

## **СЕТЕВЫЕ СТРУКТУРЫ ГУМАНИТАРНОЙ ИНФОРМАТИКИ: ТЕХНОЛОГИИ E-HUMANITIES**

**Л.И. Бородкин**

В статье отражены последние тенденции в развитии сетевых технологий в их применении к социально-гуманитарным наукам. Дается характеристика нового этапа развития e-Science, направленного на создание организационных и технологических структур, разработку соответствующего программного обеспечения для функционирования новой информационной среды с распределенными информационными ресурсами. Эти средства должны обеспечить доступ к огромным ресурсам индивидуальных пользователей, исследовательских групп, лабораторий и институтов (эти сообщества называют виртуальными организациями, collaboratories).

### **NETWORKING STRUCTURES OF THE HUMANITIES INFORMATION SCIENCE: TECHNOLOGIES OF E-HUMANITIES**

**L.I. Borodkin**

Article reflect recent trends in the development of networking technologies in their application to social and human sciences. Describes the new phase of development of e-Science building organizational and technological structures, development of software for the operation of the new information environment with distributed information resources. These funds must ensure access to the vast resources of individual users, research groups, laboratories and institutions (these community called virtual organizations, collaboratories).

Одно из инновационных направлений развития современной гуманитарной информатики связано с концепциями распределенных сетевых ресурсов и соответствующих технологий, создающими новую информационную среду гуманитарных исследований и образования. Это направление отражается, в частности, в материалах нового журнала – «International Journal of Humanities and Arts Computing» (IJHAC), появление которого (2007 г.) отражает тенденцию к усилению междисциплинарности современных исследований в области гуманитарных наук. Журнал представляет сотрудничество трех международных научных организаций: International Association for History and Computing, Cultural

Atlas Initiative, Digital Resources in the Humanities and Arts. Редакторы IJHAC – Дэвид Боденхамер (США) и Пол Элл (Великобритания); в состав редакционного совета журнала входят ученые из Австралии, Великобритании, Германии, Голландии, Норвегии, России, США, Тайваня. Приоритетная тематика, которую будет поддерживать журнал, включает технологии визуализации, цифровые медиа, моделирование, ГИС в гуманитарных приложениях, информационную архитектуру<sup>1</sup>, e-Science, Web-сервисы, опыт применения грид-технологий и другие направления современных разработок ИКТ. Журнал публикует также статьи и обзоры по электронным ресурсам и электронным публикациям (e-Resources и e-Publications) в области гуманитарных наук.

Появление здесь ряда новых для гуманитарной аудитории терминов и понятий требует комментария: ведь они в ближайшие годы войдут в активный лексикон гуманитариев. Этот процесс уже идет, его инициировали известные голландские специалисты по исторической информатике, авторы программной работы о прошлом, настоящем и будущем этой междисциплинарной области – Онно Боонстра, Питер Доорн, Лен Брёре<sup>2</sup>.

\* \* \*

Наиболее характерной чертой развития ИКТ в 2000-х гг. является формирование новых компонентов электронной среды взаимодействия. Этот процесс породил новые методологические подходы, программные и аппаратные решения, новые термины и понятия. Широкое распространение получили новые категории: e-Science, e-Social Science, e-data, e-Research, e-Publications, e-Culture, e-Learning, e-Business, e-Banking, e-Medicine, e-Government и т.д., связанные с развитием распределенной инфраструктуры информационных технологий и ресурсов. Наверное, излишним будет напоминание, что «е» означает здесь «electronic». Нас интересует прежде всего содержание термина «e-Science», возникшего изначально в Великобритании, где крупные исследовательские проекты

---

<sup>1</sup> Информационная архитектура – совокупность методов и приёмов структурирования информации (знаний или данных). Чаще всего термин «информационная архитектура» применяется в веб-разработке; в этом контексте информационная архитектура связывается прежде всего с принципами систематизации информации и навигации по ней с целью помочь пользователям более успешно находить и обрабатывать нужные им данные.

<sup>2</sup> Boonstra O., Breure L. and Doorn P. Past, Present and Future of Historical Information Science. Amsterdam, NIWI-KNAW. 2004. См. об этой книге: Бородкин Л.И. Историческая информатика в точке бифуркации: движение к Historical Information Science // Круг идей: алгоритмы и технологии исторической информатики. – Москва-Барнаул, 2005; Гарскова И.М. Историческая информатика и квантитативная история: преемственность и взаимодействие // Анализ и моделирование социально-исторических процессов. – М.: Ком-Книга, 2006.

