воспринять их во всей полноте.

Есть и другие подходы к оцифровке культурного наследия. С 2012 года в рамках Программы поддержки политики Европейской комиссии в области ИКТ был реализован пилотный проект 3D-ICONS (<http://3dicons-project.eu>). Он объединил партнеров со всей Европы, обладающих соответствующим опытом оцифровки архитектурных и археологических памятников и зданий в 3D. В ходе реализации проекта был создан полный цикл производства 3D-копий археологических памятников и исторических зданий. При этом были созданы 3D-модели с описанием как самих оцифрованных объектов, так и технологии создания этих моделей. Сами модели и их описания доступны через интерактивную карту (<http://3dicons.ceti.gr/index.php> [10, 11, 12]). Кроме этого, через агрегатор CARARE (<https://pro.carare.eu/en/>) эти цифровые объекты размещаются в европейской цифровой библиотеке Европеана (<http://europeana.eu>), что создаёт условия для их доступности [13, 14].

Вышесказанное относится к вторичным объектам цифрового наследия, являющихся оцифрованными копиями реальных объектов. Совсем другая ситуация складывается с первичным цифровым наследием, которое изначально создаётся в цифровой форме. Информация о нём не собирается, не аккумулируется и не систематизируется. Об этих объектах знают только разработчики, заказчики и пользователи. Зачастую цифровые объекты создаются в рамках исследований, исторических и иных реконструкций, деятельности отдельных энтузиастов и творческих коллективов. На основе собственного опыта и анализа цифрового информационного пространства можно перечислить следующие основные источники информации о цифровом наследии в России:

- сведения о разработках по проектам Российского государственного научного фонда (РГНФ), который с 1994 по 2016 год поддерживал разработку программного обеспечения информационных систем для научных исследований в области гуманитарных наук, способствующих распространению гуманитарных научных знаний в обществе (тип конкурса «в»). Фонд в 2016 году был упразднён путём присоединения к Российскому фонду фундаментальных исследований (РФФИ) и данный тип конкурса перестал существовать. За годы реализации проектов по этому типу конкурса было создано большое число информационных систем, сайтов, 3D-моделей, виртуальных реконструкций, которые в полной мере можно считать цифровым наследием. Источниками информации об этих объектах являются сайты университетов и научных учреждений (в которых выполнялись проекты), отчёты учёных и исследователей - участников проектов, сайты

проектов в сети Интернет (при условии их функционирования);

- сведения об инициативных или заказных разработках сайтов организаций культуры (музеев, библиотек, театров и т.п.);
- сведения о проектах, созданных членами Российской ассоциации цифровых гуманитарных наук (<http://dhrussia.ru>);
- сведения с официальных сайтов и порталов Министерства культуры, в том числе из базы открытых данных (<https://opendata.mkrf.ru>);
- сайт Российского комитета Международного совета музеев (ИКОМ России, <https://icom-russia.com>). На сайте представлена информация о деятельности Совета по цифровому развитию музеев, который координирует взаимодействие членов ICOM в России, в том числе, по цифровым объектам культуры. Помимо этого на сайте можно получить список российских музеев, которые являются членами Международного комитета по аудиовизуальным, новым мультимедийным технологиям и социальным медиа (International Committee for Audiovisual, New Technologies and Social Media, AVICOM). В настоящее время это 26 музеев, которые, в том числе участвуют в деятельности по применению мультимедийных технологий в сохранении культурного наследия;
- сведения с сайтов разработчиков программного обеспечения и информационных систем, которые представляют собой цифровое наследие;
- сведения из научных публикаций, отражающих разработки цифровых объектов. В электронной форме эти публикации доступны, например, в Научной электронной библиотеке (<https://elibrary.ru>);
- информация из электронных научных журналов, публикующих мультимедийные материалы (например, электронный мультимедийный журнал «Культура и технологии», <http://cat.ifmo.ru/>);
- сведения из материалов научных конференций, где докладывались результаты разработок в области культуры. Это могут быть, например, сайты конференций «Интернет и современное общество» (<https://ojs.itmo.ru/index.php/IMS>, <https://ojs.itmo.ru/index.php/ISESCTF>), Electronic Imaging & Visual Arts, EVA Saint-Petersburg (<http://evaspb.ifmo.ru>) и «Научный сервис в сети интернет» (<https://keldysh.ru/abrau/>);
- результаты целевых исследований инфосферы цифровых гуманитарных наук, проводимых с целью разработки информационно-справочной системы по этой инфосфере [15, 16].

Конечно, динамичное развитие цифровых технологий постоянно порождает всё новые объекты, претендующие стать цифровым наследием. И не всегда оперативно информация о них находит своё отражение в общественном или научном дискурсе. Поэтому необходимо создание сайта с общедоступной веб-формой для заявительного сбора сведений о цифровых объектах от заинтересованных лиц.

Такая деятельность уже начала осуществляться в рамках инициативы для исходного кода открытого программного обеспечения. Одной из важных для

сохранения цифрового наследия является инициатива Software Heritage (<https://www.softwareheritage.org>). Этот проект был запущен в 2016 году компанией Inria (Франция) и направлен на сбор, сохранение и обеспечение доступности исходного кода всего когда-либо написанного программного обеспечения. Основатели проекта исходят из того, что исходный код программного обеспечения является ценной частью человеческого наследия. Software Heritage - это открытая некоммерческая инициатива, миссия которой заключается в обеспечении того, чтобы этот драгоценный массив знаний сохранялся с течением времени и был доступен всем [17, 18]. Как отмечает основатель и руководитель проекта R. Di Cosmo, программное обеспечение и, в частности, исходный код должны быть заархивированы, снабжены ссылками и описаны для создания стабильного и долговечного корпуса научных знаний [19].

III. ОПИСАНИЕ И КАТАЛОГИЗАЦИЯ ОБЪЕКТОВ ЦИФРОВОГО НАСЛЕДИЯ

Выявление объектов цифрового наследия ставит следующую важную задачу - описание этих объектов и дальнейшую каталогизацию этих описаний. С одной стороны, описание должно быть жёстко унифицированным, так как при создании каталога необходимо ориентироваться на его электронное представление с возможностью поиска по определённым критериям. С другой - описание должно содержать сведения, отражающие всё разнообразие технологических решений и форматов данных, а иногда и уникальность объекта. При этом унификация должна обеспечить описание как вторичных, так и первичных объектов цифрового наследия.

В рамках деятельности Международного совета музеев (International Council of Museums, ICOM) разработано описание музейного объекта Object ID, который принят с 1997 года в качестве стандарта описания объектов культуры [20, 21]. Подавляющее число музеев во всём мире используют его. В России также используется этот стандарт и, прежде всего, при описании музейных объектов в электронных каталогах. Этот стандарт описывает реальные объекты и только отчасти подходит для описания вторичных объектов цифрового наследия.

С библиотечном сообществе как основной стандарт для описания изданий, в том числе представленных в цифровой форме (например, оцифрованные бумажные издания), используется MARC, различные интерпретации которого реализованы в подавляющем большинстве электронных каталогов библиотек.

На основе этих общепринятых стандартов возможно создание и стандарта для описания объектов цифрового наследия. Очевидно, что в этом описании должна как обязательная присутствовать следующая информация:

- название цифрового объекта
- описание цифрового объекта
- программное обеспечение, использованное для создания цифрового объекта

- описание технологий, использовавшихся при создании цифрового объекта
- описание аппаратных и программных требований, необходимых для воспроизведения (использования) цифрового объекта
- тип цифрового объекта
- формат представления цифрового объекта
- носитель цифрового объекта
- ссылка на сайт в сети Интернет, где представлен этот объект (при необходимости)
- автор(ы) цифрового объекта
- дата или период создания цифрового объекта
- права на цифровой объект
- место размещения цифрового объекта

То есть описание вторичного цифрового наследия должно содержать информацию как о реальном объекте, так и о его цифровом представлении.

