

УДК 378.147:004.9+070.1

ПРОЕКТНОЕ ОБУЧЕНИЕ: СОЗДАНИЕ ИНСТРУМЕНТАРИЯ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ ДЛЯ ИСПОЛЬЗОВАНИЯ В ДАТА-ЖУРНАЛИСТИКЕ

Ерженин Роман Валерьевич

к.э.н., генеральный директор ООО «НПЦ ГОСУЧЕТ»

e-mail: rezhenin@gmail.com

143441, Московская область, Красногорский р-н,
69 км МКАД, п/о Путилково, ООК ЗАО «Гринвуд», стр. 19

Бахвалова Зинаида Андреевна

к.т.н., доцент, доцент института информационных технологий и анализа данных

e-mail: zinand@inbox.ru

Волков Евгений Дмитриевич

студент института информационных технологий и анализа данных

e-mail: vamper.2012@gmail.com

Абзаев Александр Андреевич

студент института информационных технологий и анализа данных

e-mail: aleksandrabz792@gmail.com

Иркутский национальный исследовательский технический университет
664074, г. Иркутск, ул. Лермонтова 83.

Аннотация. В статье рассматривается проблема адаптации образовательной программы подготовки ИТ-специалистов к решению сложных инженерных и, одновременно, творческих задач, связанных с развитием дата-журналистики¹. В качестве источника больших данных используются открытые государственные данные о системе здравоохранения. В статье отражены отдельные положения проектного обучения, сформированного на стандартах проектной и системно-инженерной деятельности, а также на принципах проведения хакатонов, где главным приоритетом является интеллектуальная составляющая работы команды и ее идеи. Приведены некоторые результаты использования проектного обучения при взаимодействии студентов - будущих ИТ-разработчиков с журналистами и исследователями данных. Предложенный подход и полученные на его основе результаты проектного обучения могут использоваться при создании методических рекомендаций для разработки совместных междисциплинарных образовательных программ, объединяющих сферу подготовки будущих ИТ-специалистов и дата-журналистов.

Ключевые слова: дата-журналистика, большие данные, визуализация, отчетность учреждений.

Цитирование: Ерженин Р.В., Бахвалова З.А., Волков Е.Д., Абзаев А.А. Проектное обучение: создание инструментария обработки больших данных для использования в дата-журналистике // Информационные и математические технологии в науке и управлении. 2021. № 4 (24). С. 111-124. DOI: 10.38028/ESI.2021.24.4.011.

Введение. По данным исследователей один среднестатистический житель Земли в 2020 году генерировал 1,7 Мб информации в секунду, и этот же индивид ежедневно потреблял объем информации, соответствующий в среднем 100 500 словам и 34 гигабайтам [1].

Если учитывать, что при всесторонней объемности подобного современного информационного поля человеческий мозг имеет определенные ограничения в восприятии информации, то можно обнаружить противоречие между ростом потребности человека в информации и его умственно-физических ограничениях. Состояние, при котором объем

¹ Дата-журналистика – это “сбор, очистка, организация, анализ, визуализация и публикация данных для создания журналистских материалов”.

потенциально полезной и актуальной информации превышает возможность ее обработки человеком, называют информационной перегрузкой [2]. Перегрузка возникает, когда когнитивные способности человека становятся помехой, а не помощником [3] и особенно остро эта проблема проявляет себя, когда индивид имеет дело с «большими данными», которые для него являются не только слишком большими, но и, одновременно, слишком «сырыми» для традиционных методов обработки.

Вопрос в определении термина «большие данные» (big data) на протяжении последних лет является особо дискуссионным, однако, в данном исследовании мы будем придерживаться определения, сформированного Doug Laney [4] и расширенного нами до формулировки «6V»:

1. VOLUME (объем) – количественная характеристика, указывающая на значительный объем физической информации (данных) о фактах или событиях;
2. VARIETY (разнообразие) - наличие множества самых разнообразных форматов информации (данных);
3. VELOCITY (скорость) – характеристика высокой скорости обновления, обработки и получения информации (данных) и результатов;
4. VALUE (польза) – способность качественно удовлетворять человеческую потребность в информации и получении новых знаний;
5. VERACITY (достоверность) – свойство бесспорного, твердо обоснованного и доказательного знания, содержащегося в информации (данных);
6. VISUALIZATION (визуализация) – технический приём представления числовой информации (данных) в виде, удобном для зрительного наблюдения и анализа.

В современном обществе, испытывающем информационную перегрузку, с каждым днем повышается спрос на различные аналитические материалы и новостные приложения, а также на логики и журналистские расследования, работа над которыми подразумевает обработку большого количества информации (данных). В данном случае следует говорить о развитии медиаконвергенции и появлении совершенно новой формы журналистского материала, основанной на обработке больших данных. Подобная форма получила название «журналистика данных», или «дата-журналистика», она включает в себя различные направления: большие базы данных, компьютерные отчёты и инфографику, визуализацию данных, интерактивную визуализацию, а также структурированную журналистскую систему управления информацией, где зафиксированы основные фактические сведения [5].

Первый пример «компьютерной журналистики» восходит к 1952 г., когда впервые было отмечено использование компьютера в работе журналиста для предсказания результатов выборов в США [6]. За это время дата-журналистика прошла несколько сложных этапов своего развития, что в конечном итоге привело к активному ее распространению по всему миру, и теперь в любых крупных СМИ есть хотя бы один материал, сделанный с помощью данных и визуальных инструментов [5]. К наиболее известным относятся:

- деловые дата-издания - «The New York», «The Washington Post», «Reuters»;
- яркие творческие коллективы, исследующие различные направления, к примеру, «The Pudding» и «Axios»;
- отдельные дата-журналисты, как, например, Mona Chalabi – британская дата-журналистка, которая публикует необычные графики на своей странице в Instagram;
- российские дата-издания, к примеру, Тинькофф-журнал и ряд других.

Вопросам развития отечественной дата-журналистики посвящены несколько научных исследований [7-9] и многочисленное количество интернет-публикаций, в частности [10-12], однако в России слабо раскрытым остается вопрос взаимодействия дата-журналистов с

разработчиками программного обеспечения, предназначенного для обработки больших данных, а также вопрос специальной подготовки ИТ-специалистов для работы в сфере журналистики данных.

