

ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

С.И. Паринов

**ОНЛАЙНОВОЕ БУДУЩЕЕ НАУКИ:
НАУКОМЕТРИЧЕСКАЯ
СИГНАЛЬНАЯ СИСТЕМА**

Препринт WP2/2007/01

Серия WP2

Количественный анализ в экономике

Москва
ГУ ВШЭ
2007



Издание осуществлено в рамках
Инновационной образовательной программы ГУ ВШЭ
«Формирование системы аналитических компетенций
для инноваций в бизнесе и государственном управлении»

Редактор серии WP2
«Количественный анализ в экономике»
В.А. Бессонов

Паринов С.И. Онлайнное будущее науки: наукометрическая сигнальная система:
П 18 Препринт WP2/2007/01. – М.: Издательский дом ГУ ВШЭ, 2007. – 52 с.

Перенос деятельности ученых и научных организаций в онлайнную среду характеризуется: а) повышением сложности и комплексности информационных систем, которые обслуживают научную деятельность; б) интеграцией этих систем в единое целое на базе национальной исследовательской онлайнной инфраструктуры; в) созданием на этой основе более совершенных организационных и индивидуальных онлайнных поведенческих моделей. Последний аспект в существенной степени зависит от наличия онлайнной наукометрической сигнальной системы, характеризующей состояния и динамику различных параметров деятельности ученых, включая степень использования результатов их исследований профессиональным сообществом. Предлагаемая статья посвящена анализу данных тенденций и описанию уже действующей на базе Соционет системы сбора и обработки онлайнной статистики, характеризующей определенные аспекты деятельности ученых и исследовательских организаций Отделения общественных наук РАН. Предложена онлайнная метрика для построения наукометрических показателей. Описаны алгоритмы и примеры онлайнных показателей востребованности, а также предложены методы построения показателей активности и индексов влияния для деятельности отдельных ученых, подразделений и институтов Отделения общественных наук РАН.

УДК 004.031.4:001
ББК 32.81

Классификация JEL: C81, C88.

Ключевые слова: научная онлайнная инфраструктура, статистическая база, наукометрические исследования, институты Отделения общественных наук РАН, Соционет.

Сергей Паринов, sparinov@hse.ru
*д.т.н., главный научный сотрудник ЦЭМИ РАН, профессор ГУ ВШЭ,
руководитель рабочей группы Соционет*

Препринты ГУ ВШЭ размещаются на сайте:
<http://new.hse.ru/C3/C18/preprintsID/default.aspx>.

© Паринов С., 2007
© Оформление. Издательский дом ГУ ВШЭ, 2007

Оглавление

1. Введение.....	4
2. Институты ООН РАН как полигон для построения e-Science	13
3. Онлайнные наукометрические показатели.....	21
3.1. Статистическая база для онлайнных показателей	22
3.2. Метрика онлайнных наукометрических показателей.....	26
3.3. Показатели востребованности.....	29
3.4. Показатели активности.....	36
3.5. Показатели влияния.....	37
4. Наукометрическая сигнальная система – использование онлайнных показателей.....	41
5. Заключение	42
5.1. Статус публичного ученого.....	44
5.2. Онлайнная модель фиксации результата исследования	46
6. Литература	49

1. Введение¹

В научной среде в последние годы набирает силу вторая волна² массового внедрения Интернет-технологий, автоматизирующих и развивающих различные аспекты научно-исследовательской деятельности, что в свою очередь создает условия для изменений организационных и индивидуальных поведенческих моделей деятельности в науке. Данные инновации имеют три основные компоненты: а) комплексная автоматизация на базе Интернет-технологий научно-исследовательской деятельности (системы управления проектами, грантами, публикациями и т.д. и т.п.), называемая в международной научной среде Current Research Information Systems (CRIS)³; б) формирование онлайн-научной инфраструктуры, называемой в западной литературе «research e-infrastructure»⁴, предназначенной для интеграции CRIS отдельных исследовательских организаций в целях создания единой межнациональной онлайн-среды для исследований; в) извлечение социально-экономических выгод из данных технических инноваций путем создания более эффективных поведенческих моделей научной деятельности как для отдельных ученых, так и для исследовательских организаций, что часто определяется как развитие социальных аспектов «e-Science»⁵.

¹ Автор выражает благодарность В.Л. Макарову и М.Д. Ильменскому (ЦЭМИ РАН) за моральную и организационную поддержку, М.Р. Коголовскому (ИПР РАН) за помощь в содержательном и стилистическом улучшении данного текста, а также М.Е. Прохорову (ГАИШ МГУ) за практическую реализацию системы сбора и обработки наукометрической статистики в Соционет. Существенная часть работы выполнена при поддержке грантов Фонда Форда (проект Соционет, 2000–2007 гг.) и 6-й Рамочной программы Европейской комиссии (проект TECHBRID, 2004–2006 гг.). В 2000–2004 гг. данные гранты были получены на ИЭОПП СО РАН. Последний грант Фонда Форда – на ГУ ВШЭ.

² Первая волна была связана с овладением отдельными учеными и исследовательскими организациями электронной почтой и созданием веб-сайтов.

³ См. подробнее: <http://www.eurocris.org/>

⁴ Office of Science and Innovation. Developing the UK's e-infrastructure for science and innovation. Report of the OSI e-Infrastructure Working Group, 2007 // <http://www.nesc.ac.uk/documents/OSI/index.html>, <http://www.nesc.ac.uk/documents/OSI/report.pdf>

⁵ Термин «e-Science» появился в конце 1990-х гг. для обозначения главным образом технических аспектов: интеграции географически распределенных вычислительных систем, научных инструментов, баз данных и других источников информационных ресурсов, сенсоров и т.п., что в настоящее время относят к CRIS и research e-infrastructure. В последние годы понятие «e-Science» все чаще используется для обозначения социальных аспектов и последствий внедрения Интернет-технологий в науку.

Общая схема этих инноваций очевидна: сначала разработка и реализация более совершенной программно-технической среды (CRIS + e-infrastructure) для профессиональных взаимодействий, что затем создает возможности для изменения форм и методов профессиональных взаимодействий в науке (e-Science) в результате организационного реинжиниринга и изменения «правил игры» (социально-экономических институтов).

Важный контекст для текущих изменений процесса научных исследований и разработок на базе Интернет-технологий задают международные инициативы, связанные с реализацией идей открытого доступа к результатам научных исследований:

- Инициатива «Открытые архивы», 2000 г. (<http://www.openarchives.org/>);
- Будапештская инициатива «Открытый доступ», 2001 г. (<http://www.soros.org/openaccess/ru/read.shtml>);
- Берлинская декларация об открытом доступе к научному и гуманитарному знанию, 2003 г. [1];
- Международное соглашение «Берлин-3», 2005 г. [2];
- Международная петиция за гарантированный публичный доступ к результатам исследований, финансируемых Европейской комиссией, 2007 г. (<http://www.ec-petition.eu/>).

Организации, подписавшие в 2003 г. Берлинскую декларацию об открытом доступе к научному и гуманитарному знанию, заявили о следующих намерениях ([1]):

- «стимулировать наших исследователей/получателей грантов публиковать свои работы согласно принципам парадигмы Открытого доступа;
- стимулировать держателей культурного наследия поддерживать Открытый доступ, обеспечивая их Интернет-ресурсы;
- разрабатывать средства и способы оценки вкладов в Открытый доступ и сетевые журналы для того, чтобы поддерживать стандарты гарантии качества и хорошей научной практики;
- добиваться, чтобы публикации в системе Открытого доступа признавались при присуждении ученых степеней и решении о занятии преподавательских должностей;
- добиваться высокого качества вкладов в инфраструктуру Открытого доступа путем развития программных средств, поставки контента и создания метаданных или публикации индивидуальных статей».

В 2005 г. было принято дополнительное соглашение, названное «Берлин-3» [2], призванное детализировать практические действия, которые рекомендуется выполнить исследовательским организациям в рамках Берлинской декларации.

«В целях реализации Берлинской декларации научные организации (институты) должны:

- 1) требовать от своих научных сотрудников выкладывать в онлайн-овые архивы с открытым доступом электронные копии всех опубликованных статей;
- 2) поощрять своих научных сотрудников публиковать статьи в журналах с открытым доступом к материалам, если подходящие журналы имеются (и обеспечивать сотрудникам поддержку в данных действиях)» [3].

В дальнейшем последовали публикации (см., например, [4]), еще более конкретизирующие сценарии поведения организаций и исследователей, действующих на базе Берлинской декларации. В частности, научно-исследовательским организациям рекомендовалось для максимизации своего научного вклада взять на себя обязательство самоархивирования статей и материалов с результатами исследований. Организациям, финансирующим исследования (правительственные и частные), рекомендовалось, в свою очередь, сделать обязательным условием выделения грантов для финансируемых ими исследований требование самоархивирования результирующих материалов в открытых архивах организаций грантополучателей.

На индивидуальном уровне ученым рекомендовалась следующая последовательность действий [4]: автор (или грантополучатель) (1) сразу после принятия статьи к публикации (2) депонирует в репозиторий организации – своего места работы (3) полный текст и библиографические метаданные (автор, дата, название, журнал и т.д.) (4) окончательного авторского текста статьи (но не файл подготовленного к печати текста, принадлежащий издательству). В рекомендациях «Берлин-3» отмечается, что необходимо гарантировать только депонирование. Открытие доступа к полному тексту остается привилегией автора. Установление открытого доступа к полному тексту настоятельно рекомендуется, но не является обязательным. Эти рекомендации делают обязательства исследовательских организаций по самоархивированию абсолютно независимыми от стратегии издателей в области самоархивирования.