Помимо этого, описание должно содержать информацию, необходимую для однозначного описания цифрового объекта, например, информация о классе цифровых объектов - первичным или вторичным цифровым наследием они являются. Если такой объект связан с реальным объектом культурного наследия (существующим или утерянным), то необходимо дать описание такого объекта для того, чтобы поместить его в культурно-исторический контекст.

Другой важной задачей является возможность использования описания выявленных объектов цифрового наследия. Наиболее эффективно можно использовать такую информацию, если она представлена в машиночитаемой форме. То есть предлагается создать соответствующий электронный каталог (базу данных) с доступом к нему по сети Интернет через соответствующий веб-интерфейс. Тогда описание объектов цифрового наследия, представляющее собой набор метаданных, будут являться записями в базе данных. Также такой каталог должен обладать классификатором (например, по типам объектов или по форматам их представления) и поисковой машиной по полям описания для быстрого отбора описаний цифровых объектов по определённым параметрам.

Для обеспечения возможности поиска, интероперабельности, установления происхождения, управления правами и повторного использования цифровых объектов разрабатываются и используются стандарты метаданных, в том числе и встроенных в сами цифровые объекты [22]. В настоящее время для описания произведений искусства и архитектуры используются различные стандарты метаданных, классификаторы и контролируемые словари [23]. Наиболее распространёнными для описания цифровых объектов культуры являются стандарты Dublin Core, VRA, MODS, CDWA, IPTC, LIDO [24, 25]. Есть и специфические схемы метаданных. Так, например, в сервисе агрегирования в цифровую европейскую энциклопедию Europeana археологического и архитектурного цифрового наследия (включая изображения, тексты и 3D-объекты) разработана и используется схема метаданных CARARE [26]. Её разработчиками особое внимание уделяется решениям

для публикации 3D-моделей и виртуальной реальности.

Под эгидой и при финансовой поддержке Федерации цифровых библиотек (Digital Library Federation) разработан и в настоящее время действует «Стандарт кодирования и передачи метаданных» (Metadata Encoding and Transmission Standard, METS), представляющий собой спецификацию кодирования и передачи данных на языке XML для кодирования метаданных, необходимых для обработки цифровых объектов в репозитории и обмена такими объектами. В рамках разработки проекта создана схема XML DTD, которая определила стандарт описания цифрового объекта для кодирования административных, описательных, структурных его метаданных и основного содержания [27].

Для структурированного представления описания цифрового объекта предлагается использовать схему метаданных Дублинского ядра (Dublin Core Metadata Element Set, DCMES) [28]. Это стандарт общего назначения, который активно используется библиотеками для описания коллекций изображений. С начала 2000-х годов стандарт метаданных Dublin Core начал использоваться в музейной практике. Dublin Core используется для описания объектов при создании виртуальных музеев [29]. Схема метаданных достаточно давно адаптирована для описания произведений живописи [30] и внедрена в электронные каталоги различных музеев по всему миру. Соответственно, она используется для описания цифровых изображений не только картин, но и любых музейных экспонатов, представляющих собой трёхмерные объекты. При этом также создаются не только цифровые изображения этих объектов, но и их 3D-модели, для которых также разрабатываются схемы метаданных, что позволяет обеспечивать доступ к самим 3D-моделям [31, 32]. Набор метаданных является расширяемым [33], в связи с чем его можно применять для описания разнообразных цифровых объектов как вторичного, так и первичного цифрового наследия. Это достигается за счёт использования квалификаторов, которые составляют второй уровень метаданных и уточняют значения элементов [34, 35]. Предлагаются разработки, в которых используется расширенный набор метаданных XDC-SC (Extended Dublin Core for Software Components) схемы Dublin Core для описания программного обеспечения, обеспечивающий извлечение информации стандартными средствами поисковых систем или инструментами работы с XML [36].

В целом, использование стандарта описания метаданных Dublin Core позволяет создать каталог, обеспечивающий автоматизированный обмен метаданными с другими информационными системами, содержащими информацию о цифровом наследии, на основе протокола обмена метаданными OAI-PMH [37]. Такое объединение позволит как агрегировать в каталоге метаданные о цифровом наследии из других каталогов и коллекций, так и предоставлять метаданные внешним системам. Для внешних систем, не поддерживающих стандарт Dublin Core, возможно

использовать для обмена данными формат XML [38, 39] и соответствующие модули экспорта-импорта. Также за основу обмена метаданными между различными информационными системами может быть положена Концептуальная эталонная модель CIDOC (CIDOC Conceptual Reference Model, CIDOC CRM), которая разработана и развивается рабочей группой по стандартизации документации Международного комитета документации (International Committee for Documentation, CIDOC) Международного совета музеев (International Council of Museums, ICOM). Это расширяемая семантическая модель-посредник, являющаяся средством для объединения разнородной информации по культурному наследию, публикуемой музеями, архивами и библиотеками [40]. Уже разработаны карты соответствия схемы этой модели и различных стандартов и схем представления метаданных [41]. Такой подход нацелен на создание единого информационного пространства («экосистемы») цифрового наследия.

IV. ВЕБ-САЙТ КАК ОБЪЕКТ ЦИФРОВОГО НАСЛЕДИЯ: ТЕХНОЛОГИЧЕСКИЙ И ОРГАНИЗАЦИОННЫЙ АСПЕКТ, ВОЗМОЖНЫЕ МОДЕЛИ

Как отмечалось выше, многие существующие инициативы, направленные на сохранение цифрового культурного наследия, и их соответствующие реализации справедливо выделяют разные аспекты. Эти аспекты неразрывно связаны, однако требуют разных технологических решений. Не менее важными оказываются организационные моменты.

Не претендуя на глобальные постановки проблем и предложение соответствующих решений, выделим для дальнейшего рассмотрения вопросы сохранения объектов цифрового культурного наследия, которые были опубликованы в формате веб-сайтов. Отметим, что сохранению могли бы подлежать как сайты в целом (хотя бы на уровне возможности отобразить их исторический вид), так и их контент (в частности, мультимедийный) [2]. Отдельный вопрос: сохранение программного кода и соответствующей серверной инфраструктуры, обеспечивающей работоспособность ресурса.

Кратко можно констатировать, что вопросы сохранения объектов цифрового культурного наследия, которые были опубликованы в формате веб-сайтов, лишь отчасти получили своё решение к настоящему моменту:

- в течение последних десятилетий создавались и развивались инициативные архивы, сохраняющие «мгновенные снимки веб-страниц» на текущую дату. Одним из наиболее известных проектов подобного рода, предоставляющего корректные данные по гигантскому множеству сайтов в течение десятилетий, является веб-ресурс <https://archive.org>. Он является бесценным источником информации как для ретроспективных исследований развития WWW (стилистика исторических ресурсов, интерфейсные решения и т.д.), так и для обеспечения образовательного процесса ряда вузовских программ. Отметим, что на сегодняшний

день [web.archive](https://web.archive.org) с его уникальным содержимым остаётся свободным ресурсом. В его манифесте указано: «The Internet Archive, a non-profit, is building a digital library of Internet sites and other cultural artifacts in digital form»;

- инициативы, направленные на сохранение контента, не ассоциируют его с теми веб-сайтами, на которых контент был впервые опубликован и с авторами. Отметим, что многие платформы нацелены на представление цифровых данных, изначально обезличенных (если, например, говорить об авторах 3D моделей, авторах иллюстраций и т.д.);
- инициативы последних лет, направленные на сохранение программного кода как объекта культурного наследия, серьёзны и интересны. Однако, мы сужаем область рассмотрения до тех программных решений, которые обеспечивают функционирование ресурсов в формате веб-сайтов. Таким образом, встают вопросы сохранения не только клиентского, но и серверного программного кода, а также структур данных и соответствующей серверной инфраструктуры. Проблема сохранения потенциальной возможности восстановления функционирования сайта в целом (а именно, архивирование данных, допускающих его восстановление из резервных копий файловой структуры и базы данных), решается штатным образом в период функционирования сайта на определённом хостинге. Что же хранить в потенциальном репозитории на случай прекращения функционирования сайта? Эта проблема крайне сложна, она не сводится к сохранению программных кодов.