Цель данной статьи – обозначить основные положения проектного обучения студентов ИТ-специальностей технического вуза и показать на примерах, как формируется предметная область разработки прикладного программного обеспечения, предназначенного для накопления, обработки и визуализации открытых государственных данных². Областью больших данных выбрана система здравоохранения т.к. в период пандемии именно эти данные особо востребованы в медийной среде для удовлетворения нарастающих запросов общества.

Проектное обучение как хакатон. Главная проблема для журналиста, знающего о возможностях дата-журналистики, заключается в сложности обработки и представления этих данных в привлекательном для читателя виде. В то же самое время в среде разработчиков программного обеспечения знают, как организовать и обеспечить информационные процессы обработки больших данных, но не знают, как представить результаты этой работы в значимом для читателя виде.

Одним из способов организации взаимовыгодного сотрудничества журналистов, редакторов, ИТ-разработчиков и исследователей данных стал «хакатон»³ [13]. В течение последнего десятилетия хакатоны стали особенно популярны в таких странах, как США, Канада, Нидерланды, Дания, Аргентина, Англия, Финляндия и т.п. [5]. Участие в хакатонах принимают студенты различных университетов, заинтересованные, в том числе, и в работе с открытыми государственными данными. К таким хакатонам относятся Canadian Open Data Experience (CODE), Hackathon Open Data Brussels, International Open Data Hackathon, Open Data Mashup и ряд других.

Российский опыт проведения хакатонов, где внедрялась и проверялась на практике модель взаимодействия общества и государства, включает такие проекты, как Федеральный хакатон «Data.gov.ru Hackathon», хакатон Audithon от Счетной Палаты РФ, хакатоны по открытым данным в Москве, Санкт-Петербурге, Воронеже, Уфе, Улан-Удэ и др. Поддержка в проведении хакатонов и конкурсов по разработке программных решений на основе открытых данных прозвучала и от Аналитического центра при Правительстве Российской Федерации.

К основным достоинствам хакатона, как инструмента для создания инновационных решений, относится получение от участников новых идей за достаточно короткое время. Однако чаще всего такой формат – это соревнование для команд подготовленных специалистов, желающих освоить «чужой» бюджет, а реальность такова, что 90% – это неопытные, наименее образованные участники, которые приходят в первую очередь из любопытства с желанием создать прототип, собрать команду или завести полезные знакомства [14].

Таким образом, при всех несомненных достоинствах хакатона к основным ограничениям его использования для создания более сложного продукта следует отнести следующее:

² Открытые государственные данные – открытые данные, опубликованные государственными органами, их территориальными органами, органами местного самоуправления или организациями, подведомственными государственным органам, органам местного самоуправления [15].

³ Англ. *hackathon*, от *hack* – хакер и *marathon* – марафон) – формат общения со специалистами в области обработки данных в режиме «мозгового штурма», в результате которого может появиться новый проект или рабочий прототип программного обеспечения для решения важной задачи.

1. Сложность предметной области. Для создания любого нового продукта даже в упрощенном виде необходим детальный анализ предметной области, и чем серьезней проблема, чем шире охват ее данными, тем сложнее анализ.
2. Фактор времени. Отведенного времени в рамках хакатона чаще всего недостаточно для решения сложных проектов, предполагающих глубокое погружение в предметную область.
3. Технооптимизм. Разработчики ИТ-систем очень редко ставят себе задачу найти наиболее подходящее решение. Чаще всего у них есть только желание попробовать определенную технологию [14], возможно, только ту, которая им известна.

Таким образом, все три указанных ограничения связаны с проектным треугольником: трио времени, ресурсов и объема работ. Очевидно, что если целью является создание реального медиапродукта, а не прототипа, то его будет трудно вписать в условия проведения хакатона без потери качества, особенно когда речь идет о привлечении студентов, большинство из которых не имеют ни должной подготовки, ни достаточного опыта.

Хакатон, в первую очередь, полезен студентам с образовательной точки зрения, т.к. позволяет развивать компетенции технологического предпринимательства и проектного управления, что особенно важно для современной подготовки специалистов инженерных специальностей. В этом направлении хакатон пересекается с целями *проектного обучения*, внедряемого в последнее время в учебный процесс технических вузов в целях переориентации на подготовку специалистов, для которых приоритетом становится именно интеллектуальная составляющая работы команды и ее идеи и, в том числе, и ИТ-специалисты здесь являются в качестве наиболее приоритетных. Если для участников хакатонов задачей максимум является быстрый поиск инвестора, то проектное обучение должно быть направлено не только на развитие способностей к технологическому предпринимательству и работе в команде, но и также на системное и последовательное накопление инженерных знаний о принципах и основах разработки и реализации ИТ-проектов.

Следует обратить внимание на то, что курс «Основы проектной деятельности» введен в обучение в университетах еще в 2020 году, при этом профессорско-преподавательский состав достаточно часто сталкивается с определенными сложностями с ведением этого курса, что обусловлено, по нашему мнению, в первую очередь слабой методической подготовкой комбинированного обучения. С целью создания методического обеспечения в следующем разделе будет предложена структура организации проектного обучения будущих ИТ-разработчиков для получения навыков использования открытых государственных данных при взаимодействии с дата-журналистами, базирующаяся на опыте работы команды студентов в этом направлении.

1. Организация проектного обучения. Создание медиапродукта на основе больших открытых государственных данных – это творческая исследовательская задача, решение которой требует от обучающихся использования интегрированных знаний в различных областях, а от преподавателя необходимых компетенций, позволяющих использовать особые методические (проектные) подходы. Поскольку проектный метод обучения предусматривает разработку ИТ-проекта, от преподавателя требуется не только тщательное планирование, но и качественное управление исполнением проекта, в том числе организацию обсуждения со всеми участниками проекта: «внутренними» и «внешними». Очевидно, что эффективность образовательного проекта во многом зависит от качества методики проектного обучения, способной учитывать теоретическую, практическую и познавательную значимость полученных результатов.