в этой области начались в 2001 г. Именно там было дано первое определение e-Science, получившее в дальнейшем широкое распространение: «научно-технологическая область, в которой всё возрастающую роль играют распределенное глобальное взаимодействие посредством сети Интернет, с использованием очень больших коллекций данных, компьютерных ресурсов тера-уровня и высококачественной визуализации, доступных индивидуальному пользователю»<sup>3</sup>. Надо сказать, что в русском языке термин «e-Science» существует пока преимущественно в англоязычном варианте<sup>4</sup>.

Иногда возникает вопрос: в чем отличие концепции e-Science от хорошо известной нам глобальной сети Интернет? Ведь пользователь Интернета имеет доступ со своего компьютера к разнообразным ресурсам глобальной сети, размещенным на сайтах всех стран мира; при этом набор веб-сервисов непрерывно расширяется. Дело в том, что кроме «обычной» информации, размещенной в Интернете, специалисты, работающие в сфере науки и образования, нуждаются в доступе к крупномасштабным информационным массивам, базам данных, имеющим объемы памяти, измеряемые терабайтами. Работа с такими массивами требует вычислительных мощностей с производительностью уровня терафлоп. Обычный пользователь Интернета работать в этом режиме не сможет. Сегодня потребность в таких крупномасштабных ресурсах и компьютерных мощностях испытывают в основном ученые-естественники, но в ближайшем будущем к этому рубежу подойдут и разработчики проектов в различных областях социально-гуманитарного знания. Так, уже сейчас историками созданы огромные базы данных, например в области исторической демографии. Эти базы содержат миллионы (а в отдельных случаях – десятки миллионов) записей, основанных на первичных материалах переписей населения.

Задача e-Science, таким образом, – создание организационных и технологических структур, разработка соответствующего программного обеспечения для функционирования новой информационной среды с распределенными ресурсами (информационными и вычислительными), обеспечивающих доступ к ним индивидуальных пользователей, исследовательских групп, лабораторий и институтов (эти сообщества называют виртуальными организациями, *collaboratories*).

Основное русло реализации задач e-Science прокладывают грид-технологии. Эта концепция (нередко ее называют Grid Computing – рас-

---

<sup>3</sup> См., например: <http://www.escience.am>, <http://www.nesc.ac.uk>

<sup>4</sup> В отличие, например, от термина e-Government, который получил распространение в русском языке как «электронное правительство».

пределенные сети, или «решетки» вычислительных ресурсов) соответствует одному из ведущих и перспективных направлений развития ИКТ. В последние годы начата разработка средств поддержки Grid на уровне стандартов и продуктов. Ситуацию можно сравнить с состоянием сети WWW в конце 1990-х гг., при этом специалисты предсказывают, что Grid произведет такую же революцию в области обработки данных, какую Интернет произвел в сфере компьютерных коммуникаций. Так, корпорация Oracle развивает свои продукты в направлении более широкого применения Grid-технологий (символично, что в обозначении очередной, десятой, версии Oracle появилась латинская буква g). Следует отметить, что в платформенных решениях Oracle и других ведущих поставщиков ИКТ уже давно наметилась ориентация на распределенные вычисления на основе кластерных систем (в том числе на платформе Intel Itanium и Linux)<sup>5</sup>. Grid-технологии используют новую архитектуру программного обеспечения, предназначенную для объединения большого количества недорогих модулей памяти и серверов, создающих распределенный виртуальный ресурс, транспарентный и легко доступный для пользователя (независимо от места нахождения его компьютера).

Одна из наиболее сильных сторон Grid-технологий – новые возможности эффективного распространения информации и обучение с доступом к обширным ресурсам<sup>6</sup>. Благодаря всё более широкому применению современных ИКТ в образовании, доступ к грид-инфраструктуре получают школы и вузы, исследовательские группы и широкий круг пользователей различных приложений, что приведёт к качественным изменениям в учебном процессе, научных исследованиях и, в конечном итоге, в разных видах экономической деятельности. При этом специфика термина «e-Social Sciences» заключается не только в предметной области используемых распределенных информационных ресурсов, но и в акценте на взаимодействие ученых-обществоведов со специалистами по ИКТ<sup>7</sup>.

В этой связи представляет интерес деятельность Национального центра по развитию e-социальных наук (NCeSS)<sup>8</sup>, основанного пять лет

---

<sup>5</sup> См., например: <http://www.interface.ru/home.asp?artId=10375>.

