

УДК 004.9

Использование информационных технологий и инфраструктур для агрегации научной информации. Опыт Канады, Нидерландов, Германии¹

Биктимиров М. Р.*, Глебский В. Л.***, Долгов Б. В.***, Поликарпов С. А.*

**Вычислительный центр им. А. А. Дородницына,
119333, Россия, г. Москва, ул. Вавилова, 40*

***Ассоциация «е-Арена», 119333, Россия, г. Москва, ул. Вавилова, 40*

****ЗАО «Академинформ», 119991, Россия, г. Москва, Ленинский пр-т, 32А*

e-mail: marat@ras.ru, polik@vmail.ras.ru

получена 20 октября 2014

Ключевые слова: агрегация, научные данные, инфраструктура

В работе анализируется опыт создания и использования информационных технологий и инфраструктур для агрегации научной информации в трех странах-лидерах мирового ИТ-рейтинга «Индекс сетевой готовности», а также опыт Европейского Союза в целом. Изучались как руководящие и аналитические документы, так и конкретные действующие проекты в рассматриваемой предметной области.

Проведенное исследование позволяет выделить как характерные общие черты в подходах, в понимании актуальности и необходимости решения задачи создания национальных инфраструктур хранения и обработки данных научных исследований и их объединения, так и заметные организационные различия, основанные на принятых в каждой конкретной стране механизмах (традициях) и органах управления, а также оценить уровень продвижения в решении поставленных задач.

Были также указаны общие основные элементы построения всех инфраструктур хранения и обработки данных научных исследований.

1. Выбор стран

Анализ международного опыта в области создания и использования информационных технологий и инфраструктур для агрегации научной информации проведен на основе изучения материалов трех стран, стабильно входящих в первые двадцать позиций мирового ИТ-рейтинга «Индекс сетевой готовности» (Networked Readness Index) [1]: Канады, Нидерландов, Германии.

¹Работа проводилась при финансовой поддержке РФФИ, грант № 14-07-00783 А.

Выбор в пользу этих стран сделан как на основе анализа доступности необходимой информации в открытых источниках, основным из которых являлись различные сетевые ресурсы Интернета, так и с учетом того, что спектр рассматриваемых стран не должен быть однородным и должен содержать подходы государств, близких к России по территориальным характеристикам и характеристикам государственного устройства.

Учитывая то, что большинство стран первой десятки упомянутого выше рейтинга входят в Европейский Союз (ЕС) и должны при создании и развитии своих систем руководствоваться нормативными актами ЕС, подход Европейского Союза, выражающийся в единых рекомендациях странам-членам ЕС по созданию систем формирования, хранения и использования данных научных исследований, также был принят во внимание.

2. Актуальность вопроса о научных данных в рассматриваемых странах, наличие у них опыта построения инфраструктур для агрегации и обмена научной информацией, используемые подходы

Анализ собранной информации позволяет утверждать, что тема агрегации и использования научных данных в разрезе неимоверного роста информации, накапливаемой во всех отраслях знаний с использованием новых возможностей, появившихся с развитием информационно-коммуникационных технологий, рассматривается всеми указанными странами как актуальная с 1990-х годов, а на рубеже нового столетия она вышла на уровень стратегического планирования. За прошедший период в большинстве стран появились государственные программы, стратегии и рекомендации, нацеленные на попытку создания отдельных элементов и систем, которые бы позволили исследователям оперативно и эффективно извлекать нужную им информацию из источников ее накопления (различного рода библиотек, архивов и т.д.). Абсолютно всем участникам процесса стало понятно, что накопленные в хранилищах научные данные ложатся туда мертвым грузом, если сообщество исследователей не располагает инструментарием для их поиска, получения и обработки. В то же время очевидно, что эффективная работа возможна только с данными, облеченными в электронную форму и хранящимися на электронных носителях. Поэтому, с одной стороны, возникла проблема оцифровки различных видов данных и материалов исследований прошлых лет, что активизировало процесс создания электронных хранилищ (во всех рассматриваемых странах такие хранилища созданы и активно развиваются), а с другой стороны, более остро встал вопрос о необходимости выработки стандартов накопления все возрастающих объемов информации и создания стандартов и путей обмена этой информацией между различными заинтересованными сторонами – исследователями, исследовательскими центрами, хранилищами, специализированными центрами обработки информации (например, статистическими) государственными контролирующими организациями и т.д.