К особенностям предложенного далее подхода к организации проектного обучения можно отнести то, что разработка проекта основана на микросервисной архитектуре, т.е. за развитие, разработку и сопровождение каждого модуля отвечает отдельная команда. Кроме этого, команда ориентирована на самостоятельную деятельность обучающихся, организованную в смешанном формате выполнения самостоятельной или групповой работы, ограниченную конкретным временным отрезком.

Основные положения организации проектного обучения сгруппированы по следующим направлениям:

Стандарты, используемые для формализации методики:

- ГОСТ Р 54869-2011 «Национальный стандарт Российской Федерации. Проектный менеджмент Требования к управлению проектом».
- ГОСТ 34.601-90 «Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Стадии создания».
- ГОСТ Р ИСО/МЭК 12207-2010 «Национальный стандарт Российской Федерации. Информационная технология. Системная и программная инженерия. Процессы жизненного цикла программных средств».
- ГОСТ Р ИСО/МЭК 15288-2005 «Национальный стандарт Российской Федерации. Информационная технология. Системная инженерия. Процессы жизненного цикла систем».

Ролевая (организационная структура) образовательного проекта приведена в таблице 1:

Таблица 1. Роли в образовательном проекте.

	<i>Роль</i>	<i>Юрлицо</i>	<i>Физлицо</i>
1	Конечный заказчик проекта	Средство массовой информации (СМИ)	Журналист, редактор
2	Инициатор проекта	Университет, Консалтинговая компания	Руководитель университета или компании
3	Руководитель проекта		Руководитель образовательной программы (проекта)
4	Куратор проекта	Консалтинговая компания	Руководитель компании
5	Команда проекта		Студенты, преподаватели, аналитики, ИТ-специалисты

Схема взаимоотношений в образовательном проекте.

В команде студентов есть три роли: наставник (преподаватель), руководитель проекта (студент) и участники команды разработчиков. Поскольку суть проектного обучения заключается в эмпирическом подходе, самоорганизации и постоянном совершенствовании, эти три роли дают минимальное определение обязанностей и подотчетности для эффективного выполнения работы в команде [16]. Структура взаимоотношений ролей команды проекта показана на рис.1.

Основная роль руководителя образовательного проекта заключается в организации эффективного процесса обучения. Он учит, тренирует и наставляет команду в тех вопросах, которые команда не может решить самостоятельно.

Руководитель команды проекта (студент) поддерживает сотрудничество между участниками команды и создает положительную рабочую среду. В его обязанности входит организация встреч команды для обсуждения рабочих вопросов.

Каждый член команды просто член команды. В группе нет четких должностей, здесь не выделяются бизнес-аналитик, архитектор, лидер команды, UX/UI-дизайнер, программист.

Команда действует совместно: она самостоятельно определяет, как выполнить поставленную задачу.

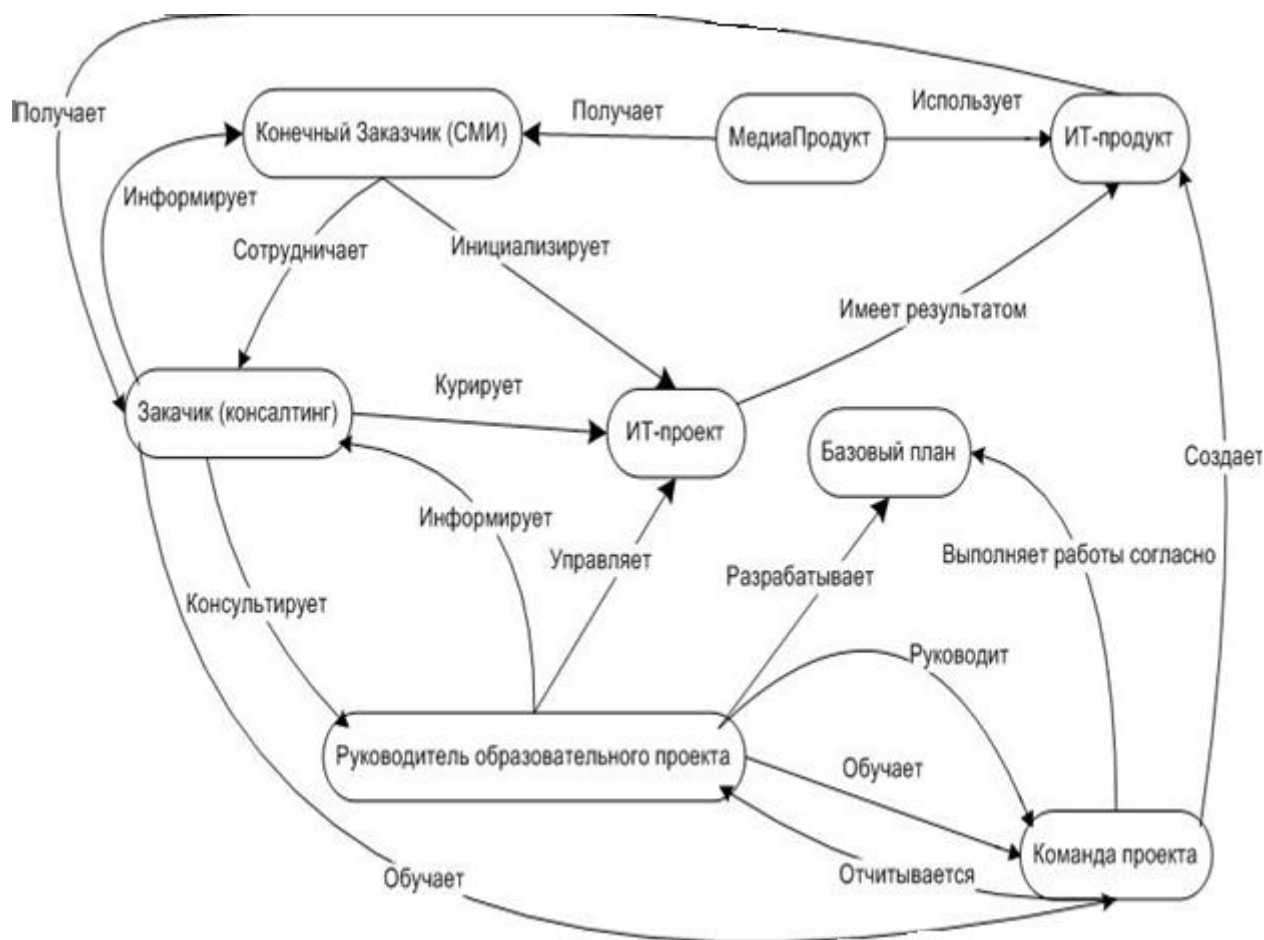


Рис. 1. Схема взаимоотношений в образовательном проекте.