<sup>6</sup> Здесь возникает известная аналогия с единой энергетической системой, включающей множество производителей энергоресурсов, поставляющих электроэнергию в общую сеть, к которой подключены миллионы потребителей («пользователей»), получающих доступ к сети, когда они в этом нуждаются (и не интересующихся, откуда конкретно пришла к ним энергия).

<sup>7</sup> О подходах к созданию структур новой информационной среды в российской науке см.: *Паринов С.И.* E-Science – онлайн-будущее науки // Приложение к журналу «Информационные технологии». – 2007. – №9.

<sup>8</sup> <http://www.ncess.ac.uk/>

назад Британским советом по экономическим и социальным исследованиям (ESRC). Задача этого центра – анализ особенностей использования грид-технологий в социальных науках, разработка средств их поддержки, подготовка рекомендаций по стратегии развития е-социальных наук. NCeSS предоставляет также консультационные услуги, обучение и доступ к онлайн-ресурсам – для помощи исследователям из различных областей социального знания в их адаптации к возможностям и реалиям е-социальных наук. NCeSS включает координационный хаб (центр) в Манчестерском университете, семь исследовательских групп и 12 небольших исследовательских проектов, ориентированных как на количественные, так и на качественные подходы тех или иных социальных наук. Так, один из этих проектов (Collaboratory for Quantitative e-Social Science – CQeSS), действующий в Ланкастерском университете, имеет целью содействие квантитативным исследованиям в различных областях социальных наук, включая развитие исследовательского аналитического инструментария, необходимого для обработки больших массивов количественных данных и доступного в режиме «распределенного использования». Речь идет о таких грид-разработках, как SABRE, GROWL, SAKAI<sup>9</sup>.

Что касается гуманитарных наук, то перспективы использования грид-технологий связываются здесь во многом с расширяющимися возможностями онлайн-доступа к огромным массивам оцифрованных книг, журналов, изобразительных материалов и т.д. Эффективность использования этих ресурсов резко возрастает, когда исследователь может связывать не только различные тексты, но и тексты с изобразительными источниками, историческими картами, записями интервью, видеоматериалами и т.д., размещенными на различных сайтах. Сегодня в Великобритании осуществляется поддержка четырех проектов по развитию виртуальной среды исследований в области гуманитарных наук, девяти регулярных научных семинаров и семи исследовательских проектов в конкретных областях гуманитарного знания (археология, музейное дело, классические исследования, история Византии)<sup>10</sup>. В этих проектах используются различные технологии e-Science, включая обработку изображений, текста древних манускриптов, интеграцию национальных баз данных по материалам археологических экспедиций, компьютерное моделирование средневековых битв, использование 3-D сканирования для анализа поверхности объектов музейного хранения, ГИС. Сущест-

---

<sup>9</sup> <http://www.ncess.ac.uk/research/quantitative/>

<sup>10</sup> Грантовая поддержка этих проектов проводится усилиями Arts and Humanities Research Council (AHRC) и UK e-Science Core Programme.

венно, что работа над этими проектами идет в тесном сотрудничестве гуманитариев и специалистов из области ИКТ. Это междисциплинарное взаимодействие необходимо еще и потому, что массивы данных, анализируемые гуманитариями (в частности, историками), имеют более сложную, «размытую» («fuzzy») структуру в сравнении с данными, полученными в естественнонаучных исследованиях.

Подобные проекты ведутся не только в Великобритании<sup>11</sup>. П. Доорн и П. Алькховен отмечают заметную роль голландской программы SATCH, развивающей возможности доступа к оцифрованным объектам культурного наследия в рамках активного сотрудничества между ИТ-специалистами, институтами культурного наследия и учеными-гуманитариями. Участники десяти проектов этой программы разрабатывают соответствующее программное обеспечение, проводят эксперименты, используя грид-технологии хранения данных, в сотрудничестве с голландским Big Grid Project. Для усиления возможностей доступа к оцифрованным данным гуманитарного профиля Королевская Академия наук и искусств Голландии (KNAW) и Голландский научный фонд (NWO) создали новый институт – DANS (Data Archiving and Networked Services). DANS функционирует как сетевая организация с центром, ответственным за организацию инфраструктуры данных<sup>12</sup>. В Германии развивается так называемая D-Grid инициатива, в рамках которой создается TextGrid программа, ориентированная на совместную деятельность сетевого сообщества ученых, проводящих редактирование, аннотирование, анализ и публикацию текстов. Дальнейшая цель программы – разработка технологий семантических сетей (Semantic Grid)<sup>13</sup>.