Следующим закономерным вопросом стал вопрос об объединении разрозненных источников накопленных данных и потребителей этих данных в единую систему,

как на внутреннем уровне стран, так и на уровне взаимодействия с мировым исследовательским сообществом, поскольку целый ряд фундаментальных исследований (астрономические, климатические, ядерные и др.) невозможно замкнуть в рамках одного государства. На настоящее время эта глобальная задача пока не решена, большинство стран прошли только этап формирования хранилищ и стандартов накопления информации, создания первоначальной инфраструктуры обмена данными и выделенных центров обработки информации. При этом самой сложной и до конца не решенной на сегодняшний день проблемой является выработка алгоритмов и методов поиска, идентификации и извлечения нужных данных из уже имеющихся систем и хранилищ, созданных на базе различных технических и организационных решений (оборудования, протоколов обмена информацией и т.п.). Квинтэссенцией выражения существующей проблемы может являться одна фраза, ставшая довольно известной: “We are drowning in data, but starving of information” («Мы захлебываемся данными, но испытываем жуткий голод в информации»). К слову сказать, весьма схожая проблема возникла в связи с ростом и развитием услуг и информации в самой сети Интернет, являющейся проводником распространения по планете информационно-коммуникационных технологий. В среде Интернета лавинообразно растет число открытых данных и публикаций ненаучного характера, касающихся всевозможных сфер человеческой деятельности. Эти данные представляют большой интерес для различных пользователей «мировой паутины» в целях развития бизнеса, поиска работы, объединения в сообщества и многих других. Для извлечения, группирования и анализа этих данных необходимы новые аналитические инструменты, которые бы могли по заданным пользователем критериям отбирать нужную информацию из всех доступных мест Интернета, включая источники на любых иностранных языках. В связи с этим возникло такое новое понятие, как «семантик Веб» (Semantic Web), с которым связано развитие нового Интернет-инструментария, основанного на принципах логистического и лингвистического подходов к сбору, анализу, селекции и объединению информации. Очевидно, что эти два направления работы с электронными данными научного и ненаучного характера в силу схожести ряда проблем будут подпитывать друг друга принципиально новыми алгоритмами обработки информации.

Выбранная тройка стран неравномерна как по географическим параметрам и возможностям бюджетов, так и по опыту создания систем для агрегации и обмена научной информацией, который будет представлен ниже.

Следует отметить, что при проведении исследований в отобранной по странам информации не раз фиксировалось, что наиболее «продвинутыми» в рассматриваемой нами теме являются США, которые также входят в первую двадцатку мирового рейтинга сетевой готовности стран. Что, впрочем, неудивительно – США располагают самым мощным бюджетом для достижения поставленной цели и, кроме того, ставят тему создания «Системы управления данными» (Data Management System – DMS) в один из главных своих приоритетов на пути обеспечения будущего процветания государства. В США существует четкое понимание того, что, государство, которое сможет наиболее эффективно использовать не только внутренние, но и мировые запасы научных данных в целях совершенствования своей системы исследований, производства, управления экономикой и выработки соответствующей внутренней и внешней политики, такое государство получит несомненные конкурентные пре-

имущества, для того чтобы играть ключевую роль в будущем мировом процессе. К сожалению, необходимого количества информации о современном уровне развития инфраструктуры данных в США в доступных источниках найти не удалось, в связи с чем США пришлось исключить из рассмотрения.

Четкое понимание важности вопроса создания мощной инфраструктуры обработки и хранения научно-исследовательских данных также присутствует в странах Европейского Союза, куда входят Германия и Нидерланды (обязанные следовать рамкам общих стратегий ЕС). В декабре 2011 года в ЕС был опубликован отчет «Взгляд на глобальную инфраструктуру исследовательских данных» [2] о подходах Европейского Союза к созданию к 2020 году Global Research Data Infrastructure (далее - ГИИД или GRDI2020) в рамках общей программы ЕС по научным исследованиям и технологическому развитию «Горизонт-2020». Отчет опубликован в рамках реализации проекта GRDI2020 (www.grdi2020.eu), реализация которого началась из средств 7-й Рамочной программы ЕС по научным исследованиям и технологическому развитию на период 2007–2013 гг. В рамках проекта была также разработана «дорожная карта» – Global Research Data Infrastructure 2020 Roadmap [3], предусматривающая последовательность шагов по реализации поставленной задачи.