Большие данные – открытые государственные данные о системе здравоохранения приведены в таблице 2.

Таблица 2. Большие открытые данные о системе здравоохранения.

	Название	Адрес	Данные
1	Официальный сайт для размещения информации об учреждениях	bus.gov.ru	Данные о плановых показателях и результатах финансово-хозяйственной деятельности учреждений, отраженные в формах бюджетной отчетности в соответствии с требованиями Приказа Минфина России №86Н от 21.07.2011 [17] и другими локальными нормативными актами органов исполнительной власти.
2	Единая государственная информационная система в сфере здравоохранения	egisz.rosminzdrav.ru	Геоданные о местонахождении учреждений; данные о доступности скорой и первичной медицинской помощи, укомплектованности кадрами, состоянием зданий.

Основные характеристики плана проектного обучения:

- Проектное обучение №1 – анализ данных о финансово-хозяйственной деятельности учреждений по отчетам, размещаемым в сети Интернет в соответствии с Приказом Минфина России №86Н от 21.07.2011, условное название - «Проект 86Н».
- Проектное обучение №2 – анализ данных об исполнении Указа Президента РФ №597, условное название - «Проект Указ 597».

- Проектное обучение №3 – анализ данных о состоянии фельдшерско-акушерских пунктов (ФАП) в рамках Указа Президента № 597, условное название - «Проект ФАП».

Итерационная модель жизненного цикла ИТ-проекта.

В учебных целях фазы жизненного цикла могут быть непоследовательными и начало работ над следующим этапом может наступить до завершения предыдущего. Это обеспечивает гибкость обучения, т.к. позволяет проходить этапы в несколько итераций, каждый раз создавая новый прототип и проверяя актуальность требований, по которым он создавался, внося технические доработки в интерфейс и функциональность. Подобная модель управления ИТ-проектом предоставляет возможность оперативно создать прототипы программного продукта для проверки его работоспособности с пользователями и, соответственно, быстрого получения комментариев и замечаний к системе [18].

Этапы жизненного цикла программного продукта:

Этап 1. Выработка системных требований (постановка задачи).

- Описание предметной области.
- Формулировка цели работы.
- Составление словарей.
- Описание функционального назначения системы.

Этап 2. Формальное описание требований к программному продукту.

- Диаграмма вариантов использования.
- Функциональные требования.
- Нефункциональные требования.
- Специальные требования.
- Инфологическая модель предметной области (нотация Чена).

Этап 3. Разработка технического задания.

Этап 4. Проектирование программного продукта.

- Выбор инструментария для разработки.
- Формирование архитектуры приложения.
- Проектирование хранилища данных.
- Подготовка макетов интерфейсных объектов.
- Разработка дизайна визуализации данных.

Этап 5. Программная реализация.

Этап 6. Разработка документации.

- Описание сценариев действий пользователя.
- Описание сценариев действий оператора.
- Описание сценариев действий администратора.

2. Результаты проектного обучения. Наиболее существенное требование к коммуникациям при разработке программного решения выражается в тезисе «Хороший ИТ-подрядчик всегда будет говорить с заказчиком на одном с ним языке». Очевидно, что язык дата-журналистики – это инфографика и визуализация, а для разработчика ИТ-систем еще важны технологии, информационные процессы и логика, которые для журналиста являются «черной магией». Чтобы разрешить эти противоречия, следует использовать профессиональную ИТ-лексику так, что она была одновременно и максимально понятна потенциальному «заказчику», и в то же самое время могла сформировать представление об информационных объектах, их атрибутах и отношениях между ними, а также о характере информационных потребностей пользователей.

Учитывая ограниченные возможности данной статьи, далее будут представлены отдельные результаты проектного обучения, отражающие указанные принципы, т.е. сосредоточенные на понимании ИТ-разработчиком потребностей дата-журналиста.

Так, например, инфологическая модель предметной области, разработанная студентами 4-го курса специальности «Информационные системы и технологии» Иркутского национального исследовательского технического университета в рамках проекта «86Н» с использованием специальных графических нотаций, отображена на рисунке 2. Инфологическая модель предметной области проекта ФАП отображена на рисунке 3.