А. Возможные модели

Фактически, основные действия профессиональных коллективов, направленные на сохранение цифрового культурного наследия, могут быть реализованы в двух основных моделях.

Первый подход предполагает поиск готовых платформ, которые представляются достаточно серьёзными с точки зрения их статуса (государственный, международный с поддержкой профессиональных сообществ и т.д.). При этом немаловажной является функциональность, предоставляемой этими платформами. Как правило, сейчас интересен вариант с онлайн представлением в формате Web и – вторично – в формате мобильного приложения. Результат: глобальные платформы представляют безликий общедоступный контент, который лишь частично и однобоко отражает предметную область, даже разнородные данные представлены единообразно. При этом у профессиональных коллективов могут быть свои прекрасные сайты, однако они считают нужным публикацию некоторых материалов на некоей «общей платформе». А на этой платформе, в частности, приводится ссылка на оригинальный и своеобразный сайт. Условно эту модель реализации можно назвать «сверху вниз».

Второй подход очень распространён и продуктивен.

Профессиональные коллективы или сообщества создают, развивают и поддерживают на протяжении длительного времени (иной раз десятков лет) собственные веб-ресурсы. Кстати, именно такие ресурсы зачастую могут быть отнесены к объектам цифрового культурного наследия. Может быть ценным как контент, так и программные решения (многие ресурсы и сами по себе представляют не только содержательную, но и историческую ценность). Включить контент таких ресурсов в некий общий репозиторий, требующий строгой формализации – сложная задача. Однако, надо задуматься о том, каким образом подобные проекты – с сохранением их своеобразности и с учетом зачастую бесценного контента – хотя бы каталогизировать на уровне уважаемых в профессиональной среде и долгоживущих онлайн ресурсов. Именно такие ресурсы и следует запланировать в дальнейшем, а далее обеспечить их посещаемость, подтверждение статуса, формальную ценность регистрации и т.д. Условно эту модель реализации можно назвать «снизу вверх».

1) Модель «сверху вниз»

На определённом этапе, обусловленном развитием технологий и инфраструктуры, появилась принципиальная возможность создавать веб-платформы (либо общедоступные, либо с частично ограниченным доступом), которые изначально обладают значительной функциональностью и дают возможность зарегистрированным пользователям публиковать свои материалы.

Некоторые подобные платформы обрели популярность и – как результат – в настоящее время продолжают развиваться. Важно, что именно активизация контентного пополнения побуждает владельцев платформ совершенствовать функциональность, инициировать применение новых технологических решений, работать над интерфейсными решениями. Получив стимулы, платформы совершенствуются, обретают определённый формальный статус и тем самым привлекают новых пользователей.

Можно привести много позитивных примеров таких платформ. Это некоторые социальные сети (а ведь были многочисленные попытки предложить практически идентичные по функциональности аналоги, но они по разным причинам популярности не обрели). Это, несомненно, Википедия: в ближайшее время ничто не может побудить к публикации своих статей «энциклопедического характера» на иной платформе.

Отдельно отметим ресурс <https://habr.com>, публикующий на протяжении многих лет статьи по тематике в сфере ИТ. Возможна также публикация программного кода, отсылки к примерам его применения (что позволяет нам отнести этот ресурс к платформам, способствующим сохранению цифрового культурного наследия). Статью может опубликовать практически любой зарегистрированный пользователь и, как следствие, уровень материалов весьма разный. Однако, существует система внутреннего ранжирования (основанная, в частности, на отзывах и оценке квалифицированных модераторов). Ресурс активно

развивается и популярен не только у тех, кто хотел бы опубликовать статью, он уважаем на уровне профессионального сообщества. Недаром по целому ряду поисковых запросов на первых позициях мы видим ссылки на habr.com (и, как правило, находим именно там адекватные материалы).

Отметим также, что есть много платформ относительно узкой направленности, позволяющих публиковать материалы определённых форматов: например, 3D модели, фотографии, видео. Возможно, часть контента, опубликованного на ряде таких платформ (например, <https://sketchfab.com>, <https://render.ru> и многих других) могла бы рассматриваться как подлежащая сохранению, но отсутствуют соответствующие механизмы.

Проблемы универсальных платформ, представляющих объекты цифрового культурного наследия:

- унификация представления контента. Материалы, представленные на универсальных платформах, оформлены единообразно. В ряде случаев это чревато потерей индивидуальности. Например, элементы музейных коллекций, опубликованные на универсальном веб-сервисе, теряются среди многих подобных (пример: Государственный каталог Музейного фонда Российской Федерации <https://givc.ru>). Возможно, многие материалы подобного рода значительно ярче воспринимаются на отдельном сайте. Вспомним про «эффект небольшой экспозиции»: каждый её экспонат будет рассмотрен с интересом, но в контексте большого музея они бы потерялись;
- коммерциализация. Любой проект, инициированный профессиональным сообществом, потенциально может быть продан вместе с доменным именем. Что при этом будет с контентом, зависит от дальнейшей политики покупателя;
- возможная политическая ангажированность. Текущие держатели ресурса могут ограничить доступ по собственным политическим мотивам либо под давлением извне. Вспомним, например, как в декабре 2015 года стал недоступным российский сегмент электронной библиотеки Европеана (<https://www.europeana.eu>);
- привлечение к пополнению контента и формальная значимость публикации. Что может побудить к публикации своих материалов на универсальной платформе? Возможно, отсутствие собственного ресурса (но в настоящее время это весьма редкая ситуация). Желание зафиксировать своё присутствие? Но в этом случае платформа должна получить должное признание на уровне профессиональных сообществ, а также обеспечить соответствующий статус публикаций и их признание на формальном уровне;
- статус платформы. Платформа может быть востребована на разных уровнях: обеспечение текущей деятельности в достаточно узкой предметной области, на региональных и национальных уровнях, на уровне международного сотрудничества в определённой предметной области;

— перспективы развития и сохранность контента. Этот крайне важный аспект должен рассматриваться с учётом ранее сформулированных проблем.

2) 5.1.2. Модель «снизу вверх»

Именно в такой модели проходит деятельность многочисленных отдельных групп и профессиональных сообществ.

Профессиональные коллективы разрабатывают собственные ресурсы, направленные на сохранение культурного наследия. Как правило, контент имеет мультимедийный характер (тексты, фотоматериалы, видео, 3D модели), материалы отдельных сайтов публикуются в соответствии с определённой структурной, стилистической и интерфейсной концепцией, зачастую оригинальной. Подобные проекты могут быть посвящены узкой предметной области либо представлять материалы широкого характера. Существенно, что многие сайты «имеют своё лицо» и сами по себе могли бы быть рассмотрены как синтетические объекты цифрового культурного наследия.

На следующем уровне профессиональные сообщества выдвигают свои манифесты и приглашают к сотрудничеству. Как правило, разрабатываются онлайн платформы, обеспечивающие деятельность в рамках инициативы (каталоги, веб-сервисы узкого характера). Те коллективы, которые присоединились к инициативе, налаживают сотрудничество, выдвигают предложения по организации совместной деятельности, совершенствуют технологические решения общей платформы. Существенно, что деятельность не регулируется свыше и протекает на условиях *самоорганизации*. Как правило, на уровне отдельных инициатив естественным образом вырабатываются внутренние принципы рецензирования и отбора материалов для публикации на общей платформе.