Укрупненный подход ЕС, излагаемый в указанных документах, заключается в создании в ближайшем будущем «Экосистемы цифровых наук» (Digital Science Ecosystem), которая должна состоять из:

- библиотек цифровых данных (Digital Libraries – DL);
- архивов цифровых данных (Digital Data Archives – DDA);
- библиотек цифровых данных по исследованиям (Digital Data Libraries - DDL);
- исследовательских сообществ (Communities of Research).

Взаимодействие этих четырех компонентов составляет окружающую среду, в рамках которой должны осуществляться будущие исследования. С организационной и технической точек зрения взаимодействие в цифровой экосистеме должно основываться и поддерживаться с помощью механизмов функциональной совместимости этих четырех компонентов. Из этого подхода в первую очередь вытекает вопрос обеспечения совместимости элементов, отвечающих за накопление и хранение данных. Три компонента экосистемы выполняют эту функцию: библиотеки цифровых данных (DL); архивы цифровых данных (DDA); библиотеки цифровых данных по исследованиям (DDL). При этом подход ЕС предполагает, что место хранения данных не обязательно должно иметь национальную привязку (это важный момент для ЕС, куда входит двадцать семь различных государств, а пространство для взаимодействия исследователей должно быть построено общим), данные могут размещаться в любом хранилище, но при условии наличия на это соответствующего мандата (разрешения), как со стороны «размещающей», так и со стороны «принимающей стороны». Такой подход может использоваться как для международной (как в случае с ЕС), так и для внутренней национальной системы хранения и использования научных данных, поскольку система выдачи разрешений может быть построена на любом уровне взаимодействующих субъектов. На современном этапе развития элементов такой экосистемы слабым звеном у многих государств является

развитие системы архивов цифровых данных (DDA), в частности, из рассматриваемых в работе стран такая проблема существует у Канады. При весьма неплохом уровне развития цифровых библиотек, недостаток архивов цифровых данных является ахиллесовой пятой канадской системы, поскольку именно архивы содержат данные о местонахождении цифровых объектов в местах их хранения (в данном случае – цифровых библиотеках) и при их отсутствии резко сужается круг данных, доступных исследователям в такой цифровой экосистеме.

Следует отметить, что при некой общности подходов терминология, используемая в различных странах при рассмотрении вопросов создания систем агрегации и использования научных данных, нередко отличается, что создает некоторые сложности при анализе и сравнении этой информации.

3. Состояние формирования системы агрегации и использования научных данных и подходы к ее развитию в Канаде, Нидерландах и Германии

1. Канада

Информация о текущем состоянии развития канадской национальной инфраструктуры исследовательских данных получена из отчета о состоянии развития проекта «Канадская национальная инфраструктура данных» (The Canadian National Collaborative Data Infrastructure (CNCI) Project Final Report) [4]. Вопрос о реализации данного проекта был поднят благодаря созданию в 2010 году по инициативе Канадской ассоциации исследовательских библиотек (Canadian Association of Research Libraries – CARL) специального управляющего комитета при Канадском инновационном фонде (Canadian Foundation for Innovation – CFI) и рабочей группы по разработке моделей данных – Data Model Working Group – DMWG).

Из указанного отчета следует, что, несмотря на наличие в Канаде достаточно большого количества исследовательских библиотек (Research Data Library, начали создаваться с 1970 года) и центров хранения исследовательских данных (Research Data Centers) и безусловной заинтересованности многих научно-исследовательских институтов в формировании единой национальной инфраструктуры хранения и обработки данных, проект CNCI так и не получил финансирования ни через CFI, ни из других фондов. В его рамках разработано лишь несколько документов об общих подходах к созданию национальной инфраструктуры исследовательских данных и проведен ряд мероприятий, на которых обсуждались вопросы общих подходов к решению задачи, возможной структуры кооперации участников будущего проекта и источниках его финансирования. CARL и CFI продолжают взаимодействовать в поисках определения источников средств для реализации национальной программы по научным данным. При этом в отчете отмечается, что необходимость создания инфраструктуры научных данных и отставание Канады от Великобритании, Австралии, Европейского Союза, Нидерландов и США, которые уже выделили средства и реализуют соответствующие стратегии и программы, осознается на достаточно высоком уровне.