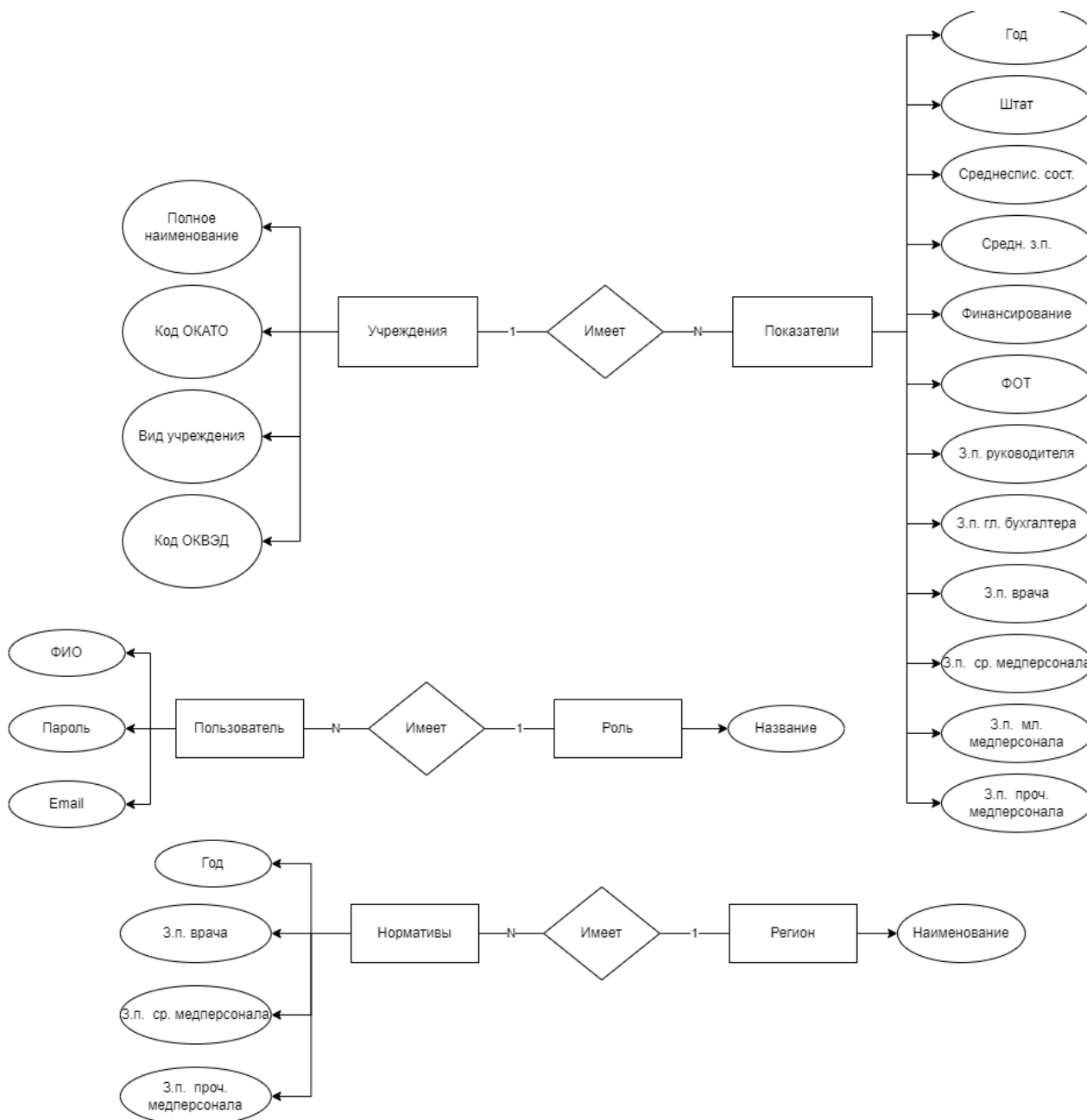


Рис. 2. Инфологическая модель предметной области для проекта «86Н».

Инфографика – неотъемлемая и наиболее креативная составляющая дата-журналистики. Визуализация данных – это отчасти искусство, отчасти наука и основная задача здесь состоит в том, чтобы сделать искусство правильным, не искажая науку, и

наоборот [19]. На рисунке 4 представлен макет экрана «Карта медицинских пунктов», разработанный студентами в рамках описания технического задания к проекту «ФАП».

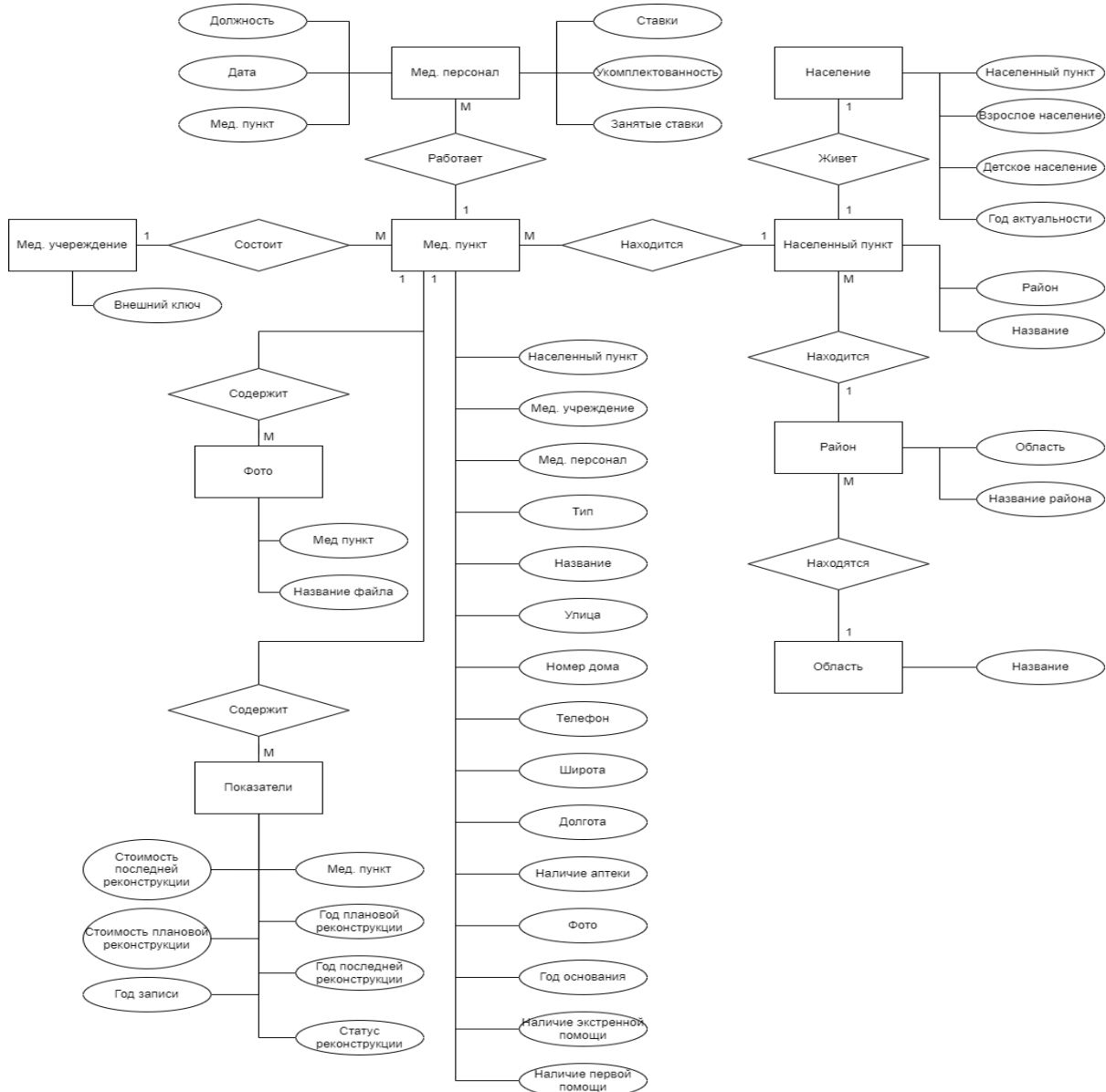


Рис. 3. Инфологическая модель предметной области проекта «ФАП».