Наконец, некоторые инициативы могут обрести значительный вес в соответствующей предметной области за счёт качества публикуемых материалов и предлагаемых решений. Тем самым привлекаются новые профессиональные коллективы, объём и качество публикаций на созданных онлайн платформах растёт, сами платформы обретают высокий статус на уровне предметной области. Возникает стимул для обеспечения дальнейшей поддержки ресурсов, их технологического совершенствования и обеспечения сохранности контента.

Отметим, что *ряд рисков* при деятельности в рамках такой модели может быть существенно снижен в сравнении с моделью «сверху вниз». Так, например, не стоит остро проблема унификации контента: многие участники инициатив имеют свои веб-ресурсы и развивают их в соответствии со своими задачами и своими представлениями о принципах отображения контента. Проблемы возможной коммерциализации и привлечения к сотрудничеству также не являются критически значимыми (по крайней мере, до обретения инициативой высокого статуса).

Основные проблемы, таким образом, фиксируются на «нижнем уровне»: это поддержка и обеспечение функционирования сайтов отдельных проектов и

соответствующего контента. Ситуация усугубляется тем, что в абсолютном большинстве случаев отсутствует финансовая поддержка и формальное признание результатов деятельности.

В. Проблемы и планируемые направления деятельности

На Кафедре информационных систем в искусстве и гуманитарных науках СПбГУ (в сотрудничестве с представителями различных предметных областей и организаций) на протяжении двух десятилетий были разработаны десятки веб-ресурсов, посвященных различным аспектам сохранения культурного наследия [2]. Большинство проектов выполнены в рамках грантов (РФФИ, РГНФ, РНФ, Фонд «Русский мир»), дальнейшая финансовая поддержка отсутствует. Авторы данной статьи работали над технологической реализацией проектов и именно вследствие этого остро воспринимают текущие проблемы, связанные с обеспечением работоспособности существующих ресурсов.

Лишь некоторые проекты активно развиваются, поэтому естественным образом для них совершенствуются технологические решения. Большинство же созданных веб-ресурсов имеют «монографический» характер, они достаточно полно освещают отдельную тему и не предполагают дальнейшего развития. Парадоксальным образом именно их поддержка представляет определённую проблему.

Рискнём сказать, что как отдельные элементы мультимедийного контента, так и сами ресурсы в целом могли бы быть отнесены к объектам цифрового культурного наследия. Программные же решения в данном случае лишь обеспечивают корректное функционирование ресурсов.

Возвращаясь к предложенным выше моделям, отметим, что наша деятельность относится к **базовому уровню модели «снизу вверх»**, на котором выявляются характерные проблемы, связанные с поддержкой отдельных веб-ресурсов.

Основные **проблемы** связаны с неизбежным устареванием исходных технологических и программных решений. В нашем случае критичным для корректного отображения в современных браузерах оказалось прекращение поддержки технологии Flash (форматы SWF и FLV). Как результат – на целом ряде сайтов, лишённых постоянной поддержки, не отображаются мультимедийные элементы, являющиеся зачастую неотъемлемой частью общего контента. Впрочем, для многих ресурсов не менее критичными могут стать иные факторы, рассмотрение которых выходит за рамки данной статьи.

Отметим определённое сходство ситуаций и предлагаемых направлений деятельности в области сохранения материальных и цифровых объектов культуры. Объекты материальной культуры, в зависимости от степени их сохранности, подлежат дальнейшему сохранению, консервации, реставрации, открытию для свободного доступа и т.д. Практически то же можно сказать и о цифровых объектах.

Понимание общих тенденций развития технологий создания веб-ресурсов и зафиксированные частные проблемы позволяет сформулировать ряд возможных направлений деятельности на уровне сохранения отдельных сайтов и их контента:

- восстановление работоспособности отдельных элементов (в основном, мультимедийных). Например, при наличии упомянутой выше «проблемы Flash» необходимо перевести видео и анимации в современный формат. Ситуация усугубляется тем, что в большинстве случаев отсутствуют исходные форматы;
- обеспечение хранения мультимедийных данных в исходном формате, позволяющим дальнейшее редактирование и последующий экспорт. Внешнему пользователю такие форматы недоступны, однако именно они могут позволить в дальнейшем восстановить работоспособность, нарушенную вследствие неизбежного развития технологий;
- редизайн сайта с целью придать ему вид, соответствующий современной стилистике;
- полный реинжиниринг сайта и перевод его на современный стек технологий. Такие действия могут быть обоснованы веской необходимостью радикально расширить функциональность: например, осознанием необходимости контентного развития с доступом через веб-интерфейс;
- организационная деятельность. Включение сайта в сферу отдельных инициатив профессиональных сообществ либо в одну из структур, работающих по модели «сверху вниз».

Коллектив авторов данной статьи предполагает инициировать проекты, направленные в первую очередь на «реставрацию» существующих сайтов и восстановление работоспособности их мультимедийных элементов.

V. ОБСУЖДЕНИЕ И ВЫВОДЫ

Возвращаясь к обсуждению общих проблем, отметим, что инициатива Software Heritage направлена только на сохранение части цифрового наследия - программного обеспечения и, притом, на уровне исходного кода. В рамках реализации проекта Software Heritage создана соответствующая инфраструктура, позволяющая архивировать информацию об исходном коде программного обеспечения (<https://archive.softwareheritage.org>). Наполнение базы данных осуществляется на заявительной основе, то есть через заполнение экранной формы на сайте. Положительным является обеспечение модерации поданных заявок. Однако, возможности пополнения архива ограничены только открытым исходным кодом, который размещается в распределённых системах управления версиями (Git, Mercurial, Subversion, CVS, Bazaar). Исходный код размещается в них по открытым лицензиям, поэтому не возникает проблем как с доступом к нему, так и при его повторном использовании. Не архивируется коммерческое, а также свободно распространяемое программное обеспечение, которое доступно в скомпилированном виде (исполняемый код). Другой проблемой этого решения

является как отсутствие классификатора, так и расширенного поиска по полям базы данных. Это затрудняет поиск в архиве и, в целом, не позволяет эффективно осуществлять повторное использование кода.

Поэтому предлагаемый нами подход к созданию электронного каталога цифрового наследия на основе разработки классификатора и обладающего возможностью поиска по полям базы данных представляется более эффективным в использовании. Из проекта Software Heritage можно позаимствовать процедуру модерации поданных заявок, так как поданные сведения должны проходить верификацию и только после этого должно приниматься решение о размещении информации в электронном каталоге.

Мы уверены, что для описания объектов цифрового наследия в каталоге нет необходимости разрабатывать собственную схему метаданных, как предлагают некоторые исследователи [42, 43], а за основу взять схему метаданных Dublin Core, так как она достаточно широко используется как для описания различных коллекций, так и цифровых объектов в области культурного наследия. При этом схема является расширяемой, а для совместимости с другими стандартами метаданных разработаны таблицы соответствий элементов, например [28].

В качестве технологического решения для создания каталога мы предлагаем использовать свободно распространяемое программное обеспечение DSpace (<https://dspace.lyrasis.org>). Это одно из самых популярных решений для создания институциональных архивов открытого доступа. Так, например, по данным открытого каталога Registry of Open Access Repositories (http://roar.eprints.org/cgi/roar_search/advanced?software=dspace&type=institutional) на момент написания данной статьи на этом ПО работают 1930 зарегистрированных в каталоге институциональных архивов по всему миру, что составляет 49% от всех зарегистрированных в каталоге институциональных архивов. Кроме этого, как отмечают исследователи, в DSpace можно гибко настраивать реализацию схемы метаданных Dublin Core для создания цифровой системы архивирования музейных и культурных ресурсов [28]. При помощи этого решения производится интеграция информации о различных объектах, например музейной коллекции с архивными и библиотечными фондами [38]. Также в DSpace возможно интегрировать и другие схемы метаданных [28].