Таким образом, в настоящее время работа в направлении создания элементов бу-

дущей инфраструктуры данных Канады ведется только по проектам, реализуемым некоторыми отдельными научно-исследовательскими институтами или несколькими институтами в кооперации.

Проработка в рамках проекта CNCDI в целом завершилась выделением следующих трех основных уровней будущей инфраструктуры исследовательских данных Канады.

Уровень 1 – Резервированные хранилища (репозитории) данных (Redundant storage repositories). Планируется, что будет создано от двух до четырех крупных хранилищ с системами надежного резервирования сохраняемых данных, которые будут обеспечивать хранение и постоянный доступ к объектам данных, связанных с ними метаданных, данных по контролю за предоставлением доступа и прочую связанную с объектами данных информацию. Исследователи и пользователи библиотек будут получать доступ к данным через вторичные порталы.

Уровень 2 – Центры доступа к исследовательским инструментам (Disciplinary ingest centers). Предполагается, что будет создано небольшое количество координирующих центров, которые будут определять принадлежность задач к тем или иным видам исследований, управлять работой инструментов и предоставлять исследователям инструменты для скачивания необходимых данных и получения информации об их размещении в хранилищах.

Уровень 3 – Местные центры обслуживания (Microservice sites). Большое количество локальных центров (библиотек, центров хранения данных и т.п.) будут предоставлять услуги доступа, рекомендации по работе с данными, услуги локальных хранилищ и средств обработки данных и другую поддержку местным исследователям. Инструменты, обучающие материалы и прочая информация будут координироваться и предоставляться центрами второго уровня.

В целом можно отметить, что имеется существенный разрыв между Канадой и такими странами, как США, Германия, Нидерланды и др. в уровне готовности к формированию национальной инфраструктуры исследовательских данных. Сами канадцы в ряде публикаций отмечают, что «ахиллесовой пятой» на этом пути является острый недостаток архивов, которые должны содержать информацию (метаданные), описывающие характер и место нахождения цифровых объектов. Незрелость инфраструктуры архивов делает невозможным получение доступа к накопленной исследовательской информации.

2. Нидерланды

В основу анализа опыта построения национальной системы агрегации и использования научных данных Нидерландов положены материалы доклада о подходе Нидерландов к вопросам создания инфраструктуры исследовательских данных (The Dutch Approach to Research Data Infrastructure) [5], подготовленного сотрудниками центра архивирования и сетевых услуг DANS (Data Archiving and Networked Services), учрежденного Королевской академией искусства и науки Нидерландов и национальным исследовательским советом Нидерландов, а также сотрудниками некоммерческого объединения нидерландских научно-исследовательских и научно-прикладных организаций в сфере ИКТ – SURF (в настоящее время объединяет шестьдесят две организации, за счет ежегодных взносов которых оно существует;

это объединение реализует три основных программы: электронные исследования, электронное обучение и электронные администрации).

Согласно указанному докладу попытка Нидерландов выстроить к 2012 году логически связанную государственную политику развития этой сферы потерпела неудачу. Тем не менее, был создан целый ряд элементов и приняты документы, создающие предпосылки для формирования нидерландской инфраструктуры хранения и обработки исследовательских данных:

- образованы организация SURF (1986 год) и центр архивации данных DANS (2005 год);
- в 2008 году была принята национальная дорожная карта по развитию средств высокопроизводительных вычислений для нужд исследований (Large-scale Research Facilities);
- в 2008 году была создана нидерландская ассоциация по хранению цифровых данных (Netherlands Coalition for Digital Preservation);
- в 2008 – 2012 годах реализованы проекты «3TU.Datacentrum» по созданию специализированного архива научных данных, содержащего информацию не только локального, но и международного характера, и «Big Grid», в рамках которого были проведены работы по созданию сетевой инфраструктуры высокопроизводительных вычислений, включившей в себя вычислительные кластеры, системы хранения данных, а также программное обеспечение, позволяющее пользователям проводить исследования на этих мощностях;
- в 2011 году создан нидерландский центр электронных исследований (Netherlands eScience Center).

Национальная дорожная карта Нидерландов предусматривает создание двадцати пяти центров работы с данными в пяти исследовательских кластерах (research domains), из которых семь уже созданы, финансирование на эти цели предусматривается в размере 20 миллионов евро ежегодно. Выполняемая центрами работа с данными подразумевает выполнение функций управления, хранения, оценки качества, стандартизации, компоновки и формирования метаданных, необходимых для функционирования инфраструктуры хранения и обработки исследовательской информации. В 2012 году были выделены дополнительные 80 миллионов евро на модернизацию устаревшего оборудования и программного обеспечения центров.