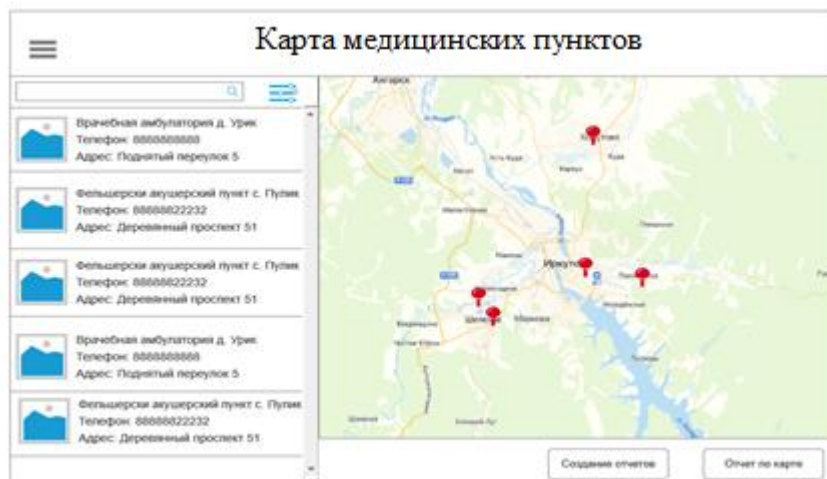


Рис. 4. Макет экрана «Карта медицинских пунктов».

К одной из наиболее информативных форм инфографики, позволяющих создавать простую систему из сложных и разрозненных отчетов и данных, относится **дашборд**⁴. Подобная форма представления данных и информации позволяет упростить процесса визуализации отчетов, объединить различные показатели в единую структурированную систему, создать иерархию данных, выявить критические точки, предотвратить риски, выявить проблемы, выбрать управленческое решение.

На рисунке 5 представлен макет экрана экрана-дашборда, разработанный в рамках создания функциональных требований к визуализации информации по проекту «Указ 597».

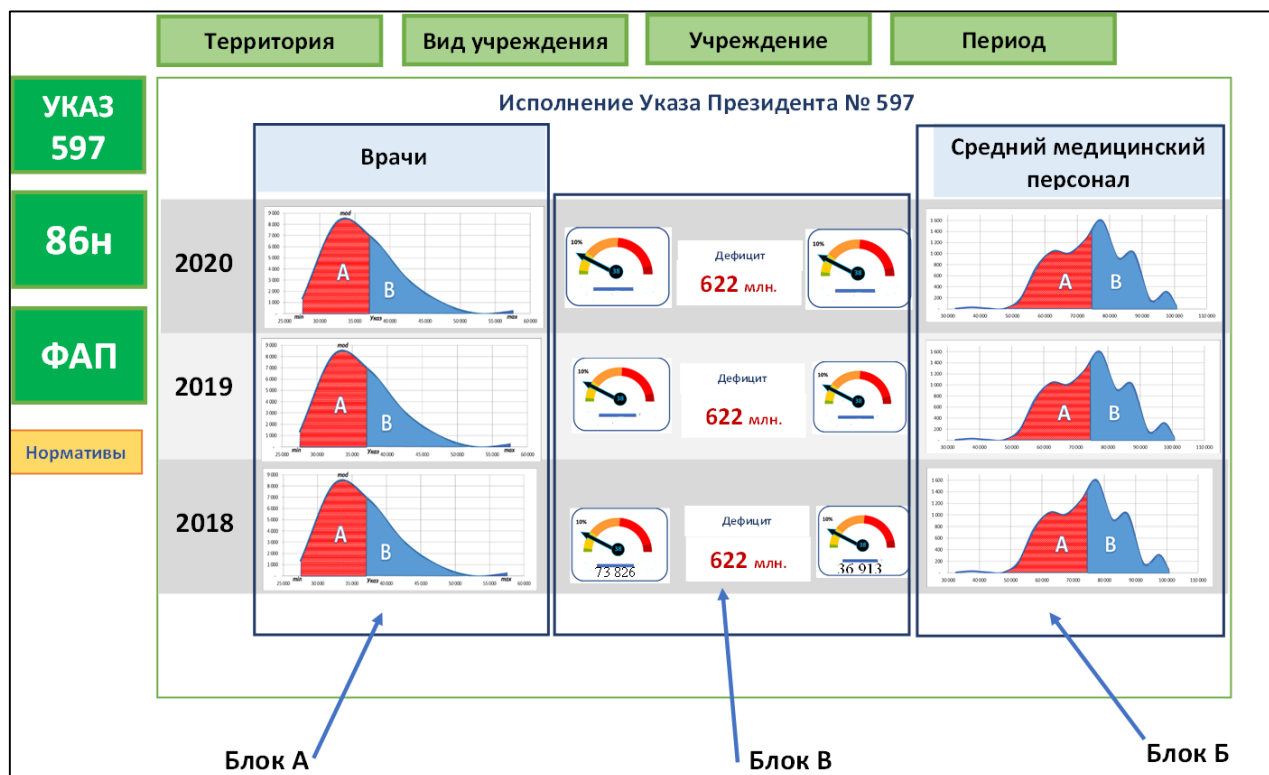


Рис. 5. Макет экрана-дашборда «Исполнение Указа Президента № 597».