К 2006 г. определение открытого доступа (в трактовке JISC – одной из основных организаций, финансирующих научные исследования в области информатики в Великобритании) выглядит следующим образом⁶:

«Что есть открытый доступ:

Исследовательская литература в открытом доступе представляет собой бесплатные доступные в он-лайн копии статей рецензируемых журналов

⁶ См. документ на сайте Joint Information Systems Committee (JISC) // http://www.jisc.ac.uk/publications/publications/pub_openaccess.aspx

и статьи из трудов конференций, а также технические отчеты, диссертации и препринты. В большинстве случаев не существует лицензионных ограничений для доступа к ним на чтение. Они могут быть свободно использованы для исследований, обучения и других целей».

«Что не является открытым доступом:

...Это не самиздат, не путь для уклонения от рецензирования и контроля качества, принятого для публикации, не вид второсортного и дешевого опубликования материалов. Это просто средство сделать результаты исследований доступными в он-лайн для всего сообщества исследователей».

В 2006 г. рекомендации «Берлин-3» были еще более детализированы. Для организации открытого доступа к результатам исследований в настоящее время предлагается два пути⁷:

- 1) «золотой» – публикация в журналах, обеспечивающих открытый доступ к своим статьям (либо за счет взимания платы с автора-организации за рецензирование / публикацию выпускаемых им статей, а не с организации-пользователя за доступ к статьям, либо сделав свободным для всех доступ к онлайн-версии журнала);
- 2) «зеленый» – самоархивирование в открытых архивах (ОА), в процессе которого авторы обеспечивают открытый доступ к их собственным уже опубликованным статьям, делая бесплатным свои е-принты (eprints – электронные материалы) для всех.

Уточняются и основные понятия: «Самоархивирование в ОА не является самиздатом, это не онлайн-публикация без контроля качества (рецензирования), оно не подходит для материалов, за которые авторы хотят получить вознаграждение, как книги или заказные статьи в газеты и “глянцевые” журналы. Самоархивирование в ОА предназначено для прорецензированных исследовательских статей, написанных в большей степени для распространения идей и достижений, чем для получения вознаграждения за них»⁸.

С 2007 г. участники различных международных и национальных инициатив за обеспечение свободного доступа к результатам исследований проводят знаменательную акцию – онлайн-подписание документа «Петиция за гарантированный публичный доступ к результатам исследований, финансируемых из общественных фондов»⁹, обращенного к Европейской комиссии (по состоянию на 08.02.2007 г. под петицией стояло 18932 подписей, из них 67 из России¹⁰).

⁷ Раздел «What is Open Access?» // <http://www.eprints.org/openaccess/>

⁸ Там же.

⁹ Petition for guaranteed public access to publicly-funded research results // <http://www.ec-petition.eu/>

¹⁰ <http://www.ec-petition.eu/res/files/statistics.pdf>

Петиция призывает Европейскую комиссию незамедлительно ввести в действие первую рекомендацию из отчета «Study on the Economic and Technical Evolution of the Scientific Publication Markets of Europe», опубликованного в январе 2006 г. как официальный документ Еврокомиссии:

«Рекомендация А1. Гарантировать свободный доступ к результатам исследований, финансируемых из общественных источников, сразу после их публикации»¹¹.

«Следующие мероприятия должны быть реализованы на общеевропейском уровне: (i) принять европейскую политику, обязывающую выкладывать опубликованные исследования, профинансированные из средств Европейской комиссии, после заданного временного периода в архивы открытого доступа, и (ii) совместно с государствами — членами Европейского союза и с европейскими исследовательскими и академическими ассоциациями рассмотреть, как еще данная политика и открытые репозитории могут применяться»¹².

В петиции также содержится рекомендация, согласно которой любое потенциальное «эмбарго» на открытый доступ к результатам исследований не может устанавливаться больше чем на 6 месяцев с момента их публикации.

Если рассмотреть отдельно только область инноваций, связанных с открытым доступом, то приметам второй волны интернетизации сферы научных исследований и разработок будет следующая цепочка логически взаимосвязанных действий и их результатов, часть из которых уже практически реализована в некоторых странах, другие еще только обсуждаются в международных научных блогах¹³ и форумах¹⁴:

Отправная точка второй волны интернетизации:

- 1) исследовательские организации создают свои *онлайн-репозитории*¹⁵, через которые открывают свободный доступ к статьям с результатами исследований и другим электронным материалам;

¹¹ RECOMMENDATION A1. GUARANTEE PUBLIC ACCESS TO PUBLICLY-FUNDED RESEARCH RESULTS SHORTLY AFTER PUBLICATION, см. <http://www.ec-petition.eu/>

¹² <http://www.ec-petition.eu/>

¹³ Open Access Archivangelism // <http://openaccess.eprints.org/>

¹⁴ American Scientist Open Access Forum // <http://amsci-forum.amsci.org/archives/American-Scientist-Open-Access-Forum.html>

¹⁵ На конец мая 2007 г. реестр ROAR (<http://roar.eprints.org/>) содержал 887 таких репозиториев. Из них 5 — российских. Главное отличие онлайн-репозиториев от веб-сайтов заключается в том, что каждый размещенный в них материал имеет стандартизованное описание (метаданные), к которым открыт технический «вход» для систем автоматической интеграции метаданных в информационные коллекторы.

- 2) сотрудники этих организаций (или назначенные ответственные) *самоархивируют (выкладывают)*¹⁶ в онлайн-репозитории статьи с результатами своих исследований, а также другие научные материалы;
- 3) для создания действенных мотиваций к самоархивированию финансирующие исследования организации (исследовательские институты, центры и фонды) *обязывают ученых*¹⁷ выкладывать результаты исследований в форме статей¹⁸ в онлайн-репозитории организаций¹⁹ (например, по месту работы), что, в том числе, означает немедленное электронное депонирование содержащегося в статьях научного приоритета автора, а также создает благоприятные предпосылки для 100%-го представления результатов исследований отдельных ученых в открытом доступе;
- 4) научное сообщество получает от интеграции метаданных институционных онлайн-репозиториев профессиональное *информационное пространство*²⁰ с общей системой навигации, благодаря которой пользователи имеют удобный доступ к аккумулярованному содержанию онлайн-репозиториев разных организаций, а также возможность децентрализованно развивать набор онлайн-средств для работы с содержимым информационного пространства, включая, например, средства автоматического контроля авторских прав (антиплагиат)²¹ и т.п.;
- 5) одно из новых качеств, которое возникает от объединения содержания репозиториев в общее информационное пространство, — ученые получают возможность создавать над множеством научных материа-

¹⁶ На момент написания статьи в онлайн-репозитории было выложено около 4 млн материалов. Российские репозитории содержали 3724 материала. См.: <http://roar.eprints.org/index.php?action=browse>

¹⁷ Перечень организаций, заявивших о принятии данной политики, см.: <http://www.eprints.org/openaccess/policysignup/>, в нем пока одна российская организация: ЦЭМИ РАН (май 2007 г.).

¹⁸ Обязывание ученых самоархивировать результаты своих исследований в форме статей не уменьшает их коммерческих возможностей, так как обычно ученые не получают денег за публикацию своих статей.

¹⁹ Например, в соответствии с принятыми в Австралии программы Research Quality Framework (RQF) и Higher Education Research Data Collection (HERDC) все университеты этой страны должны иметь онлайн-репозитории к концу 2007 г.

²⁰ Примерами информационных пространств являются системы ЕНИП РАН (<http://enip.gas.ru/>) и Соционет (<http://socionet.ru/>)

²¹ Контроль соблюдения авторских прав с технической точки зрения гораздо проще реализуем, когда все контролируемые тексты находятся в электронной форме в общем доступе. В Интернете уже существуют примеры подобных онлайн-сервисов (например, <http://www.antiplagiat.ru/>).

лов сети логических связей, например, между своим персональным профилем и своими статьями, попадающими в информационное пространство из разных независимых источников, или связи типа электронного цитирования, и т.п., подобные сети связей дают возможность визуализации сложившихся в научном сообществе *профессиональных социальных сетей*²²;

- 6) создание связей между объектами информационного пространства, таких как, с одной стороны, профиль организации или ученого, и с другой – материалы, которые имеют к ним отношение (статьи, коллекции, соавторы, коллеги, сотрудники, место работы и т.д.), в условиях обязательного самоархивирования учеными результатов своих исследований означает появление в информационном пространстве полноценных *профессиональных самопрезентаций*²³;
- 7) полноценное профессиональное представление ученых в виде их онлайн-образов придает смысл процедурам сбора и обработки *онлайн-статистики*²⁴, фиксирующей различные параметры их онлайн-активности, включая количества ежедневных просмотров абстрактов статей, скачиваний их полных текстов и т.п., наличие мотивации для 100%-го представления учеными результатов своих исследований в он-лайн определяет достаточно высокую содержательность и значимость подобной статистики для наукометрических исследований;
- 8) получение на основе наукометрической онлайн-статистики реалистичных *показателей влияния* результатов работы отдельных ученых, деятельности исследовательских организаций, отдельных статей и т.п. на профессиональное научное сообщество требует разработки *онлайн-метрики результатов научной деятельности*²⁵;
- 9) наличие онлайн-наукометрических показателей, рассчитываемых на основе «прозрачных» алгоритмов в реальном времени и с учетом сложившихся профессиональных социальных сетей, создает условия для модернизации системы *оценки профессиональной дея-*

²² Система Соционет предлагает пользователям такие возможности. См. <http://socionet.ru/socio-net.htm>

²³ См. примеры профессиональных самопрезентаций ученых // <http://socionet.ru/datatype.xml?h=person> и организаций // <http://socionet.ru/datatype.xml?h=institution>

²⁴ Описание практического примера сбора и обработки онлайн-статистики: <http://socionet.ru/rates.htm>

²⁵ Под онлайн-метрикой мы здесь понимаем правила вычисления индексов влияния результатов исследований для различных объектов профессионального информационного пространства, определения «расстояния» между ними и их упорядочивания по отношению друг к другу.