Нидерландская ассоциация по хранению цифровых данных включила в себя библиотеки, DANS, национальные архивы, Нидерландский институт звука и изображения, музеи и культурные центры. В задачи ассоциации входит консультационная и методологическая помощь ее участникам в проведении работ по оцифровке хранимых в них объектов и организации их хранения.

Проект Big Grid (www.biggrid.nl) был реализован в кооперации между национальным компьютерным фондом Нидерландов NCF (National Computing Facilities Foundation, существует с 1990 года), Нидерландским институтом фундаментальных исследований в области субатомной физики Nikhef и Нидерландским центром биоинформатики NBIC. С их помощью и непосредственным участием в Нидерландах создана национальная GRID-инфраструктура, объединившая центры хранения

(data storage) и центры обработки информации (data processing facilities). Инфраструктура также включает в себя центры поддержки и консультаций, оказывающих поддержку исследователям в организации исследований по широкому спектру дисциплин – от области физических процессов до исследований в общественной и социальной сферах.

Проект 3TU.Datacentrum (www.datacentrum.3tu.nl, TU – Technical Universities) был реализован на базе трех технических университетов. В результате проекта создан архив данных по направлению технических исследований, позволяющий извлекать и использовать данные, расположенные в хранилищах не только Нидерландов, но и других стран. Основная задача создания архива – обеспечить постоянный доступ к данным исследований, средства архива позволяют находить данные, управлять процессами получения к ним доступа и извлекать данные с возможностью их повторного использования.

Нидерландский центр электронных исследований был создан по инициативе SURF и нидерландского фонда поддержки исследований NOW. Ежегодный бюджет центра составляет 6 миллионов евро. Деятельность центра сфокусирована на проведении мультидисциплинарных исследований, включая исследования на уровне школ. Задача центра – стимулирование инноваций в исследованиях на базе информационно-коммуникационных технологий.

На данный момент усилия SURF сконцентрированы на следующих приоритетах политики формирования нидерландской инфраструктуры хранения и обработки данных:

- развитие механизмов управления в данной сфере (хранение, защита, обеспечение доступа, обеспечение совместного использования);
- обеспечение должного развития «облачных» решений и применения их исследовательским сообществом;
- дальнейшее стимулирование развития кооперации между исследовательскими организациями на национальном уровне;
- развитие кооперации с международными (преимущественно европейскими) партнерами.

В свою очередь DANS определены следующие стратегические цели на период до 2015 года:

- проведение исследований в сфере изучения жизненного цикла данных (Data Life Cycle) от момента их возникновения до момента их уничтожения;
- развитие систем и механизмов, объединяющих потребителей и источники данных (исследователи, проекты, исследовательские организации);
- расширение инфраструктуры данных за пределы Нидерландов;
- развитие международной кооперации с европейскими и международными исследовательскими инфраструктурами.

3. Германия

Разработка немецкой национальной системы агрегации и использования научных данных ведется с 1999 года по заказу Федерального Министерства Образования и Исследований Германии. Подход к решению данного вопроса рассмотрим на примере развития системы хранения и обработки исследовательских данных социального и статистического характера, имеющих в Германии один из самых высоких приоритетов.

По решению указанного министерства для выработки соответствующих критериев развития и осуществления постоянного контроля над созданием инфраструктуры научных данных в 2004 году образована специальная Комиссия по усовершенствованию инфраструктуры сбора и обработки данных социальных исследований и статистики – German Data Forum (GDF или RatSWD). Основными задачами GDF являются:

- выработка рекомендаций по вопросам дальнейшего обеспечения безопасности и улучшения доступа к данным, в частности путем создания, стандартизации и оценки результатов научных исследований постоянными центрами хранения и обработки данных;
- выработка рекомендаций по улучшению использования данных путем предоставления соответствующей документации и научно-статистических данных (данные исследовательских порталов, метаданные);
- консультирование научных учреждений и организаций по включению генерируемых ими данных в образовательную и исследовательскую инфраструктуру;
- выработка рекомендаций по предметам исследований и задачам, которые относятся к концептуальным вопросам развития инфраструктуры данных на национальном, европейском и международном уровнях;
- выработка рекомендаций о том, как сделать производство и предоставления данных, относящихся к социальным исследованиям, более эффективными;
- консультирование министерств и уполномоченных государственных органов по вопросам дальнейшего развития инфраструктуры научных данных;
- консультирование общественных (и частных) производителей данных;
- консультирование производителей данных по вопросам проверки (сертификации) научно-исследовательских учреждений;
- подготовка и проведение конференций по вопросам социальных и экономических данных.