VI. ЗАКЛЮЧЕНИЕ

ЮНЕСКО в вопросах сохранения цифрового наследия особое внимание уделяет тому, что эта деятельность должна быть организована на национальном уровне (статья 10 Хартии). Это предполагает определение учреждений, «которые несут ответственность за координацию в вопросах сохранения цифрового наследия и необходимое ресурсное обеспечение. При определении задач и ответственности можно исходить из имеющегося распределения ролей и опыта». Организационная и координационная деятельность

предполагает привлекать к сотрудничеству по вопросам сохранения цифрового наследия, наряду с национальными библиотеками, архивами, музеями и иными публичными организациями в области культурного наследия разработчиков аппаратного и программного обеспечения, создателей, издателей, производителей и распространителей цифровых материалов. И как одно из важных направлений выделяется подготовка «*кадров и проведение научных исследований, обмен опытом и знаниями между заинтересованными учреждениями и профессиональными ассоциациями*». Без чего, в принципе, невозможно на долгосрочной основе осуществлять сохранение не только существующего цифрового наследия, но и разрабатывать подходы и методы для сохранения новых форм цифрового наследия, порождаемых перманентно развивающимися информационно-коммуникационными технологиями.

Следует отметить, что данный вид деятельности как целенаправленный и скоординированный в России в настоящее время находится в зачаточном состоянии. Проблемы выявления объектов цифрового наследия в России связаны с отсутствием как нормативной базы на государственном уровне, так и координации этой деятельности между отдельными инициативными группами, которые представляют собой в основном библиотеки и музеи. Созданный в Санкт-Петербургском государственном университете Центр по изучению, сохранению, реставрации и актуализации объектов культурного наследия согласно Положению, в том числе, занимается деятельностью по выявлению, изучению, сохранению объектов культурного наследия, создаваемых с использованием информационных технологий. Поэтому он может организовать деятельность по выявлению, систематизации и классификации цифрового наследия, созданного в нашей стране. Для каталогизации цифрового наследия возможно использовать существующий архив открытого доступа СПбГУ (<https://dspace.spbu.ru>), который реализован на ПО DSpace. Дальнейшая деятельность будет направлена на разработку классификатора и схемы метаданных для описания объектов цифрового наследия и интеграцию этих решений в открытый архив СПбГУ.

Отдельной задачей деятельности по сохранению цифрового наследия видится разработка проекта по описанию веб-сайтов как объектов цифрового наследия, нацеленного на возможность повторного использования в условиях постоянного технологического развития и связанных с ним проблем устаревания технологий.

БИБЛИОГРАФИЯ

- [1] Хартия о сохранении цифрового наследия (2003) // ООН. Конвенции и соглашения. URL: https://www.un.org/ru/documents/decl_conv/conventions/digital_heritage_charter.shtml (дата обращения: 01.04.2023)
- [2] Борисов Н.В., Захаркина В.В., Мбого И.А., Прокудин Д.Е., Щербаков П.П. Проблемное поле сохранения цифрового культурного наследия // Информационное общество: образование, наука, культура и технологии будущего. 2022. № 6. С. 9-26. DOI: 10.17586/2587-8557-2022-6-09-26.
- [3] Непомнящий К.Л., Городищева А.Н. Стандарты цифрового аудиовизуального архивирования культурного наследия // Актуальные проблемы авиации и космонавтики. 2012. Т. 2, № 8. С. 311-312.
- [4] Воробьева Н.В. Оцифровка региональных книжных памятников как возможность представить образ территории в мировом информационном пространстве (на примере Алтайской краевой универсальной научной библиотеки им. В.Я. Шишкова) // Книжные памятники в цифровой среде: Сборник научных трудов, Санкт-Петербург, 06–09 мая 2020 года / Научный редактор Е.Д. Жабко. Санкт-Петербург: Президентская библиотека имени Б.Н. Ельцина, 2020. С. 52-61.
- [5] Костицина А.В., Шпакова М.В. Опыт оцифровки книжных памятников Пермской государственной краевой универсальной библиотеки им. А. М. Горького в рамках федерального проекта «Цифровая культура» // Библиотека и культурное пространство региона: материалы Всероссийской научно-практической конференции (Пермь, 11–12 ноября 2021 г.) / Министерство культуры Российской Федерации; Пермский государственный институт культуры; Центр непрерывного образования и повышения квалификации творческих и управленческих кадров в сфере культуры; ответственный редактор Вафина Е. М.; редакционная коллегия: Чуприн К. П., Шепелева С. В. Пермь, 2021. 324 с. С. 218-223.
- [6] Лихоманов А.В. Реализация проекта «Оцифровка книжных памятников» в Российской национальной библиотеке в 2019 г // Книжные памятники в цифровой среде: Сборник научных трудов, Санкт-Петербург, 06–09 мая 2020 года / Научный редактор Е.Д. Жабко. Санкт-Петербург: Президентская библиотека имени Б.Н. Ельцина, 2020. С. 136-140.
- [7] О проекте // ЭБ Научное наследие России. URL: <http://www.e-heritage.ru/Catalog/About> (дата обращения: 04.04.2023).
- [8] Государственный каталог Музейного фонда Российской Федерации. URL: <https://goskatalog.ru/portal> (дата обращения: 04.04.2023).
- [9] Государственный каталог Музейного фонда Республики Беларусь. URL: <https://www.dkmf.by> (дата обращения: 04.04.2023).
- [10] Guidi G., Gonizzi Barsanti S., Loredana Micoli L., Russo M. Massive 3D Digitization of Museum Contents // Built Heritage: Monitoring Conservation Management Research for Development / Toniolo, L., Boriani M., Guidi G. (eds). 2014. P. 335-346. DOI:10.1007/978-3-319-08533-3_28.
- [11] Gonizzi Barsanti, S., Guidi G. 3D Digitization of Museum Content within the 3d-icons Project // ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 2013. Vol. II-5/W1. P. 151–156. DOI: 10.5194/isprannals-II-5-W1-151-2013.
- [12] Corns A., Deevy A., Devlin G., Kennedy L., Shaw R. 3D-ICONS: Digitizing Cultural Heritage Structures // New Review of Information Networking. 2015. Vol. 20. Iss. 1-2. P. 59-65. DOI: 10.1080/13614576.2015.1115232.
- [13] Papatheodorou C., Dallas C., Ertmann-Christiansen C., Fernie K., Gavrilis D., Masci M.E., Constantopoulos P., Angelis S. A New Architecture and Approach to Asset Representation for European Aggregation: The CARARE Way // Metadata and Semantic Research. MTSR 2011 / Garcia-Barriocanal E., Cebezi Z., Okur M.C., Öztürk A. (eds). Springer, Berlin, Heidelberg. Communications in Computer and Information Science. 2011. Vol 240. P. 412-423. DOI: 10.1007/978-3-642-24731-6_41.
- [14] D'Andrea A., Fernie K. CARARE 2.0: A metadata schema for 3D cultural objects // 2013 Digital Heritage International Congress (DigitalHeritage). Marseille, France. 2013. P. 137-143. DOI: 10.1109/DigitalHeritage.2013.6744745.
- [15] Антопольский А.Б. Границы и содержание инфосферы цифровой гуманитаристики // Информационное общество: образование, наука, культура и технологии будущего. 2022. № 6. С. 62-70. DOI: 10.17586/2587-8557-2022-6-62-70.
- [16] Антопольский А.Б., Володин А.Ю. Справочно-информационная система по цифровой гуманитаристике: опыт описания интернет-ресурсов российских архивов // Историческая информатика. 2022. № 2(40). С. 50-66. DOI: 10.7256/2585-7797.2022.2.38236.
- [17] Di Cosmo R., Zacchiroli S. Software Heritage: Why and How to Preserve Software Source Code // iPRES 2017 - 14th International Conference on Digital Preservation, Sep 2017, Kyoto, Japan. 2017. P. 1-10. URL: <https://hal.science/hal-01590958> (дата обращения: 04.04.2023).
- [18] Di Cosmo R. Software Heritage: Why and How We Collect, Preserve and Share All the Software Source Code // 2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS). Gothenburg, Sweden. 2018. P.