*тельности*²⁶ ученых и исследовательских организаций; подобные онлайн-показатели в первую очередь могут использоваться в исследовательских организациях в процессе аттестации ученых, определения академических надбавок и т.п.;

- 10) ученые и исследовательские подразделения/организации, с одной стороны, имеют возможности для 100%-го представления результатов своих исследований в профессиональном информационном пространстве и развития на этой основе своей профессиональной социальной сети, а с другой – наукометрические показатели, характеризующие в реальном времени степень своего влияния на профессиональное научное сообщество; наличие подобной *сигнальной системы*²⁷ позволяет ученым лучше корректировать свои действия для достижения максимально возможного профессионального эффекта, в том числе это создает дополнительные мотивации ученым для популяризации и *продвижения своих результатов* в целях их более частого цитирования в статьях других ученых;
- 11) организации, финансирующие исследования (научные фонды, исследовательские институты и др.), в свою очередь могут использовать онлайн-наукометрические показатели результатов исследований ученых и научных организаций для принятия *решений о финансировании* их исследований;
- 12) описанные выше факторы в совокупности создают условия для появления *новых поведенческих моделей* действующих лиц сферы научных исследований и разработок с меньшими транзакционными издержками и большей эффективностью.

Синтетическим результатом реализации подобной серии логически связанных разработок, организационных решений и мероприятий является значительная модернизация всей отрасли научных исследований и разработок. Ожидания, связанные с развитием социальных аспектов e-Science в результате второй волны внедрения Интернет-технологий, включают:

- 1) существенное улучшение публичной доступности результатов научных исследований;
- 2) более высокую оперативность в распространении и использовании новых научных знаний, в том числе направленное информирование ученых о появлении материалов, которые могут представлять для них интерес, автоматический контроль соблюдения авторских прав и т.п.;

²⁶ См. публикации по данной теме [7], [8].

²⁷ Мы понимаем термин «сигнальная система» как наличие внешних индикаторов, характеризующих отклик внешней среды на деятельность ученых в процессе принятия решений по поводу их профессиональной активности.

- 3) совершенствование системы моральной и материальной мотивации в науке;
- 4) повышение прозрачности сложившейся структуры профессиональных социальных сетей, состояния научных направлений, действующих специализаций в профессиональной деятельности и системы разделения труда в данной отрасли в целом;
- 5) повышение скоординированности научных исследований и др.

Сказанное выше свидетельствует об особой роли для успешной реализации ожиданий, связанных с e-Science, эффективной профессиональной сигнальной системы на основе комплекса наукометрических показателей. На наш взгляд, с учетом уже имеющихся заделов задачи построения эффективной сигнальной системы должны включать:

- разработку информационной модели профессиональных взаимодействий исследовательских организаций, отдельных ученых и других действующих лиц, адаптированной к новым онлайн-возможностям, с последующим проектированием и конструированием соответствующих онлайн-сервисов, в том числе позволяющих организациям и индивидам представлять в максимально полном и выгодном виде результаты своей научной деятельности в профессиональном онлайн-информационном пространстве;
- исследование и разработку онлайн-метрики для построения показателей использования и влияния результатов исследований, а также практическую реализацию механизмов подсчета и визуализации соответствующих показателей в реальном времени;
- модернизацию (на основе использования онлайн-показателей) организационных процедур оценки результатов научно-исследовательской деятельности ученых и организаций, применяемых на различных уровнях в различных управляющих и финансирующих науку организациях.

Общая научная проблематика построения и использования онлайн-показателей для анализа различных объектов и видов социально-экономической деятельности имеет более чем 10-летнюю историю²⁸. Разработка на этой основе методологии создания онлайн-профессиональной *сигнальной системы* фактически только начинается.

Изложенное ниже может рассматриваться как результат проведенных подготовительных работ к решению перечисленных выше трех задач мо-

²⁸ Алексеев А.В., Киселев А.В., Паринов С.И. Анализ деловой активности в регионах России в 1994 г. // Вопросы экономики. 1995. № 3; Паринов С.И. Трафик сети Интернет как источник экономических данных // РАРДЭС. 1999 Май (<http://rvles.ieie.nsc.ru/parinov/net-data/intro.htm>); Паринов С.И. К теории сетевой экономики. Новосибирск, ИЭОПП СО РАН, 2002 (<http://socionet.ru/publication.xml?h=repcc:rus:ieieli:97421>).

дернизации российского сектора научных исследований и разработок и является, по всей видимости, первым опытом построения и практического использования онлайн-наукометрических показателей.

2. Институты ООН РАН как полигон для построения e-Science

Объектом наших усилий по развитию научных информационных систем (CRIS), связыванию их в единое целое на основе онлайн-инфраструктуры (e-infrastructure) и формированию на этой основе e-Science стали 29 институтов Отделения общественных наук РАН (ООН РАН). Данные работы выполняются межинститутской группой Соционет, в которую входят специально назначенные представители всех институтов ООН РАН в рамках реализации программы «Открытый доступ к результатам исследований» (ОДРИ), инициированной в 2006 г. в ООН РАН [11].

Вторая волна интернетизации оказывает достаточно сильное влияние на институты ООН РАН благодаря реализации ряда проектов, осуществляемых, в том числе, межинститутской рабочей группой Соционет на гранты Фонда Форда и других организаций.

Институты ООН РАН активно участвуют в одном из проектов программы «Информатизация» Президиума РАН, цели которого характеризуются следующим образом:

«...инициатива по организации Единого научного информационного пространства РАН (ЕНИП РАН) призвана помочь научным коллективам сделать ряд шагов в направлении интеграции разнородных научных информационных и программных ресурсов отдельных научных учреждений, предоставления пользователям более эффективных средств интеграции и поиска информации, научной коммуникации, сотрудничества и совместной работы. Под единым пространством понимается не формирование централизованной системы, не навязывание всем одних и тех же решений, а стремление последовательностью практических шагов, совместными усилиями научных коллективов РАН:

- сформулировать взаимосогласованный набор соглашений, правил и открытых стандартов;
- приготовить совокупность макетов и типовых решений для реализации адаптеров прикладных систем, инфраструктурных служб, поддерживающих разные уровни интероперабельности распределенных гетерогенных данных и приложений;
- создать ряд информационных систем общего назначения, следующих этим соглашениям, использующих эти реализации, допускаю-

щих модульную организацию, наращивание функциональных возможностей;

- применить эти результаты для решения соответствующих задач учреждений РАН». [12]

По состоянию на начало 2007 г. в рамках проекта ЕНИП РАН получены значительные результаты по созданию базовой онлайн-инфраструктуры, объединяющей в единую информационную сеть многодисциплинарные ресурсы научных организаций Российской академии наук²⁹. Отдельными узлами информационной сети ЕНИП РАН являются достаточно крупные информационные ресурсы, представленные различными организациями от отдельных научных институтов до отделений Академии. Одним из узлов этой сети является упомянутая ранее система Соционет.

Основным проектом, на котором в данное время основывается реализация программы ОДРИ, является «Соционет». Проект продвигается инициативной группой, ядро которой до 2005 г. составляли сотрудники Института экономики и ОПП СО РАН, а в настоящее время сотрудники ЦЭМИ РАН. Проект с 2000 г. финансировался Фондом Форда. Исходно проект ориентировался на исследователей в областях общественных наук. В последние годы наиболее активными пользователями Соционет стали институты Отделения общественных наук РАН, для которых система Соционет, в том числе, играет роль шлюза для передачи данных в ЕНИП РАН.

К настоящему времени старый образ системы Соционет³⁰ — «информационное пространство + личная зона» — эволюционирует в новый: социальная сеть участников профессиональной деятельности в областях науки и образования. Одно из следствий этого — система Соционет вышла за пределы общественных наук. В ней уже появилась дисциплина «Информатика», а также сформировалось параллельное информационное пространство участников инновационной деятельности — СИНИН (<http://sinin.nsc.ru/>).