В основу национальной модели системы управления исследовательскими данными GDF положил создание двух типов субъектов хранения и обработки информации: Центров хранения исследовательских данных (Research Data Centers – RDC) и Центров обработки данных (Data Services Centers – DSC).

В общей сложности к настоящему времени в Германии существует сто сорок четыре RDC и DSC, накапливающих и позволяющих обрабатывать данные по исследованиям различной направленности. Из них двадцать пять центров, работающих с данными социальных и статистических исследований, прошли аккредитацию в GDF на соответствие разработанным GDF критериям формирования инфраструктуры исследовательских данных – Criteria of the German Data Forum (RatSWD) for the Establishment of Research Data Infrastructure (приняты в сентябре 2010 года). Число аккредитованных центров постоянно растет. Следует отметить, что аккредитация в German Data Forum не предполагает получение какого-либо доступа к государственному финансированию, интерес к участию формируется только благодаря предоставлению возможности использования данных хранилищ других участников German Data Forum.

С технической точки зрения вопрос доступа к данным решается благодаря использованию уже упоминавшейся технологии «метаданных» (metadata или микро-данных (microdata) – массивов структурированной информации, содержащих описание материалов исходного объекта данных, помещенного в центр хранения данных), что обеспечивает возможность предварительного анализа данных и извлечение из хранилища только информации, соответствующей запросу (например, только один абзац из всего документа), а также поиск по разным формам представления данных в одном запросе. Свои наработки по технологиям получения, обработки и представления исследовательской информации представил в 2011 году Технический университет Дортмунда в обзоре Vision of a Virtual Infrastructure for Storing and Processing Scientific Data [6], в котором рассматриваются вопросы создания технологий виртуальной среды (виртуальной инфраструктуры) для упрощения и унификации доступа к данным с использованием гибких параметров запросов и форм вывода информации. По замыслу разработчиков такая виртуальная среда призвана избавить исследователей от операций с данными, которые могут быть выполнены автоматически (посредством операционных средств этой среды) и, тем самым, позволить им сконцентрировать свою деятельность непосредственно на научных разработках.

Исходя из принятой в Германии модели использования научных и статистических данных, рабочее место исследователя должно оборудоваться программно-аппаратным интерфейсом работы с данными, позволяющим взаимодействовать с интерфейсом центра обработки данных (постановка задачи и получение результата), который в свою очередь взаимодействует с интерфейсом хранилища данных (обработка данных в соответствии с поставленной задачей). Алгоритмы работы и внутренние стандарты работы с данными различных центров обработки и хранилищ данных могут отличаться, поскольку, как правило, создавались в разное время и на базе технологий различных производителей. Интерфейсы призваны обеспечить перевод внутренних алгоритмов обмена данными на некий общий стандарт обмена данными на уровне виртуальной инфраструктуры, включающей в себя указанные интерфейсы, взаимодействующие через внешнюю сеть обмена данными. Виртуальность этой инфраструктуры заключается в том, что для различного вида задач архитектура работающей сети будет меняться, как по составу взаимодействующих центров обработки и хранения информации, так и по задействованным линиям пере-

дачи данных. При этом самые большие скорости должны обеспечиваться на линиях связи, соединяющих центры обработки и центры хранения данных.

4. Выводы

Изучение материалов об опыте и подходах Канады, Нидерландов и Германии к вопросу о создании национальных инфраструктур агрегации и обмена научной информацией позволяет сделать следующие заключения:

- все рассматриваемые страны осознают необходимость создания инфраструктур хранения и обработки данных (ИХОД) научных исследований и связывают с ними перспективы своего устойчивого экономического развития и укрепления конкурентных позиций на мировом рынке;
- ни одна из стран не считает, что национальная ИХОД уже построена, идет интенсивный процесс поиска ее оптимальной структуры и необходимых для ее работы технологических и организационных решений;
- все страны считают необходимым развитие национальных ИХОД с обязательной их интеграцией в ИХОД других государств-партнеров;
- вопросы создания национальной ИХОД, включая выбор технологий, определения источников финансирования и формирования кооперации исполнителей, каждая страна решает индивидуально в зависимости от достигнутого уровня развития отдельных элементов будущей инфраструктуры, доступных сетевых ресурсов и вычислительных мощностей. Организационные решения также базируются на принятых в каждой конкретной стране механизмах (традициях) и органах управления. Таким образом, не существует универсального прототипа национальной ИХОД. Ориентируясь на опыт других стран, можно использовать только схожие элементы и технологии, но не всю структуру полностью;
- к основным элементам построения всех ИХОД можно отнести (эти элементы могут иметь различные названия, но смысл их функционирования одинаков):
 - а) цифровые библиотеки;
 - б) хранилища данных (репозитории);
 - в) архивы данных (информационные центры, накапливающие метаданные с описанием исходных источников данных, их местоположения в сети, условий получения к ним доступа);
 - г) центры обслуживания пользователей, обеспечивающие исследователей необходимыми консультациями и инструментарием, помогающие осуществить постановку исследовательской задачи и получить доступ к источникам данных и ресурсам их обработки;
 - д) ресурсы обработки данных (суперкомпьютерные и др. центры обработки данных);
 - е) инфраструктуру обмена электронными данными между элементами ИХОД (GRID – сети).

Список литературы

1. The Global Information Technology Report 2014. The Networked Readiness Index 2014 // www3.weforum.org/docs/GITR/2014/GITR_OverallRanking_2014.pdf
2. GRDI 2020. A Vision for Global Research Data Infrastructures // www.grdi2020.eu/Repository/FileScaricati/fe80cfc2-aa24-4afb-8b32-0acfa9b76f15.pdf
3. GRDI 2020. Final Roadmap Report // www.grdi2020.eu/Repository/FileScaricati/e2b03611-e58f-4242-946a-5b21f17d2947.pdf
4. *Copeland L., Whitehead M.* The Canadian National Collaborative Data Infrastructure Project Final Report // www.carl-abrc.ca/uploads/pdfs/carl_cncdi_final_report.pdf
5. *Doorn P., Dupuis M., Vanderfeesten M.* The Dutch Approach to Research Data Infrastructure // www.slideshare.net/pkdoorn/surf-dans-prato-april-2011
6. Vision of a Virtual Infrastructure for Storing and Processing Scientific Data // www.cs.tu-dortmund.de/nps/de/Forschung/Publikationen/Graue_Reihe1/index.html

Information Technologies and Infrastructures for Research Data Aggregation. The Experience of Canada, Netherlands and Germany

Biktimirov M. R.* , Glebsky V. L.** , Dolgov B. V.*** , Polikarpov S. A.*

**Dorodnicyn Computing Centre of RAS, Vavilova str., 40, Moscow, 119333, Russia*

***Association "e-Arena", Vavilova str., 40, Moscow, 119333, Russia*

****ZAO Academinform, Leninsky av., 32A, Moscow, 119991, Russia*

Keywords: aggregation, research data, e-infrastructure

In the paper we focus on the experience of some Networked Readiness Index (NRI) leading countries. We are interested in their actions aimed at creating and using e-infrastructures and IT for the research data aggregation. As far as quite many of NRI leaders are EU countries, we also take into account EU approaches to the task mentioned. We study some steering papers and some particular IT-projects in the field as well. As a result of the analysis we ascertain that some approaches in different countries have much in common. All of the countries examined acknowledge the problem of e-infrastructures for research data storage and processing to be vital. The countries work hard in order to build and integrate it with the others. We can clearly see some differences caused by traditions and authorities. Sometimes we even could estimate the level of success of a particular country on the way to the problem solution. We also indicate some common elements for all e-infrastructures for research data storage and processing.

Сведения об авторах:

Биктимиров Марат Рамилевич,

Вычислительный центр им. А. А. Дородницына,

канд. техн. наук, зам. директора

Глебский Владимир Леонидович,

Ассоциация «е-Арена»,

советник по вопросам международного сотрудничества

Долгов Борис Вячеславович,

ЗАО «Академинформ»,

специалист

Поликарпов Сергей Алексеевич,

Вычислительный центр им. А. А. Дородницына,

канд. физ.-мат. наук, вед. науч. сотр.