На европейском уровне подобные институты и программы работают вместе под эгидой DARIAH (Digital Research Infrastructure for the Arts and Humanities). Цель этого центра – обеспечить координацию усилий по сохранению культурного наследия в Европе и обеспечению доступа к материалам гуманитарных исследований<sup>14</sup>.

Эти разработки дают основания для введения термина «e-Humanities» (менее принятого пока, чем e-Science)<sup>15</sup>. В работе Й. Кирца

---

<sup>11</sup> Alkhoven P., Doorn P. New Research Perspectives for the Humanities // International Journal of Humanities and Arts Computing. – 2007. – Vol. 1, №. 1. – P. 35–47.

<sup>12</sup> Ibid. – P. 45. Отметим, что директором-организатором DANS является д-р П. Доорн, президент АНС.

<sup>13</sup> Ibid.

<sup>14</sup> Ibid.

<sup>15</sup> Ibid. – P. 35.

рассмотрены десять направлений разработок, развитие которых приведет к повышению роли e-Humanities в развитии гуманитарных наук<sup>16</sup>.

1) создание больших коллекций оцифрованных материалов (таких, например, как корпус источников);

2) смена акцентов в образовательных программах, ориентированных на овладение навыками работы со стандартными пакетами: предлагается изучать компьютерные методы и технологии, связанные с решением исследовательских задач, характерных для соответствующей области гуманитарного знания;

3) helpdesks: их роль на гуманитарных факультетах должна измениться, они должны содержать больше возможностей для работы студента-гуманитария в новой исследовательской информационной среде, обеспечивать использование специализированного программного обеспечения;

4) установление более тесных связей гуманитарной информатики (humanities computing) с «базовой» информатикой;

5) обеспечение более широкого доступа к массивам электронных данных; они должны быть приложенными к соответствующей публикации; архивы машиночитаемых данных должны принимать их на хранение; необходимо разработать стандарты метаданных для упорядочения электронных библиотек, архивов и других ресурсов;

6) создание библиотек программного обеспечения, прошедшего апробацию в исследовательских проектах; они должны быть хорошо документированными;

7) развитие веб-сервисов, которые повысят возможности использования данных и ресурсов в различных исследованиях;

8) разработка схем метаданных становится критическим аспектом развития e-humanities в ближайшие годы;

9) разработка основанных на онтологиях авторских программных средств (а также программного обеспечения, созданного в коллективных проектах) – важная составляющая процесса; структура метаданных при этом должна быть согласована с соответствующей структурой, принятой в архивах машиночитаемых данных;

10) развитие удобного интерфейса и средств визуализации – для повышения эффективности исследований мультимедийных данных, демографических и социально-экономических рядов, материалов археологических экспедиций и т.д.

---

<sup>16</sup> *Kitcz J.* E-based Humanities and E-Humanities on a SURF platform. (A report commissioned by SURF-DARE) // KRA publishing research. – 1 June 2004. – Amsterdam.

Как отмечают П. Доорн и П. Алькховен, наиболее высокие оценки экспертов получили пп. 1, 4 и 2, что подчеркивает приоритетность создания коллекций оцифрованных материалов, трансформации образовательных программ в области ИКТ-приложений и более тесной связи «отраслевой» гуманитарной информатики с «базовой» информатикой<sup>17</sup>. Ш. Росс, директор Центра гуманитарной информатики Института информации Университета Глазго также рассматривает в качестве важнейшей задачи развития e-Humanities в ближайшем десятилетии обеспечение хранения оцифрованных данных и онлайн-доступа к ним<sup>18</sup>.

\* \* \*

Одним из наиболее известных проектов в области грид-технологий является проект EGEE – Enabling Grids for E-sciencE, направленный на создание международной грид-инфраструктуры<sup>19</sup>. Проект выполняется консорциумом из 70 институтов в 27 странах, объединенных в региональные гриды. В результате создается самый крупный в мире Грид с суммарной вычислительной мощностью 20000 ЦПУ. Европейское научно-исследовательское сообщество получает в распоряжение общий рынок компьютеринга, обеспечивающий круглосуточный доступ к крупнейшим вычислительным ресурсам. Доступ не будет зависеть от места расположения потребителей и будет основываться на научных сетях Geant и NRNs. Российские участники образовали региональный консорциум РДИГ (Российский грид для интенсивных операций с данными – Russian Data Intensive GRID, RDIG)<sup>20</sup>. Наиболее известное приложение проекта – это обработка данных на Грид в глобально распределенном режиме с Большого адронного коллайдера (Large Hadron Collider – LHC), который построен в ЦЕРНе (Женева). Эксперименты на коллайдере LHC порождают потоки данных с экстремальными характеристиками: скорость 100 Мбайт/с, общий объем – несколько петабайтов в год. По прогнозам специалистов, в полном варианте концепция Grid Computing будет реализована в рамках архитектуры Open Grid Services Architecture, которая фактически должна привести к созданию динамических Web-сервисов.