Представители институтов ООН РАН создали в Соционет свои личные зоны³¹, которые позволяют им реализовывать достаточно большой набор сценариев по созданию коллекций научных материалов и выкладывания их в открытый доступ. В этом качестве функциональность личных зон играет роль институтских CRIS, так как позволяет автоматизировать определенные виды научно-исследовательской деятельности. Кроме личных зон, Соционет в другой своей сущности выполняет роль информационно-

го хаба [10], интегрируя разнородную информацию, создаваемую децентрализованно представителями институтов (как в личных зонах Соционет, так и на своих сайтах), в единую базу данных с функциями сквозного поиска, глобальной навигации и т.п. в рамках профессионального информационного пространства. Тем самым фактически Соционет выполняет роль научной онлайн-инфраструктуры (research e-infrastructure), так как предлагает институтам стандартизованные сервисы по автоматизации своих частных информационных задач (позволяя экономить средства на создание собственных сервисов), при этом превращая отдельные информационные продукты институтов в единую взаимосвязанную систему.

По состоянию на февраль 2007 г. институты ООН РАН создали свои онлайн-репозитории и таким образом выложили в открытый доступ в рамках профессионального информационного пространства Соционет (<http://socionet.ru/>) около 30 тыс. материалов, включая 41 профиль³² организаций/подразделений, 466 профилей ученых. Практически все институты ведут ленты новостей, коллекции статей своих сотрудников, описания значимых опубликованных материалов (книги и монографии), авторефераты защищаемых диссертаций и т.п.

Между выложенными в открытый доступ в Соционет профилями организаций, ученых и описаниями статей/материалов сотрудники институтов, входящие в рабочую группу Соционет, создают сети связей, которые отражают различные виды взаимозависимостей между ними. В результате развития сети связей происходит следующее:

- Электронные версии статей и материалов имеют связи с профилем автора и профилем организации — местом работы автора. Статья имеет связи с материалами, перечисленными в разделе «Литература» (если они присутствуют в информационном пространстве), а также обратные связи от статей, в которых данная статья цитируется.
- Персональные профили авторов имеют связи с описаниями статей и других материалов, выложенными их авторами в Соционет, с коллекциями, которые ими созданы, а также с профилями организаций, с которыми себя данные ученые ассоциируют.
- Профили организаций (институтов) имеют связи с коллекциями статей и материалов, которые в них созданы, с персональными профилями ученых, которые в них работают, а также с профилями своих подразделений.

²⁹ См. текущее состояние информационной сети на <http://enip.ras.ru/>

³⁰ <http://socionet.ru/>

³¹ Личная зона Соционет // <http://spz.socionet.ru/index-ru.shtml>

³² В данном контексте «профиль» означает стандартизованное описание организации или ученого, создаваемое ответственными от институтов или самими учеными путем заполнения карточек в личной зоне Соционет (<http://spz.socionet.ru/index-ru.shtml>).

Если продолжить в будущее данные тенденции [8], то:

- с определенного времени метаданные всех статей и других материалов с результатами исследований будут частью определенных открытых онлайн-репозиторий;
- все материалы этих репозиторий будут вписаны в различные тематические и профессиональные рубрикаторы и онтологии;
- за счет этого каждый материал в отдельности будет иметь прямые и обратные связи разного типа с другими материалами, и все это будет образовывать профессиональную социальную сеть;
- в частности, каждое цитирование в исходном материале будет позволять по ссылке прямо перейти к использованному материалу и обратно — если исходный материал процитирован, то при его просмотре будут доступны ссылки на цитирующие его материалы;
- все вместе это создаст дополнительную систему навигации по профессиональному информационному пространству типа «социальная сеть», позволит отслеживать в реальном времени индекс онлайн-цитирования и количество просмотров аннотации/полного текста каждого материала;
- создаст основу для подсчета и агрегирования данных о востребованности материалов, авторов и организаций в рамках данной профессиональной социальной сети.

Реализуя шаг за шагом программу открытого доступа, рабочая группа Соционет использует рекомендации «Берлин-3» [2] для исследовательских организаций:

- 1) каждая исследовательская организация, входящая в ООН РАН, должна иметь свой официальный онлайн-репозиторий, включенный в международные реестры и каталоги открытого доступа (Eprints, OAIster и др.);
- 2) организации должны требовать от своих научных сотрудников самоархивировать (выкладывать) в институтские онлайн-репозитории с открытым доступом электронные копии всех своих опубликованных статей;
- 3) необходимо стимулировать научных сотрудников публиковать статьи в журналах с открытым доступом к материалам при условии существования таковых (и обеспечивать сотрудникам поддержку в данных действиях).

В западной научной литературе уже опубликованы [3] статистические исследования последствий реализации рекомендаций «Берлин-3». Результаты показывают, что количество цитирований материалов, выложенных в репозитории открытого доступа, увеличивается в 2 раза по сравнению с обычной ситуацией. Кроме того, реализация самоархивирования и дру-

гих мероприятий, предлагаемых Берлинской декларацией, эквивалентна дополнительным инвестициям в науку. Например, для Великобритании это соответствует дополнительным 1,5 млрд ф. ст. [3].

Выполнение этих рекомендаций институтами ООН РАН по состоянию на май 2007 г. имело следующие результаты:

- 22 института (из 29, входящих в ООН РАН) создали онлайн-репозитории, которые представлены в Соционет (<http://socionet.ru/>) и ЕНИП РАН (<http://enip.ras.ru/>), а также практически готовы для регистрации в международных реестрах репозиторий открытого доступа;
- онлайн-репозиторий пока только одного института — ЦЭМИ РАН — имеет международный протокол доступа, что позволило зарегистрировать его в международном реестре репозиторий открытого доступа (п. 1 рекомендаций), с учетом этого Россия занимает пока 27-е место в списке стран по количеству зарегистрированных репозиторий (ожидается, что только после регистрации остальных институтов ООН РАН, Россия может переместиться на 12-е место);
- пока только научные сотрудники ЦЭМИ РАН обязываются приказом директора этого института размещать в онлайн-репозитории все законченные результаты исследований (п. 2 рекомендаций);
- примеры реализации п. 3 рекомендаций «Берлин-3» пока нам не известны.

ЦЭМИ РАН первым из институтов ООН РАН в апреле 2007 г. полностью присоединился к международной инициативе «Открытый доступ к результатам исследований». В институте создан онлайн-Открытый архив³³ (по содержанию — полный аналог упомянутого выше онлайн-репозитория). Приказом директора³⁴ научные сотрудники института обязываются депонировать в нем электронные версии всех своих законченных результатов исследований. Для создания надлежащей мотивации к электронному депонированию результатов исследований в приказе содержится положение о подготовке к использованию онлайн-статистики востребованности электронных материалов, депонированных исследователями в ОА ЦЭМИ РАН, в процедуре установления персональных надбавок научным сотрудникам института.

Общий статус Открытого архива (ОА) ЦЭМИ РАН определен следующим образом.

Институтский ОА объединяет различные тематические коллекции электронных материалов, созданные и поддерживаемые при участии ЦЭМИ

³³ <http://cemi.socionet.ru/>

³⁴ <http://www.cemi.rssi.ru/rus/news/initiat-rus.htm>

РАН, а также отдельные материалы с результатами исследований подразделений и сотрудников института.

ОА ЦЭМИ РАН имеет, в том числе, интерфейс OAI-PMH³⁵, рекомендованный положением «Берлин-3» для научных электронных ресурсов, открытых для свободной интеграции в международные ОА-реестры и информационные хабы. В этом качестве ОА ЦЭМИ РАН зарегистрирован в реестре ROAR³⁶, а также в некоторых других российских и международных системах.

В международном реестре ROARMAP³⁷ также зарегистрирована политика ЦЭМИ РАН по открытому доступу, введенная в действие приказом директора, которая уточняет организационный статус институтского ОА:

а) содержание ОА представляет в открытый онлайн-доступ *все* законченные результаты исследований, выполненных научными сотрудниками института;

б) все имеющиеся в ОА расширенные библиографические описания со ссылками на полные тексты статей (метаданные) свободно открыты для использования в научных тематических подборках, системах поиска и других интегрированных информационных ресурсах.

В результате для сотрудников ЦЭМИ РАН действуют следующие правила электронного депонирования материалов в институтском ОА:

Все законченные исследования, включая результаты по грантам научных фондов, должны быть опубликованы в рецензируемых научных изданиях или, если это не сделано в течение 6 месяцев с момента их готовности, материалы должны быть депонированы в институтском ОА (в виде электронного препринта, включая его полный текст).

Если материал принят рецензируемым научным изданием для публикации, то авторская версия материала должна быть депонирована в институтском ОА (в виде электронного постпринта, включая его полный текст).

Более 90% научных журналов³⁸ в мире официально объявили о согласии с электронным депонированием авторских версий материалов в форме препринтов и постпринтов. Если научное издание все-таки требует от автора эксклюзивных прав на материал с результатами исследований или требует не выставлять материал в открытый доступ определенное время после его публикации в издании, то автору ре-

³⁵ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

³⁶ См. соответствующую строчку в российском разделе реестра: <http://roar.eprints.org/index.php?action=home&country=ru>

³⁷ <http://www.eprints.org/openaccess/policysignup/>

³⁸ <http://romeo.eprints.org/stats.php>, см. список издательств: <http://romeo.eprints.org/publishers.html>

комендуется³⁹ депонировать в институтском ОА расширенное библиографическое описание материала (метаданные), в которое следует включить (например, в поле «Статус материала») уведомление о возможности получения полного текста материала по личному запросу на адрес электронной почты автора.

Материалы с результатами исследований прошлых лет также должны быть представлены в институтском ОА. В идеале научные сотрудники ЦЭМИ должны иметь в открытом доступе 100% результатов своих исследований. Это позволит полнее учесть в индексе востребованности автора результаты его работы.