2. URL: <https://ieeexplore.ieee.org/abstract/document/8445152> (дата обращения: 04.04.2023).
- [19] Di Cosmo R. Archiving and Referencing Source Code with Software Heritage // *Mathematical Software – ICMS 2020. ICMS 2020 / Bigatti A., Carette J., Davenport J., Joswig M., de Wolff T. (eds). Springer, Cham. Lecture Notes in Computer Science. 2020. Vol. 12097. P. 362-373. DOI: 10.1007/978-3-030-52200-1_36.*
- [20] Object ID // ICOM. <https://icom.museum/en/resources/standards-guidelines/objectid/> (дата обращения: 04.04.2023).
- [21] Yasaitis K.E. Object ID: A Model of Global Collaboration // *Journal of Museum Management and Curatorship. 2005. Vol. 20. No. 1. P. 21-39. DOI: 10.1080/09647770500402001.*
- [22] Christensen S., Dunlop D. The case for implementing core descriptive embedded metadata at the Smithsonian // *Proceedings of the 2010 International Conference on Dublin Core and Metadata Applications (DCMI '10). Dublin Core Metadata Initiative. 2010. P. 80–87. DOI: 10.5555/1891793.1891804.*
- [23] Baca M. Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information // *Cataloging & Classification Quarterly. 2003. Vol. 36. No. 3-4. P. 47-55. DOI: 10.1300/J104v36n03_05.*
- [24] Saleh E.I. Image embedded metadata in cultural heritage digital collections on the web: An analytical study // *Library Hi Tech. 2018. Vol. 36. No. 2. P. 339-357. DPO: 10.1108/LHT-03-2017-0053.*
- [25] Miller S.J. *Metadata for digital collections. Second edition. Chicago: ALA Neal-Schuman, 2022.*
- [26] Mukherjee S., Das R. Integration of Domain-Specific Metadata Schema for Cultural Heritage Resources to DSpace: A Prototype Design // *Journal of Library Metadata. 2020. Vol. 20. No. 2-3. P. 155-178. DOI: 10.1080/19386389.2020.1834093.*
- [27] METS: Стандарт кодирования и передачи метаданных: [пер. с англ. / Digital Library Federation; рук. рабочей группы по пер. Ю. Г. Селиванова]. Санкт-Петербург: Президентская библиотека, 2018. 309 с.
- [28] ГОСТ Р ИСО 15836-2011 Информация и документация. Набор элементов метаданных Dublin Core [Электронный ресурс] // Национальный стандарт Российской Федерации. Москва: Стандартинформ. 2014. URL: https://standartgost.ru/g/ГОСТ_Р_ИСО_15836-2011 (дата обращения: 04.04.2023).
- [29] Nevile L., Lissonet S. Dublin core and museum information: metadata as cultural heritage data // *International Journal of Metadata, Semantics and Ontologies. 2006. Vol. 1. No. 3. P. 198-206. DOI: 10.1504/IJMSO.2006.012344.*
- [30] Zavalin V., Zavalina O.L. Exploration of Accuracy, Completeness and Consistency in Metadata for Physical Objects in Museum Collections. // *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity. iConference 2023. Lecture Notes in Computer Science. 2023. Vol. 13972. Springer, Cham P. 83-90. DOI: 10.1007/978-3-031-28032-0_7.*
- [31] Mi X., Pollock B.M. Metadata Schema to Facilitate Linked Data for 3D Digital Models of Cultural Heritage Collections: A University of South Florida Libraries Case Study // *Cataloging & Classification Quarterly. 2018. Vol. 56. No. 2-3. P. 273-286. DOI: 10.1080/01639374.2017.1388894.*
- [32] Homburg T., Cramer A., Raddatz L., Mara H. Metadata schema and ontology for capturing and processing of 3D cultural heritage objects. *Herit Sci. 2021. No. 9. 91. DOI: 10.1186/s40494-021-00561-w.*
- [33] Wan J., Zhou Y., Chen G., Yi J. Designing a Multi-level Metadata Standard based on Dublin Core for Museum Data // *International Conference on Dublin Core and Metadata Applications. 2014. P. 31–36. URL: https://dcpapers.dublincore.org/pubs/article/view/3712* (дата обращения: 04.04.2023).
- [34] Квалификаторы Dublin Core (Дублинского ядра) // RUSMARC, российская версия UNIMARC. Российская национальная библиотека. <http://www.rusmarc.info/soft/dcq.html> (дата обращения: 10.04.2023)
- [35] DCMI Qualifiers // Dublin Core Metadata Initiative. <https://www.dublincore.org/specifications/dublin-core/dcmes-qualifiers/> (дата обращения: 10.04.2023).
- [36] González R., Van Der Meer K. Standard Metadata Applied to Software Retrieval // *Journal of Information Science. 2004. Vol. 30(4). P. 300–309. DOI: 10.1177/0165551504045850.*
- [37] Jackson A.S., Han M.-J., Groetsch K., Mustafoff M., Cole T.W. Dublin Core Metadata Harvested Through OAI-PMH // *Journal of Library Metadata. 2008. Vol. 8. No. 1. P. 5-21. DOI: 10.1300/J517v08n01_02.*
- [38] Barroso I., Hartmann N., Ribeiro C. Metadata Crosswalk for a Museum Collection in a Thematic Digital Library, *Journal of Library Metadata. 2015. Vol. 15. No. 1. P. 36-49. DOI: 10.1080/19386389.2015.1011025.*
- [39] Andresen L. Dublin Core as a tool for interoperability: Common presentation of data from archives, libraries and museums // *International Conference on Dublin Core and Metadata Applications. DCMI Proceedings . 2006. URL: https://dcpapers.dublincore.org/pubs/article/view/844* (дата обращения: 04.04.2023).
- [40] CIDOC-CRM (CIDOC Conceptual Reference Model) // Президентская библиотека им. Б.Н. Ельцина. Стандарты метаданных в области культурного наследия. URL: https://www.prlib.ru/cidoc_crm (дата обращения: 10.04.2023).
- [41] Reports about mappings // CIDOC-CRM. URL: https://cidoc-crm.org/report_mappings_res (дата обращения: 10.04.2023).
- [42] Giannoulakis S., Tsapatsoulis N., Grammalidis N. Metadata for Intangible Cultural Heritage - The Case of Folk Dances // *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2018). 2018. P. 634-645. DOI: 10.5220/0006760906340645.*
- [43] Wijesundara C., Sugimoto S. Metadata model for organizing digital archives of tangible and intangible cultural heritage, and linking cultural heritage information in digital space // *LIBRES. 2018. No. 28(2). P. 58-80. DOI: 10.32655/LIBRES.2018.2.2.*
- Борисов Николай Валентинович**, доктор физико-математических наук, профессор, Санкт-Петербургский Государственный Университет, заведующий Кафедрой информационных систем в искусстве и гуманитарных науках факультета искусств, ORCID 0000-0002-1671-5525 (n.borisov@spbu.ru).
- Захаркина Валентина Валентиновна**, кандидат физико-математических наук, Санкт-Петербургский Государственный Университет, доцент, ORCID 0000-0002-4950-2410 (v.zaharkina@spbu.ru).
- Мбого Ирина Анатольевна**, Санкт-Петербургский Государственный Университет, старший преподаватель, ORCID 0000-0002-5942-7143 (i.mbogo@spbu.ru).
- Прокудин Дмитрий Евгеньевич**, докт. филос. наук, доцент, Санкт-Петербургский государственный университет, ORCID 0000-0002-9464-8371 (e-mail: d.prokudin@spbu.ru).
- Щербаков Павел Петрович**, кандидат физико-математических наук, старший научный сотрудник, Санкт-Петербургский Государственный Университет, доцент, ORCID 0000-0003-1158-7460, (p.scherbakov@spbu.ru).