Между тем веб-сервисы в последние годы получили новый импульс к развитию, который привел к формированию концепции Web 2.0. По-

---

<sup>17</sup> Alkhoven P., Doorn P. Op. cit. – P. 38–39.

<sup>18</sup> Ibid. – P. 42.

<sup>19</sup> Проект финансировался Европейской комиссией в рамках 6-й Рамочной программы IST FP6. Подробнее о проекте см.: <http://www.eu-egee.org/>

<sup>20</sup> <http://www.egee-rdig.ru>

явление термина Web 2.0 принято связывать со статьёй Тима О'Рейли «What Is Web 2.0» от 30 сентября 2005 г.<sup>21</sup>, впервые опубликованной на русском языке в журнале «Компьютерра» (№37 (609) и 38 (610) от 14 и 19 октября 2005 г., соответственно) и затем выложенной под заголовком «Что такое Веб 2.0» на веб-сайте «Компьютерра online»<sup>22</sup>. В этой статье Т. О'Рейли связал появление большого числа сайтов, объединённых некоторыми общими принципами, с общей тенденцией развития интернет-сообщества и назвал это явление Web 2.0 в противовес «старому» Web 1.0. Несмотря на то, что значение этого термина до сих пор является предметом многочисленных споров, те исследователи, которые признают существование Web 2.0, подчеркивают фактор широкого участия «обычных» пользователей глобальной Сети в формировании и развитии коллективных интернет-ресурсов и выделяют несколько основных аспектов и технологий этого явления: Wiki – технология коллективного создания гипертекста; Flickr – социальный фотосервис; YouTube – социальный видеосервис; Google Maps – карты и спутниковые снимки Земли; Blogs – сетевые дневники; Del.icio.us – он-лайн закладки; Netvides – сетевой десктоп и др.

\* \* \*

Какое место занимают в последние годы отмеченные выше тенденции в разработке научно-образовательных ресурсов, использовании новых ИКТ в обучении студентов-историков? Рассмотрим этот вопрос на примере тематики ежегодных конференций американской ассоциации «History and Computing». Состоявшаяся в апреле 2008 г. конференция американских специалистов в области исторической информатики имела характерное название: Web 2.0 / History 2.0: Making History Together. В центре внимания участников были следующие вопросы:

- Что включает «исторический сегмент» Web 2.0?
- В чем его отличия от соответствующего сегмента Web 1.0?
- Что заставляет нас выходить за пределы возможностей Web 1.0?
- Каковы приложения Web 2.0 в исторических исследованиях и образовании?
- Каковы «плюсы» и «минусы» Web 2.0 в этих приложениях?

Рассмотрение данных вопросов потребовало обращения к конкретным аспектам сетевого взаимодействия, указанным организаторами конференции: программы «Open source», используемые для интерпретации истории; тэги в исторических приложениях; технологии Wiki, блоги и RSS в исторических приложениях; право собственности на дан-

---

<sup>21</sup> <http://www.oreilynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

<sup>22</sup> <http://www.computerra.ru/think/234100/>

ные; базы данных и исторический анализ; картографирование истории (документация, основанная на возможностях Google); создание звуковых и видео-передач (подкастов), цифровое видео (например, YouTube), виртуальные миры в исторических приложениях.

Следующая конференция этой ассоциации, состоявшаяся в апреле 2009 г., имела название «Frontiers in Digital History»; авторы большинства докладов использовали термин «цифровая история» в связи с разработкой больших тематических ресурсов и их применением в образовании и исследованиях. Ряд докладов был обращен к анализу новых возможностей доступа к оцифрованным архивным фондам (Online Digital Archives), к распределенным программным средствам «цифровой истории» (Digital History Tools), к использованию распределенных цифровых медиа-ресурсов в обучении студентов-историков (Teaching History with Digital Media).

Как нам представляется, одна из задач, стоящих перед преподавателями кафедр, связанных с методологией и методами гуманитарно-социальных наук, – отразить отмеченные выше тенденции в соответствующих учебных планах, включив отечественный опыт развития концепций e-Humanities.