Краткая инструкция по электронному депонированию материалов в ОА ЦЭМИ РАН

Во всех подразделениях ЦЭМИ должны быть назначены ответственные за электронное депонирование в институтском ОА, которые прошли обучение, администрируют коллекции материалов от своих подразделений и могут проконсультировать сотрудников своего подразделения по самостоятельному электронному депонированию материалов.

Ответственные от подразделений создают в своих личных зонах Соционет необходимое количество тематических коллекций, в которые выкладываются (депонировются) материалы с результатами исследований и др. Администратор ОА ЦЭМИ РАН подключает созданные коллекции к профилю института по мере их готовности. Установление связи между коллекцией и профилем ЦЭМИ РАН⁴⁰ в Соционет означает включение данной коллекции в институтский ОА.

Сотрудники ЦЭМИ РАН могут также выполнять электронное депонирование своих материалов самостоятельно. Эта операция занимает 2–3 минуты (на один материал), если уже готовы библиографическое описание материала и файл с его полным текстом. При первом электронном депонировании необходимо создать свою личную зону Соционет. В уже созданной личной зоне можно размещать материалы в а) коллекциях Открытого архива Соционет⁴¹ или б) создавать свои собственные коллекции⁴². Чтобы материалы, депонированные сотрудниками самостоятельно, попали в институтский ОА, необходимо установление связи между коллекцией ученого и профилем ЦЭМИ РАН. При этом сотрудник должен размещать материалы в собственной коллекции (пункт «б» выше), либо ответственный от соответствующего

³⁹ http://www.eprints.org/news/features/request_button.php

⁴⁰ http://socionet.ru/publication.xml?h=repec:rus:ecoorg:semi-ras_admin.45009-org1&type=institution

⁴¹ <http://spz.socionet.ru/add-docs.htm>

⁴² http://socionet.ru/docs/instr1/instruction_04.htm

подразделения должен скопировать метаданные материала сотрудника в коллекцию подразделения, которая уже соединена с профилем ЦЭМИ.

Для полного учета индекса востребованности результатов исследований сотрудника ЦЭМИ РАН все его материалы, находящиеся в открытом доступе в Соционет, должны быть связаны с персональным профилем данного ученого в Соционет.

Подобные организационные меры предполагается реализовать во всех институтах ООИ РАН. Новое качество, которое получают институты ООИ РАН от участия в этой работе, состоит в следующем:

а) развитие функциональности институтских CRIS и расширение таким образом автоматизированных областей научной деятельности за счет использования в сочетании со своими веб-ресурсами еще и возможностей системы Соционет, выполняющей связующую роль исследовательской онлайн-инфраструктуры (e-infrastructure);

б) повышение доступности для потенциальных потребителей результатов исследований института за счет обязывания научных сотрудников размещать в открытом доступе все законченные исследования;

в) формирование профессиональных социальных сетей над межинститутским междисциплинарным информационном пространством статей и материалов, что демонстрирует комплексность результатов его работы и степень вовлеченности института в деятельность научного сообщества;

г) участие в международной системе распространения и использования результатов исследований и научной информации на основе стандартов OAI-PMH за счет включения институтских Открытых архивов в реестры репозиториях открытого доступа и информационные хабы.

Накопленный в настоящее время рабочей группой Соционет опыт показывает, что ключевым фактором успешности данных технических и организационных инноваций, а также формирования e-Science в целом, является (при наличии необходимых технических условий) еще и действенная работа мотивационного механизма для всех участников этих процессов. Приказ директора ЦЭМИ РАН о необходимости 100%-го размещения в открытом доступе результатов исследований является первой необходимой частью механизма мотивации научных сотрудников. Другой необходимой частью является оценка результатов деятельности научных сотрудников и моральное/материальное стимулирование их труда с использованием данных о выполнении данного приказа, а также внешних индикаторов результативности исследователей, отражающих отклик внешней среды на онлайн-активность ученых.

Именно последняя компонента механизма мотивации, названная нами сигнальной системой, является на данный момент наименее понятной и

проработанной. Создание эффективной сигнальной системы основывается на онлайн-наукометрических показателях, которые, в свою очередь, строятся на основе онлайн-статистики использования информационных объектов, размещенных учеными в открытом доступе в институ-ских ОА, онлайн-репозиториях и других веб-ресурсах.

3. Онлайн-наукометрические показатели

В западной научной литературе отмечается, что вызванный международными инициативами открытого доступа к результатам исследований рост количества публикаций в открытом доступе в институ-ских онлайн-репозиториях создает в настоящее время уникальный исторический момент для разработки и использования онлайн-метрики в оценке научной деятельности (Research Metrics). Использование онлайн-показателей в процессе принятия решений, например, при профессиональной переквалификации (научной аттестации) или при выделении грантов и т.п., создает ряд важных для научной деятельности мотиваций и новых возможностей [8]:

- 1) мотивировать большее количество исследователей открыть доступ к своим статьям путем их самоархивирования (размещения в онлайн-репозиториях);
- 2) отслеживать рост открытого доступа по дисциплинам, странам и языкам;
- 3) осуществлять навигацию по статьям в открытом доступе, используя связи цитирования и их ранжирование по индексу влияния;
- 4) измерять научную деятельность и производительность;
- 5) выбирать кандидатов для научных грантов;
- 6) выбирать результаты исследований для научных грантов;
- 7) отслеживать основные научные направления в переложении на отдельных ученых, институты, журналы, области науки, нации и т.п.;
- 8) анализировать и предсказывать текущие и будущие научные направления и тенденции;
- 9) обеспечивать студентов учебными и обучающими ресурсами (через навигацию по индексу влияния) в более эффективном виде, чем это делают Google и другие поисковые системы общего назначения.

В качестве первого шага к построению онлайн-метрики для оценки результатов исследований ниже рассматривается возможная статисти-

ческая база и примеры онлайн-показателей, которые могут быть построены на основе этой базы.

3.1. Статистическая база для онлайн-показателей

Все запросы к статьям и материалам в онлайн-информационных системах стандартным образом фиксируются в лог-файлах (электронных журналах) соответствующих Интернет-серверов. Система Соционет собирает и обрабатывает данные из стандартных лог-файлов, а также сведения из других разделов Соционет, с целью построения максимально полной статистической базы для наукометрических исследований.

Онлайновая статистика в полном виде собирается в системе Соционет с 01.01.2007 г.⁴³ Это означает, что в начале каждого календарного суток: а) отфильтровываются данные о загрузках/просмотрах из лог-файла сервера socionet.ru, а также б) из базы данных Соционет собираются другие необходимые сведения. Эти данные записываются в файлы и хранятся на ftp://socionet.ru/raw-stats/ в простом текстовом формате. Для каждой календарной даты могут существовать файлы с данными нескольких разных типов (если на заданную дату какие-то файлы отсутствуют, это означает, что соответствующие данные в этот день отсутствовали). Имена файлов с данными статистики разных типов различаются первой буквой, после которой в имени стоит календарная дата, для которой файлы построены. Файлы могут начинаться с: «s» — статистика просмотров/загрузок; «r» — данные о размере коллекций и количестве публикаций у авторов; «x» — сведения о связях между материалами, которые необходимы для агрегирования статистики просмотров/загрузок. Ниже приводится описание формата данных для каждого типа.

Файлы исходной статистики Соционет свободно доступны на ftp://socionet.ru/raw-stats/ всем желающим для контроля правильности расчета (независимого воспроизведения) показателей, приводимых в Соционет, а также для использования в наукометрических исследованиях и конструирования собственных показателей.

⁴³ Это относится к данным о просмотрах/загрузках (тип файлов S). Ранее (начиная с 2001 г.) аналогичные данные собирались только как ежемесячная статистика по методике LogEc (<http://logec.repec.org/about.htm>). Данные о размерах коллекций (тип файлов R) начали сохраняться в ftp://socionet.ru/raw-stats/ с 07.03.2007 г. До этого текущие размеры коллекций брались из базы данных Соционет. Данные о связях между объектами (тип файлов X) начали сохраняться в ftp://socionet.ru/raw-stats/ с 2007.04.21. До этого все данные о связях также брались из базы данных Соционет.

Тип данных S. В файлах, имя которых начинается с буквы «s», находится статистика просмотров и загрузок статей/материалов в базе данных Соционет. Подобные сведения получены путем фильтрации лог-файла сервера socionet.ru и содержат статистику запросов к статьям и материалам в Соционет (за основу взята методика системы LogEc⁴⁴).