Website as an object of digital heritage: problems of identification and systematization

N. V. Borisov, V. V. Zakharkina, I. A. Mbogo, D. E. Prokudin, P. P. Shcherbakov

Abstract— The development of information and communication technologies has led to the emergence of digital cultural forms, which are cultural heritage as well as traditional objects of cultural heritage. At the same time, modern trends in the development of culture in the information society generate both digitized objects of cultural heritage (secondary digital heritage) and various digital objects that have no analogues in the real world (primary digital heritage). The dynamic development of information and communication technologies has posed as an urgent problem of preserving the digital heritage, which has been solved since the beginning of the XXI century both at the international and national levels. This is due, among other things, to the fact that digital objects exist in various formats and, with the development of information technology, may become unavailable due to the obsolescence of technological solutions, with the termination of support for certain data formats used in their creation and other reasons. This situation poses problems not only of preservation, but, above all, of identifying objects of digital heritage. The issues of classification, description and systematization of identified objects become relevant. The solution of these problems, in the future, will allow organizing and carrying out continuous systematic activities to preserve the digital heritage and ensure effective access to it. In this study, based on the analysis of world experience, approaches to the identification, description and systematization of digital heritage objects are proposed. Special attention is paid to such a digital object as a website (information system), which is due to its complex structure and dependence on many factors. The possible role of the "Center for the Study, Preservation, Restoration and Actualization of Cultural Heritage Objects" created at St. Petersburg State University in theoretical, methodological and practical activities in the field of description and preservation of digital heritage is considered.

Keywords— digital heritage, website, preservation, identification, systematization, description, metadata, Dublin Core, DSpace.

REFERENCES

- [1] Charter on the Preservation of Digital Heritage // UN. Conventions and Agreements. URL: https://www.un.org/ru/documents/decl_conv/conventions/digital_heritage_charter.shtml (accessed date: 01.04.2023).
- [2] Borisov N.V., Zakharkina V.V., Mbogo I.A., Prokudin D.E., Shcherbakov P.P. Problems of digital cultural heritage preservation // Information Society: Education, Science, Culture and Technology of Future. Vol. 6. P. 9-26. DOI: 10.17586/2587-8557-2022-6-09-26.
- [3] Nepomnyashchii K.L., Gorodishcheva A.N. Standarty tsifrovogo audiovizual'nogo arkhivirovaniya kul'turnogo naslediya // Aktual'nye problemy aviatsii i kosmonavтики. 2012. Vol. 2, No. 8. P. 311-312.
- [4] Vorobyova N.V. Regional historical book digitalization as a way of representing the territory in the global information space (Altai Regional Universal Scientific Library named after v. Ya. Shishkov case study) // Book monuments in the digital environment: Collection of research works / Scientific editor E. D. Zhabko. Issue 10. St. Petersburg: The Presidential Library, 2020. P. 52-61.
- [5] Kostitsina A.V., Shpakova M.V. Experience in the digitization of book-monuments of Perm state regional universal library named after A. M. Gorky within the federal project «Digital Culture» // Biblioteka i kul'turnoe prostranstvo regiona: materialy Vserossiyskoy nauchno-prakticheskoy konferentsii (Perm', 11–12 noyabrya 2021 g.) / Ministerstvo kul'tury Rossiyskoy Federatsii; Permskiy gosudarstvennyy institut kul'tury; Tsentr nepreryvnogo obrazovaniya i povysheniya kvalifikatsii tvorcheskikh i upravlencheskikh kadrov v sfere kul'tury; otvetstvennyy redaktor Vafina E. M.; redaktsionnaya kollegiya: Chuprin K. P., Shepeleva S. V. Perm', 2021. P. 218-223.
- [6] Likhomanov A.V. Implementation of the "Digitization of book monuments" project in the Russian National Library in 2019 // Book monuments in the digital environment: Collection of research works / Scientific editor E. D. Zhabko. Issue 10. St. Petersburg: The Presidential Library, 2020. P. 136-140.
- [7] About the Project // Digital Library "Scientific Heritage of Russia". URL: <http://www.e-heritage.ru/Catalog/About> (accessed date: 04.04.2023).
- [8] Gosudarstvennyy katalog Muzeynogo fonda Rossiyskoy Federatsii. URL: <https://goskatalog.ru/portal> (accessed date: 04.04.2023).
- [9] Gosudarstvennyy katalog Muzeynogo fonda Respubliki Belarus'. URL: <https://www.dkmf.by> (accessed date: 04.04.2023).
- [10] Guidi G., Gonizzi Barsanti S., Loredana Micoli L., Russo M. Massive 3D Digitization of Museum Contents // Built Heritage: Monitoring Conservation Management Research for Development / Toniolo, L., Boriani M., Guidi G. (eds). 2014. P. 335-346. DOI:10.1007/978-3-319-08533-3_28.
- [11] Gonizzi Barsanti, S., Guidi G. 3D Digitization of Museum Content within the 3d-icons Project // ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 2013. Vol. II-5/W1. P. 151–156. DOI: 10.5194/isprsannals-II-5-W1-151-2013.
- [12] Corns A., Deevy A., Devlin G., Kennedy L., Shaw R. 3D-ICONS: Digitizing Cultural Heritage Structures // New Review of Information Networking. 2015. Vol. 20. Iss. 1-2. P. 59-65. DOI: 10.1080/13614576.2015.1115232.
- [13] Papatheodorou C., Dallas C., Ertmann-Christiansen C., Fernie K., Gavrilis D., Masci M.E., Constantopoulos P., Angelis S. A New Architecture and Approach to Asset Representation for Europeana Aggregation: The CARARE Way // Metadata and Semantic Research. MTSR 2011 / Garcia-Barriocanal E., Cebeci Z., Okur M.C., Öztürk A. (eds). Springer, Berlin, Heidelberg. Communications in Computer and Information Science. 2011. Vol 240. P. 412-423. DOI: 10.1007/978-3-642-24731-6_41.
- [14] D'Andrea A., Fernie K. CARARE 2.0: A metadata schema for 3D cultural objects // 2013 Digital Heritage International Congress (DigitalHeritage). Marseille, France. 2013. P. 137-143. DOI: 10.1109/DigitalHeritage.2013.6744745.
- [15] Antopolskii A.B. The Boundaries and Content of the Digital Humanities Infosphere // Information Society: Education, Science, Culture and Technology of Future. 2022. Vol. 6. P. 62-70. DOI: 10.17586/2587-8557-2022-6-62-70.
- [16] Antopolskii A.B., Volodin A.U. Information System on Digital Humanities: an experiment of describing the Internet resources of Russian archives // Historical informatics. 2022. No. 2(40). C. 50-66. DOI: 10.7256/2585-7797.2022.2.38236.
- [17] Di Cosmo R., Zacchiroli S. Software Heritage: Why and How to Preserve Software Source Code // iPRES 2017 - 14th International Conference on Digital Preservation, Sep 2017, Kyoto, Japan. 2017. P. 1-10. URL: <https://hal.science/hal-01590958> (accessed date: 04.04.2023).
- [18] Di Cosmo R. Software Heritage: Why and How We Collect, Preserve and Share All the Software Source Code // 2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS). Gothenburg, Sweden. 2018. P. 2. URL: <https://ieeexplore.ieee.org/abstract/document/8445152> (accessed date: 04.04.2023).
- [19] Di Cosmo R. Archiving and Referencing Source Code with Software Heritage // Mathematical Software – ICMS 2020. ICMS 2020 / Bigatti A., Carette J., Davenport J., Joswig M., de Wolff T. (eds).