Заключение. Развитие дата-журналистики, как нового медиатренда в информационной культуре российского общества, можно рассматривать одновременно и как появление новых алгоритмов деятельности современного дата-журналиста, главная задача которого состоит в «совмещении роли переводчика с экономического/государственного/цифрового языка в язык, понятный всем, и, одновременно, роли аналитика, представляющего материалы в предельно наглядной форме» [20]. В то же самое время наиболее сложным остается пока вопрос инструментального обеспечения дата-журналиста и, в первую очередь, это связано со стоимостью разработки, которая в некоторых случаях может сравниться с рекламным бюджетом СМИ. Очевидно, хакатоны по-прежнему будут играть важную роль не только в развитии дата-журналистики, привлекая дизайнеров, программистов и открывая новые таланты, но и в улучшении качества образования студентов. Однако, системная подготовка специалистов на базе высших учебных заведений, переориентированная и адаптированная для решения междисциплинарных, сложных и, одновременно, творческих задач имеет намного больше перспектив для развития информационной, гражданской и активистской культуры общества.

⁴ В переводе с английского dashboard — это интерактивная панель с важной информацией, сгруппированной на одном экране, где отображены показатели из систем аналитики, таблиц и т.п.

СПИСОК ЛИТЕРАТУРЫ

1. Infographic. Data Never Sleeps 6.0. Режим доступа: <https://www.domo.com/learn/data-never-sleeps-6> (дата обращения: 12.12.2021).
2. Общество и пандемия: опыт и уроки борьбы с COVID-19 в России. Москва. 2020. 744 с.
3. Bohn R., Short J.E. Measuring Consumer Information. International Journal of Communication. 6. 2012. Pp. 980-1000.
4. Laney D. 3D Data Management: Controlling Data Volume., Velocity, and Variety. 06.02.2001. Режим доступа: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocityand-Variety.pdf> (дата обращения: 12.12.2021).
5. Неренц Д.В. «Журналистика данных» в США. США & Канада: экономика, политика, культура. 2019. 49(4). С. 75-91. DOI: 10.31857/S032120680004360-0.
6. Houston B. Computer-Assisted Reporting: A Practical Guide, Fourth Edition. 2015. New York City. Routledge.
7. Лисицин М.Е. Понятие журналистика данных в освещении российских и зарубежных СМИ: дис. докт. техн. наук. М. 2019. 202с.
8. Симакова С.И. Дата-журналистика как медиатренд // Вестник ННГУ. 2014. №2-2. Режим доступа: <https://cyberleninka.ru/article/n/data-zhurnalistika-kak-mediatrend> (дата обращения: 02.12.2021).
9. Шилина М.Г. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакommunikации: к вопросу формирования теоретических исследовательских подходов // Медиаскоп. 2013. № 1.
10. Что такое журналистика данных. Режим доступа: <https://texterra.ru/blog/chto-takoe-zhurnalistika-dannykh.html> (дата обращения: 12.12.2021).
11. Сервис работы с документами и таблицами от Google. Режим доступа: <https://docs.google.com/> (дата обращения: 12.12.2021).
12. Каталоги открытых данных DataGov.ru (Россия). Режим доступа: <http://www.datagov.ru> (дата обращения: 10.12.2021).
13. Хакатон по журналистике данных: как находить в Интернете сенсации. Режим доступа: <https://news.itmo.ru/ru/startups-and-business/initiative/news/5599/> (дата обращения: 02.12.2021).
14. Бут. А. Почему хакатоны (почти) никогда не приносят компаниям новые идеи? Режим доступа: <https://rb.ru/opinion/hakatony-ne-prinosyat-idei/> (дата обращения: 12.12.2021).
15. Протокол заседания Правительственной комиссии по координации деятельности Открытого правительства от 25 декабря 2014 г. № 10.
16. Гибкая методология Agile. Режим доступа: <https://www.atlassian.com/ru/agile/scrum/roles> (дата обращения: 12.12.2021).
17. Приказ Минфина России от 21 июля 2011 г. № 86н «Об утверждении порядка предоставления информации государственным (муниципальным) учреждением, её размещения на официальном сайте в сети "Интернет" и ведения указанного сайта». Режим доступа: <https://roskazna.gov.ru/dokumenty/gis/elektronnyy-byudzhet/ofitsialnyy-sayt-gmu/535198/> (дата обращения: 12.12.2021).
18. Управление ИТ-проектами: введение, жизненный цикл и окружение проекта: Учеб.-метод. пособие к курсу «Проектирование информационных систем управления» для бакалавров направления 09.03.02 «Информационные системы и технологии» / Т.Ф. Шамсутдинов. Казань: Изд-во Казанск. гос. архитектур.-строит. ун-та. 2017. 103 с.

19. Claus O. Wilke Fundamentals of Data Visualization. 2019. 389 p.

20. Бегтин И. Памятка по журналистике данных // Редакторский портал. Режим доступа: <http://www.redactor.in.ua/internet/3126.html> (дата обращения: 12.12.2021).

UDC 378.147:004.9+070.1

PROJECT LEARNING: CREATING A BIG DATA PROCESSING TOOLKIT FOR USE IN DATE JOURNALISM

Roman Yerzhenin

Candidate of Economic Sciences,

General Director of OOO NPTs GOSUCHET,

e-mail: rerzhenin@gmail.com

143441 Moscow region, Krasnogorsk district, 69 km MKAD, p/o Putilkovo,

OOK CJSC "Greenwood", p. 19

Zinaida Bakhvalova

Ph.D., Associate Professor, Associate Professor

of the Institute of Information Technology and Data Analysis

e-mail: zinand@inbox.ru

Evgeniy Volkov

student of the Institute of Information Technology and Data Analysis

e-mail: vamper.2012@gmail.com

Alexander Abzaev

student of the Institute of Information Technology and Data Analysis

e-mail: aleksandrabz792@gmail.com

Irkutsk National Research Technical University

664074, Irkutsk, Lermontov st., 83.

Abstract. The article discusses the problem of adapting the educational program of training IT specialists to solving complex engineering and, at the same time, creative problems associated with the development of data journalism. Open government data on the health care system is used as a source of big data. The article proposes the main provisions of the approach to project training, formed on the standards of project and system engineering activities, as well as on the principles of conducting hackathons, where the main priority is the intellectual component of the team's work and its ideas. Some results of the use of project-based learning in the interaction of students - future IT developers with journalists and data researchers are presented. The proposed approach and the results of project training obtained on its basis can be used to create guidelines for the development of joint interdisciplinary educational programs that combine the field of training future IT specialists and data journalists.