Файлы этого типа имеют следующую структуру:

1) дата	2) сервер	3) код материала	4) вид	5) исходящий адрес
2007-02-18	RuPEc	repec:aal:abbspw:99-1	a:	62.158.21.139
2007-02-18	RuPEc	repec:anp:en2005:047	a:	66.36.241.193
2007-02-18	RuPEc	repec:apr:aprpdp:dp0002	d:	72.36.176.130 203.94.243.2
2007-02-18	RuPEc	repec:bar:bedcje:2006146	a:	66.36.229.227
2007-02-18	RuPEc	repec:bca:bocawp:05-18	a:	74.52.83.122
2007-02-18	RuPEc	repec:bep:mchbio:1025	d:	85.140.199.95

- где: 1) дата (date) — год-месяц-день обращения к материалу;
2) сервер (base) — имя сервера (RuPEc означает код сервера socionet.ru), на котором получено и обработано обращение к статье/материалу, в общем случае данные о запросах статей/материалов могут собираться с разных серверов;
3) код материала (handle) — уникальный код статьи или материала, который является единым для данного информационного объекта в информационных системах на разных серверах (base), экспонирующих статьи/материалы с информационного хаба Соционет [10];
4) вид (request) — определяет вид запроса и может быть только «a:» или «d:»:
 «a:» — запрошены только метаданные (карточка, описатель статьи/материала);
 «d:» — запрошен полный текст статьи/материала, т.е. при просмотре метаданных пользователь воспользовался ссылкой «полный текст»;
5) исходящий адрес (ip) — уникальный идентификатор компьютера в сети Интернет (IP-адрес), с которого поступил запрос на данный материал.

При обработке суточной порции лог-файлов в них фиксируются запросы от сетевых «роботов» и «индексаторов» (их IP-адреса приводятся в начале каждого файла с суточной порцией исходных данных) и отфильтровываются «дубли». Считается, что IP-адрес принадлежит сетевому «роботу».

⁴⁴ <http://logec.repec.org/>

ту» или «индексатору», если от него исходит серия запросов через каждые 10–30 секунд. Устранение дублей означает, что в отфильтрованный файл попадает только по одному запросу из имеющих одинаковый набор всех описанных выше пяти параметров.

Тип данных R. В файлах, имя которых начинается с буквы «r», находятся сведения о размерах коллекций (количество в них статей/материалов) и количестве статей/материалов, указанных в связях от персональных профилей авторов на определенную дату. В файл этого типа заносятся только те коллекции/авторы, у которых на текущую дату зафиксированы изменения в количестве принадлежащих им статей/материалов.

Файлы этого типа имеют следующую структуру:

1) дата	2) кол-во	3) код объекта	4) вид
2007-03-15	165	repec:vlg:vlgwps	coll
2007-03-15	255	repec:wej:wldecn	coll
2007-03-15	6	repec:rus:socper:zagorulko_elena.62537-17	auth
2007-03-15	1	repec:rus:socper:zagorulko_elena.62537-47	auth
2007-03-15	41	repec:rus:doahtw:lbn-1	auth
2007-03-15	2	repec:rus:doahtw:lbn-10	auth

где: 1) дата в формате год-месяц-день;

2) количество статей/материалов на дату из столбца 1 для объекта информационного пространства Соционет, соответствующего коду из столбца 3, значение «–1» в этом столбце означает отсутствие на текущую дату коллекции/автора, код которого указан в столбце 3, при условии, что в предыдущий день данный объект существовал;

3) код объекта, для которого зафиксировано количество принадлежащих ему статей/материалов на дату из столбца 1;

4) вид объекта может быть только «coll» – коллекция, или «auth» – персональный профиль:

- если «coll», то в столбце 2 указано количество статей/материалов коллекции с кодом из столбца 3;

- если «auth», то в столбце 2 указано количество статей/материалов, связанных с персональным профилем с кодом из столбца 3.

Тип данных X. В файлах, имя которых начинается с латинской буквы «x», находятся данные о связях материала с другими статьями/материалами в базе данных Соционет. В файл данного типа заносятся только те материалы, у которых на текущую дату зафиксированы изменения в количестве и содержании связей. Об использовании этих данных для построения показателей см. в следующих разделах.

Файлы этого типа имеют следующую структуру:

1) дата	2) код объекта	3–4) тип связи	5) коды связанных объектов
2007-05-20	repec:rus:ecoper:parinov_sergey.56054-1	publication 1	repec:rus:rardes:1762001 ...
2007-05-08	repec:rus:ecoorg:semi-ras_admin.45009-org	collection 6	book_ekonomika_rus_cemnew ...

Каждая строчка в файлах этого типа относится к одному объекту и имеет 5 колонок с разделителем "|" (вертикальная черта);

Содержание данных в колонках:

1) дата в формате год-месяц-день, на которую зафиксированы изменения в связях;

2) код объекта (статьи/материала), связи которого изменились на текущую дату;

3) название поля в метаданных описания объекта, в котором изменились связи (если у объекта существуют другие поля со связями, в которых не было изменений, то в файл они не записываются);

4) ID-поля, в котором изменились связи (значение поля id относительно таблицы datatypes базы данных Соционет);

5) перечень кодов связанных объектов, с которыми установлены связи исходного объекта, разделителем является пробел (в примерах выше в этой колонке приведено только по одной связи), если в этой колонке стоит слово «DEL», то это означает отсутствие связей на текущую дату при условии, что в предыдущий день они были.

Суточные порции статистики из описанных выше файлов сразу после их создания загружаются в общую базу данных, которая затем используется для расчета и хранения требуемого множества показателей для заданных периодов времени на накопленных временных рядах.

3.2. Метрика онлайн-научно-метрических показателей

Для использования описанной выше статистической базы необходимо определить правила вычисления онлайн-научно-метрических показателей, которые будут содержательно характеризовать различные аспекты онлайн-научно-исследовательской деятельности. Например, представляют интерес индексы влияния результатов исследований, а также определение на этой основе меры удаления друг от друга объектов из одинаковых классов (например, внутри трех классов «статьи», «ученые» и «организации»), а также правила их упорядочивания по различным критериям.

Пространство данных, создаваемое собираемой статистикой, имеет пять основных осей координат:

- временная ось (переменная date с суточной шкалой изменений, начальная точка – 01.01.2007 г.);
- ось объектов профессионального информационного пространства (статьи, материалы, профиль автора и организации и т.д.; перечень типов объектов см. на <http://swb.socionet.ru/datatype.htm>), переменная handle;
- ось пользователей, представленных адресами своих компьютеров в Интернете (профессиональное сообщество и случайные посетители), переменная ip;
- ось видов загрузки материалов (просмотр карточки и загрузка полного текста), переменная request;
- ось имен Интернет-серверов, с которых на компьютеры пользователей загружается требуемый материал (материал с одним и тем же handle может просматриваться пользователями на разных Интернет-серверах, система сбора статистики позволяет объединять такую статистику в единое целое), переменная base.

Точка в этом пространстве означает следующее событие: на дату date на сервере base для объекта с кодом handle поступил запрос вида request от пользователя ip. Этот запрос означает проявление интереса к соответствующему объекту handle. Или иначе: объект handle востребован, если к нему проявил интерес пользователь ip. Правда, не известен статус данного интереса, имеет ли он отношение к использованию содержания соответствующей статьи/материала или был случайным. Статус интереса можно считать серьезным, если пользователем ip было произведено скачивание полного текста объекта handle, т.е. вид запроса – request = d. Поэтому при расчете ряда наукометрических показателей имеет смысл выделять отдельно количество загрузок соответствующих полных текстов.

Построение показателей на этой статистике в общем случае означает задание правил выборки, агрегирования и преобразования данных по каждой из пяти осей координат. В ряде случаев для агрегирования данных используется информация о связях между родственными материалами, которые формируются ответственными от институтов ООН РАН в рамках создания своих профессиональных социальных сетей. Чем больше, например, у заданного персонального профиля автора связей с его статьями, тем больше показателей востребованности статей суммируется для вычисления аналогичного показателя для данного автора. При добавлении в профиль автора связи с новой статьей его показатель востребованности немедленно увеличивается (если показатель востребованности данной статьи не нулевой). Следовательно, динамика такого показателя востребованности зависит от истории изменения связей соответствующего объекта.

Текущие данные о всех установленных на данный момент связях исходно находятся в базе данных Соционет. В целях восстановления из исходных статистических данных идентичной динамики показателей востребованности потребовалось сохранять историю изменений связей между объектами (см. выше описание файлов статистики типа X).

Некоторые варианты показателей востребованности включают различные нормировки. Например, показатель востребованности организации может нормироваться на количество сотрудников, которые в нем работают, или показатель востребованности автора нормируется на количество статей, которое указано в связях в его персональном профиле. В целях восстановления из исходных статистических данных идентичной динамики для нормированных показателей потребовалось также сохранять историю изменений размеров различных объектов информационного пространства Соционет (см. выше описание файлов статистики типа R).

Наиболее очевидными показателями, которые могут быть сравнительно легко построены на основе онлайн-статистики, являются:

показатели «востребованности» статей и материалов, рассчитываемые как количества уникальных запросов на просмотр/скачивание статей и материалов в Соционет.

На основе этих показателей можно рассчитать вторичные показатели востребованности:

показатели «востребованности» для авторов и организаций, построенные суммированием показателей востребованности статей/материалов на основе связей, установленных для заданного профиля.

Например, показатель востребованности ученого считается как сумма показателей востребованности для статей и материалов, автором которых он является и связи с которыми установлены в его персональном профиле. Показатель востребованности организации подсчитывается аналогично, но с учетом более широкого набора связей, устанавливаемых в профиле организации. Для целей сравнения показатель востребованности организации может нормироваться по числу ее сотрудников, чьи персональные профили указаны в связях профиля организации.