- Springer, Cham. Lecture Notes in Computer Science. 2020. Vol 12097. P. 362-373. DOI: 10.1007/978-3-030-52200-1_36.
- [20] Object ID // ICOM. <https://icom.museum/en/resources/standards-guidelines/objectid/> (accessed date: 04.04.2023).
- [21] Yasaitis K.E. Object ID: A Model of Global Collaboration // *Journal of Museum Management and Curatorship*. 2005. Vol. 20. No. 1. P. 21-39. DOI: 10.1080/09647770500402001.
- [22] Christensen S., Dunlop D. The case for implementing core descriptive embedded metadata at the Smithsonian // *Proceedings of the 2010 International Conference on Dublin Core and Metadata Applications (DCMI '10)*. Dublin Core Metadata Initiative. 2010. P. 80–87. DOI: 10.5555/1891793.1891804.
- [23] Baca M. Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information // *Cataloging & Classification Quarterly*. 2003. Vol. 36. No. 3-4. P. 47-55. DOI: 10.1300/J104v36n03_05.
- [24] Saleh E.I. Image embedded metadata in cultural heritage digital collections on the web: An analytical study // *Library Hi Tech*. 2018. Vol. 36. No. 2. P. 339-357. DPO: 10.1108/LHT-03-2017-0053.
- [25] Miller S.J. *Metadata for digital collections*. Second edition. Chicago: ALA Neal-Schuman, 2022.
- [26] Mukherjee S., Das R. Integration of Domain-Specific Metadata Schema for Cultural Heritage Resources to DSpace: A Prototype Design // *Journal of Library Metadata*. 2020. Vol. 20. No. 2-3. P. 155-178, DOI: 10.1080/19386389.2020.1834093.
- [27] METS: Metadata Encoding and Transmission Standard: [per. s angl. / Digital Library Federation; ruk. rabochey gruppy po per. Yu. G. Selivanovaj]. Sankt-Peterburg: Prezidentskaya biblioteka, 2018. 309 p.
- [28] GOST R ISO 15836-2011 Informatsiya i dokumentatsiya. Nabor elementov metadannykh Dublin Core // *Natsional'nyy standart Rossiyskoy Federatsii*. Moskva: Standartinform. 2014. URL: https://standartgost.ru/g/ГОСТ_P_ICO_15836-2011 (accessed date: 04.04.2023).
- [29] Nevile L., Lisonnet S. Dublin core and museum information: metadata as cultural heritage data // *International Journal of Metadata, Semantics and Ontologies*. 2006. Vol. 1. No. 3. P. 198-206. DOI: 10.1504/IJMSO.2006.012344.
- [30] Zavalin V., Zavalina O.L. Exploration of Accuracy, Completeness and Consistency in Metadata for Physical Objects in Museum Collections. // *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity*. iConference 2023. Lecture Notes in Computer Science. 2023. Vol. 13972. Springer, Cham P. 83-90. DOI: 10.1007/978-3-031-28032-0_7.
- [31] Mi X., Pollock B.M. Metadata Schema to Facilitate Linked Data for 3D Digital Models of Cultural Heritage Collections: A University of South Florida Libraries Case Study // *Cataloging & Classification Quarterly*. 2018. Vol. 56. No. 2-3. P. 273-286. DOI: 10.1080/01639374.2017.1388894.
- [32] Homburg T., Cramer A., Raddatz L., Mara H. Metadata schema and ontology for capturing and processing of 3D cultural heritage objects. *Herit Sci*. 2021. No. 9. 91. DOI: 10.1186/s40494-021-00561-w.
- [33] Wan J., Zhou Y., Chen G., Yi J. Designing a Multi-level Metadata Standard based on Dublin Core for Museum Data // *International Conference on Dublin Core and Metadata Applications*. 2014. P. 31–36. URL: <https://dcpapers.dublincore.org/pubs/article/view/3712> (accessed date: 04.04.2023).
- [34] Kvalifikatory Dublin Core (Dublinskogo yadra) // RUSMARC, Rossiyskaya versiya UNIMARC. Rossiyskaya natsional'naya biblioteka. URL: <http://www.rusmarc.info/soft/dcq.html> (accessed date: 10.04.2023)
- [35] DCMI Qualifiers // Dublin Core Metadata Initiative. URL: <https://www.dublincore.org/specifications/dublin-core/dcmes-qualifiers/> (accessed date: 10.04.2023).
- [36] González R., Van Der Meer K. Standard Metadata Applied to Software Retrieval // *Journal of Information Science*. 2004. Vol. 30(4). P. 300–309. DOI: 10.1177/0165551504045850.
- [37] Jackson A.S., Han M.-J., Groetsch K., Mustafoff M., Cole T.W. Dublin Core Metadata Harvested Through OAI-PMH // *Journal of Library Metadata*. 2008. Vol. 8. No. 1. P. 5-21. DOI: 10.1300/J517v08n01_02.
- [38] Barroso I., Hartmann N., Ribeiro C. Metadata Crosswalk for a Museum Collection in a Thematic Digital Library, *Journal of Library Metadata*. 2015. Vol. 15. No. 1. P. 36-49. DOI: 10.1080/19386389.2015.1011025.
- [39] Andresen L. Dublin Core as a tool for interoperability: Common presentation of data from archives, libraries and museums // *International Conference on Dublin Core and Metadata Applications*. DCMI Proceedings . 2006. URL: <https://dcpapers.dublincore.org/pubs/article/view/844> (accessed date: 04.04.2023).
- [40] CIDOC-CRM (CIDOC Conceptual Reference Model) // Presidential Library. Standarty metadannykh v oblasti kul'turnogo naslediya. URL: https://www.prlib.ru/cidoc_crm (accessed date: 10.04.2023).
- [41] Reports about mappings // CIDOC-CRM. URL: https://cidoc-crm.org/report_mappings_res (accessed date: 10.04.2023).
- [42] Giannoulakis S., Tsapatsoulis N., Grammalidis N. Metadata for Intangible Cultural Heritage - The Case of Folk Dances // *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2018)*. 2018. P. 634-645. DOI: 10.5220/0006760906340645.
- [43] Wijesundara C., Sugimoto S. Metadata model for organizing digital archives of tangible and intangible cultural heritage, and linking cultural heritage information in digital space // *LIBRES*. 2018. No. 28(2). P. 58-80. DOI: 10.32655/LIBRES.2018.2.2.
- Borisov Nikolay Valentinovich**, doctor of physico-mathematical Sciences, Professor, St. Petersburg State University, head of Department of informational systems in arts and Humanities, faculty of arts, ORCID 0000-0002-1671-5525 (n.borisov@spbu.ru)
- Zakharkina Valentina Valentinovna**, candidate of physico-mathematical Sciences, St. Petersburg State University, associate Professor, ORCID 0000-0002-4950-2410 (v.zaharkina@spbu.ru)
- Mbogo Irina**, Saint Petersburg State University, senior lecturer, ORCID 0000-0002-5942-7143 (i.mbogo@spbu.ru)
- Prokudin Dmitry Evgen'evich**, doctor of Philosophy Sciences, associate Professor, Saint Petersburg State University, ORCID 0000-0002-9464-8371 (e-mail: d.prokudin@spbu.ru).
- Pavel P. Shcherbakov**, candidate of physical and mathematical Sciences, senior researcher, Saint Petersburg State University, associate Professor, ORCID 0000-0003-1158-7460, (p.scherbakov@spbu.ru)