Keywords: data- journalism, big data, visualization, reporting of institutions.

REFERENCES

1. Infographic. Data Never Sleeps 6.0. Available at: <https://www.domo.com/learn/data-never-sleeps-6> (accessed 12.12. 2021).
2. Obshchestvo i pandemiya: opyt i uroki bor'by s COVID-19 v Rossii [Society and the pandemic: experience and lessons of the fight against COVID-19 in Russia]. Moskva = Moscow. 2020. 744 p. (in Russian).

3. Bohn R., Short J.E. Measuring Consumer Information. *International Journal of Communication*. 6. 2012. Pp. 980-1000.
4. Laney D. 3D Data Management: Controlling Data Volume., Velocity, and Variety. 06.02.2001. Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocityand-Variety.pdf> (accessed 12.12.2021).
5. Nerents D.V. «Zhurnalistika dannykh» v SSHA. SSHA & Kanada: ekonomika, politika, kul'tura ["Data Journalism" in the United States. USA & Canada: economics, politics, culture]. 2019. 49(4). Pp. 75-91. DOI: 10.31857/S032120680004360-0. (in Russian).
6. Houston B. *Computer-Assisted Reporting: A Practical Guide*, Fourth Edition. 2015. New York City. Routledge.
7. Lisitsin M.Ye. Ponyatiye zhurnalistika dannykh v osveshchenii rossiyskikh i zarubezhnykh SMI [The concept of data journalism in the coverage of Russian and foreign media]: dis. doct. tech. sciences. 2019. 202 p. (in Russian).
8. Simakova S.I. Data-zhurnalistika kak mediatrend [Data journalism as a media trend]// *Vestnik NNGU = Bulletin of NNGU*. 2014. No. 2-2. Available at: <https://cyberleninka.ru/article/n/data-zhurnalistika-kak-mediatrend> (accessed 02.12.2021). (in Russian).
9. Shilina M.G. Data Journalism – data-zhurnalistika, zhurnalistika metadannykh – v strukture mediakommunikatsii: k voprosu formirovaniya teoreticheskikh issledova-tel'skikh podkhodov [Data Journalism - data journalism, metadata journalism - in the structure of media communication: on the formation of theoretical research approaches] // *Mediascope = MediaScope*. 2013. No. 1. (in Russian).
10. Chto takoye zhurnalistika dannykh [What is data journalism]. Available at: <https://texterra.ru/blog/chto-takoe-zhurnalistika-dannykh.html> (accessed 02.12.2021) (in Russian).
11. Servis raboty s dokumentami i tablitsami ot Google [Service for working with documents and spreadsheets from Google]. Available at: <https://docs.google.com/> (accessed 10.12.2021). (in Russian).
12. Katalogi otkrytykh dannykh DataGov.ru (Rossiya) [Open data catalogs DataGov.ru (Russia)]. Available at: <http://www.datagov.ru> (accessed 10.12.2021). (in Russian).
13. Khakaton po zhurnalistike dannykh: kak nakhodit' v Internete sensatsii [Data Journalism Hackathon: How to Find Internet Sensations]. Available at: https://news.itmo.ru/ru/startups_and_business/initiative/news/5599/ (accessed 02.12.2021). (in Russian).
14. But A. Pochemu khakatony (pochti) nikogda ne prinisyat kompaniyam novyye idei? [Why do hackathons (almost) never bring new ideas to companies?]. Available at: <https://rb.ru/opinion/hakatony-ne-prinisyat-idei> (accessed 12.12.2021). (in Russian).
15. Protokol zasedaniya Pravitel'stvennoy komissii po koordinatsii deyatel'nosti Otkrytogo pravitel'stva ot 25 dekabrya 2014 g [Minutes of the meeting of the Government Commission for the Coordination of the Activities of the Open Government dated December 25, 2014]. No. 10. (in Russian).
16. Gibkaya metodologiya Agile [Agile methodology]. Available at: <https://www.atlassian.com/ru/agile/scrum/roles> (accessed 12.12.2021). (in Russian).
17. Prikaz Minfina Rossii ot 21 iyulya 2011 g. No. 86n «Ob utverzhdenii poryadka predostavleniya informatsii gosudarstvennym (munitsipal'nym) uchrezhdeniyem, yeyo razmeshcheniya na ofitsial'nom sayte v seti "Internet" i vedeniya ukazannogo sayta» [Order of the Ministry of Finance of Russia dated July 21, 2011. No. 86h "On approval of the procedure

- for the provision of information by a state (municipal) institution, its posting on the official website on the Internet and maintaining the specified site"]. Available at: <https://roskazna.gov.ru/dokumenty/gis/elektronnyy-byudzhnet/ofitsialnyy-sayt-gmu/535198/> (accessed 12.12.2021). (in Russian).
18. Shamsutdinov T.F. Upravleniye IT-proyektami: vvedeniye, zhiznennyy tsikl i okruzhe-niye proyekta: Ucheb.-metod. posobiye k kursu «Proyektirovaniye informatsionnykh si-stem upravleniya» dlya bakalavrov napravleniya 09.03.02 «Informatsionnyye sistemy i tekhnologii» [IT project management: introduction, life cycle and project environment: Textbook. Method. manual for the course "Designing management information systems" for bachelors of the direction 09.03.02 "Information systems and technologies"]. Kazan': Izdatel'stvo Kazanskogo gosudarstvennogo arkhitekturno-stroitel'nogo universiteta = Publishing house of the Kazan State University of Architecture and Civil Engineering. 2017. 103 p. (in Russian).
 19. Claus O. Wilke Fundamentals of Data Visualization. 2019. 389 p.
 20. Begtin I. Pamyatka po zhurnalistike dannykh [Memo on Data Journalism] // Redaktorskiy portal = Editorial Portal. Available at: <http://www.redactor.in.ua/internet/3126.html> (accessed 12.12.2021). (in Russian).

Статья поступила в редакцию 13.12.2021; одобрена после рецензирования 20.12.2021; принята к публикации 24.12.2021.

The article was submitted 13.12.2021; approved after reviewing 20.12.2021; accepted for publication 24.12.2021.