При использовании динамических рядов размеров коллекций, которые также сохраняются в статистической базе Соционет, может быть рассчитан еще один тип показателей. Данные, отражающие изменение количества материалов, выложенных в разное время администраторами коллекций в открытый доступ, позволяют построить:

показатели активности действующих лиц (в режиме редактирования), рассчитываемые как частота производимых ими обновлений и(или) интенсивность выкладывания (количество статей за заданный период

времени) новых материалов в информационное пространство Соционет, где действующие лица — ученые и исследовательские организации (в случае организации активность включает сумму активностей ученых, принадлежащих к данной организации).

Показатели активности дают своего рода оценку «затрат» действующих лиц, а показатели востребованности — оценку «результатов» научно-исследовательской деятельности ученых и исследовательских организаций, от которых зависит как появление новых материалов, так и просмотр всех существующих. Однако для научных исследований существует еще более точный показатель их результативности, который тоже может быть построен на основе онлайн-статистики.

При использовании информации о связях между материалами типа «цитирование» возможен расчет еще одного типа показателей:

влияния статей и материалов, а также действующих лиц как степени полезности статьи (или нескольких статей одного автора) для других ученых в генерировании ими новых результатов исследований и разработок, включая учет того, как часто данная работа была прочитана, использована, процитирована в других исследованиях, а также в обзорных, технологических, культурных, социальных и практических приложениях.

Показатель первого типа («востребованность» статьи) прямо входит в расчет индексов влияния результатов исследований, поскольку в наших условиях исходным источником влияния является статья. В общем случае в состав показателя влияния статьи входят следующие компоненты:

- количество просмотров аннотации и скачиваний полного текста статьи (востребованность);
- количество цитирований статьи (исключая самоцитирование);
- индекс влияния цитирующих авторов;
- рейтинг изданий (научных журналов), принявших статью для публикации.

Далее описаны специфические детали построения соответствующих показателей.

3.3. Показатели востребованности

Показатели востребованности подсчитываются специфическим образом (разный способ учета связей между материалами) для следующих типов объектов информационного пространства Соционет:

- 1) статья/материал;
- 2) коллекция статей/материалов;
- 3) архив коллекций организации;

- 4) профиль ученого;
- 5) профиль организации/подразделения;
- 6) сервер, на котором пользователи просматривают и пополняют профессиональное информационное пространство.

Общим в алгоритмах расчета показателя востребованности (*uz*) для перечисленных выше объектов является:

uz определяет количество уникальных запросов для заданного кода материала (*handle*) за заданный временной период (*date* — день, неделя, месяц и год) для указанного вида запроса (*request*).

Для разных типов объектов этот показатель представляет собой функцию *uz(handle, date, request)* и рассчитывается для конкретных случаев следующим образом:

1) *uz_item* — показатель *uz* статьи/материала определяется прямым подсчетом уникальных запросов и к этому может добавляться с определенным коэффициентом сумма *uz_item* для статей/материалов, цитирующих исходную статью/материал, исключая случаи самоцитирования;

2) *uz_series* — показатель *uz* для коллекции определяется как сумма *uz_item* статей/материалов, составляющих данную коллекцию, плюс *uz_item(handle=series_id)* описателя самой коллекции; нормирование на размер коллекции не делается, так как в данном случае подсчитывается абсолютный размер «привлеченного внимания» к материалам коллекции или их абсолютная востребованность;

3) *uz_archive* — показатель *uz* для архива подсчитывается как сумма *uz_series* входящих в него коллекций плюс *uz_item(handle=archive_id)* для описателя архива;

4) *uz_person* — показатель *uz* профиля ученого включает кроме суммы уникальных просмотров самого профиля *uz_item(handle=person_id)* еще суммы⁴⁵ *uz_item* статей/материалов, указанных в связях от данного профиля (статьи/материалы данного автора и созданные им коллекции), а также результат объединения *uz_item* из персональных профилей, которые указаны в исходном профиле ученого как устаревшие (замещаемые профили);

5) *uz_institution* — показатель *uz* профиля организации/подразделения включает кроме суммы уникальных просмотров самого профиля *uz_item(handle=institution_id)* еще и суммы *uz* по его связям с профилями сотрудников, с коллекциями или архивами, подразделения и мате-

⁴⁵ В данном случае суммы просмотров статей/материалов (*uz_item*) могут быть умножены на определенные коэффициенты (веса), отражающие статус (например, импакт-фактор) коллекции, которым они принадлежат.

риалами⁴⁶, при подсчете исключаются все повторы в статьях и других объектах, а также этот показатель нормируется на число сотрудников, указанных в связях;

б) *uz_base* – показатель *uz* сервера в целом подсчитывается как общее количество уникальных запросов за заданный промежуток времени, принятых соответствующим сервером.

Динамические ряды показателей *uz* для всех шести описанных выше разновидностей объектов заранее рассчитываются и хранятся в статистической базе данных.

Ниже представлены реальные примеры динамических рядов показателей, построенных на основе исходных данных, хранящихся на <ftp://socio-net.ru/raw-stats/>, для временного периода 06.04.2007– 5.06.2007 гг., по оси Y – величины соответствующих показателей. Нижняя зона столбцов – количество просмотров, верхняя – количество загрузок.

Рис. 1 демонстрирует общее количество уникальных обращений за последние 60 дней (период 06.04.2007–05.06.2007 гг.) к статьям и материалам информационного пространства Соционет. В среднем за данный период зафиксировано около 400 уникальных запросов в сутки. Из этих данных удалены повторы и запросы от сетевых «роботов» и «индексаторов».

Обычно запросы от сетевых «роботов» и «индексаторов» составляют большую часть, примерно 80% в общем потоке запросов. В результате этих запросов обновляются сведения об информационных ресурсах Соционет в глобальных поисковых системах типа Google, Яндекс и др., через которые пользователи обычно ищут интересующие их материалы. Если учесть и эти запросы, то среднее число уникальных запросов к Соционет в целом будет составлять примерно 2–3 тыс. в сутки.

На рис. 2, 3 показано, что ежедневный просмотр материалов, принадлежащих сотрудникам ЦЭМИ РАН, составляет порядка 15 статей, а загрузка полных текстов – примерно 2 материала в среднем за день. Всего с 01.01.2007 г. зафиксировано более 5 тыс. просмотров и загрузок. Администратор профиля организации в Соционет может развивать связи с сотрудниками и пополнять коллекции статей, создавая условия для роста показателя востребованности своей организации.

Рис. 4, 5 позволяют оценить средний и накопленный уровень востребованности статей и материалов отдельного автора. По графикам можно анализировать пики и спады интереса к материалам автора. В дальнейшем в систему будет добавлена возможность анализировать вклады различных

⁴⁶ Суммируются *uz_person*, *uz_series* или *uz_archive*, *uz_institution* для подразделений данной организации, а также *uz_item* статей/материалов, указанных в связях от профиля организации.

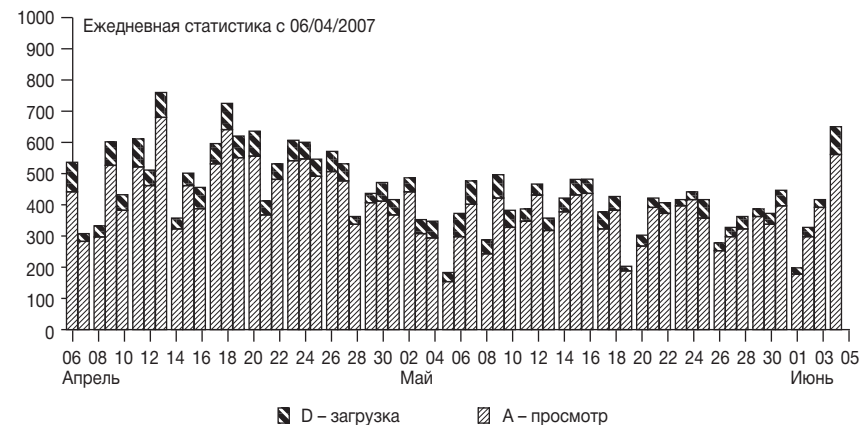


Рис. 1. Динамика полных уникальных ежедневных обращений (*uz_base*) <http://socio-net.ru/stat-base.xml>

статей в показатель востребованности автора в целях выявления наиболее интересных его работ. Сам автор может создавать условия для роста своего показателя востребованности, увеличивая количество собственных статей и материалов в открытом доступе и не забывая устанавливать с ними связи в своем персональном профиле.

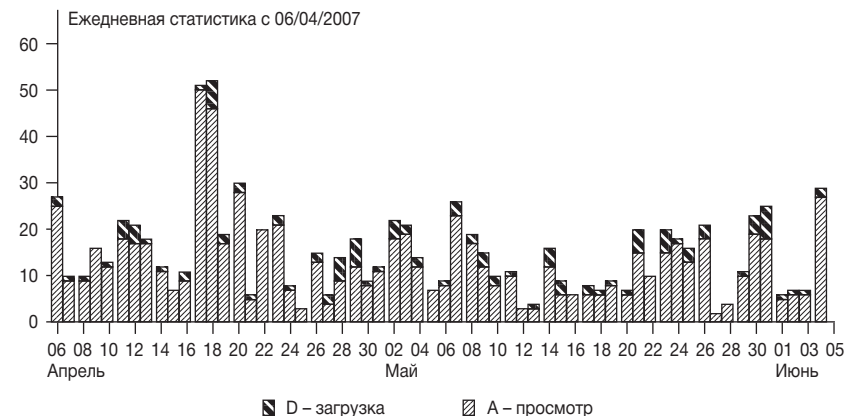


Рис. 2. Динамика востребованности для организации (*uz_institution* для ЦЭМИ РАН), http://socio-net.ru/stat-inst-repec:rus:ecoorg:cemi-ras_admin.45009-org1.xml

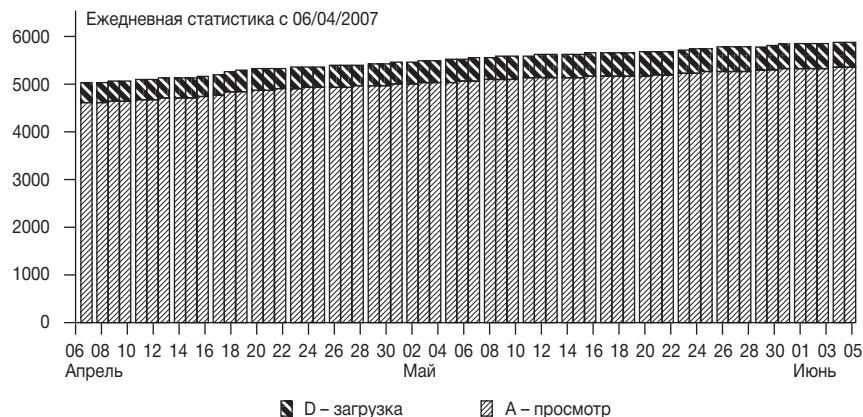


Рис. 3. Данные, аналогичные предыдущему графику, но накопленным итогом

Важно, что количество загрузок полного текста имеет более высокий вес при формировании показателя востребованности автора (прибавляются к количеству просмотров). Это мотивирует авторов выкладывать в открытый доступ не только метаданные (расширенное библиографическое описание) статей, но и их полные тексты.

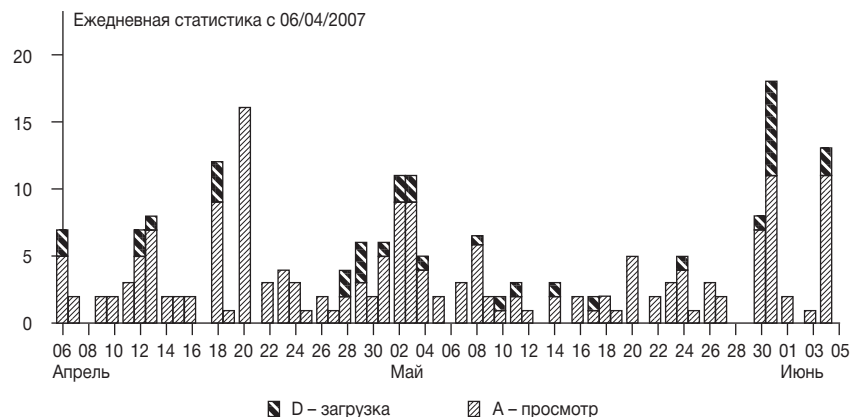


Рис. 4. Динамика востребованности автора (uz_person для Паринов С.И.), http://socionet.ru/stat-auth-repec:rus:escoper:parinov_sergey.56054-1.xml

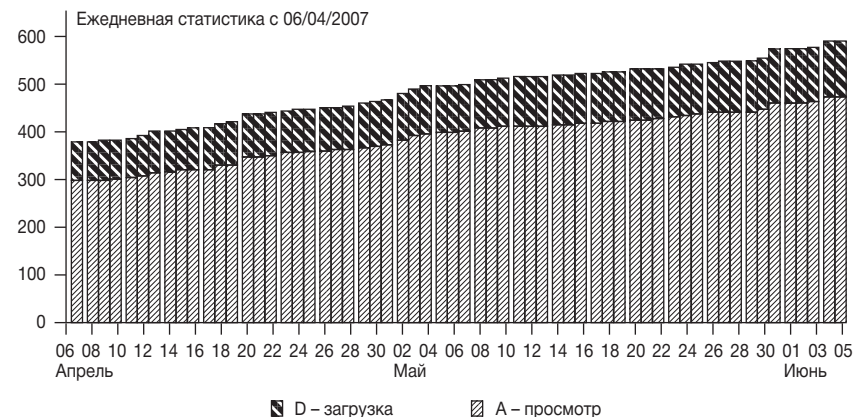


Рис. 5. Данные, аналогичные предыдущему графику, но накопленным итогом

На рис. 6 показана динамика востребованности статьи, имеющей самый высокий в Соционет показатель загрузок полного текста (май 2007 г.).

Поскольку для всех информационных объектов в информационном пространстве Соционет подсчитываются накапливаемые (с начала периода построения статистики) показатели востребованности, то это позволяет также строить таблицы их рейтингов по данному показателю. Таблицы текущих рейтингов по всем типам объектов можно найти по ссылке <http://socionet.ru/stats.xml>.

Не являются исключением и авторы. На странице <http://socionet.ru/statlist-auth.xml> представлен список из 2868 авторов (на май 2007 г.), упорядоченных по аккумулярованному количеству загрузок статей авторов, начиная с 01.01.2007 г. В табл. 1 перечислены первые 20 авторов из этого списка.

Таблица 1

Объекты	Загрузки	Просмотры
Паринов С.И. [профиль автора]	115	468
Гусев А.А. [профиль автора]	46	147
Моткин Г.А. [профиль автора]	38	138
Евсеенко А.В. [профиль автора]	31	26
Унтура Г.А. [профиль автора]	30	51
Прохоров Б.Б. [профиль автора]	29	82

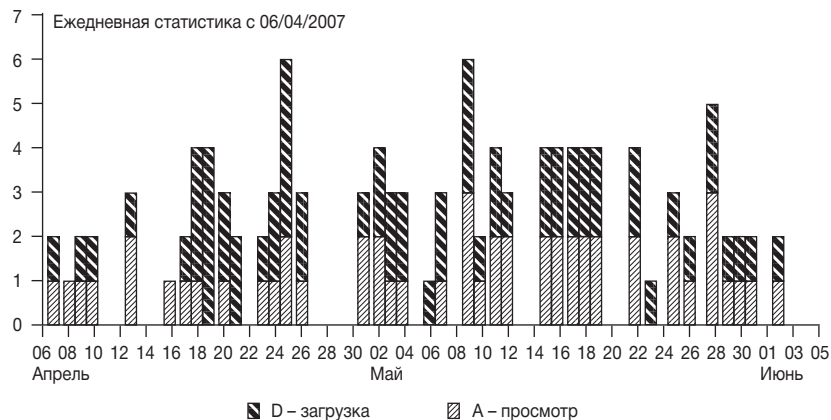


Рис. 6. Динамика востребованности статьи (uz_item – «Анализ социологических данных с применением статистического пакета SPSS»*), <http://socionet.ru/stat-doc-repec: rus:nsusme:rostovtsev.xml>

* *Ростовцев П.С., Ковалева Г.Д.* Анализ социологических данных с применением статистического пакета SPSS. Новосибирск: НГУ, 2001 (<http://socionet.ru/publication.xml?h=repec:rus:nsusme:rostovtsev>).

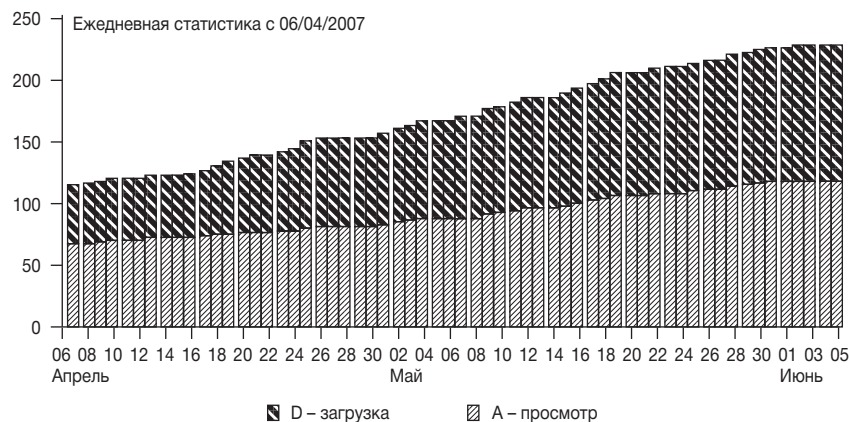


Рис. 7. Данные, аналогичные предыдущему графику, но накопленным итогом

Окончание табл. 1

Объекты	Загрузки	Просмотры
Зверев В.С. [профиль автора]	28	16
Рюмина Е.В. [профиль автора]	27	86
Варшавский А.Е. [профиль автора]	24	142
Цветков В.А. [профиль автора]	24	109
Вальтух К.К. [профиль автора]	24	28
Львов Д.С. [профиль автора]	23	178
Ларичкин Ф.Д. [профиль автора]	23	104
Зулькарнай И.У. [профиль автора]	18	75
Коровкин А.Г. [профиль автора]	17	49
Ляпичева Н.Г. [профиль автора]	15	53
Когаловский М.Р. [профиль автора]	14	58
Дементьев В.Е. [профиль автора]	13	59
Маневич В.Е. [профиль автора]	13	46
Некрасов А.С. [профиль автора]	13	27

В связи появлением в открытом доступе рейтинга ученых возникает новый общественный феномен: авторы, находящиеся в данном списке, приобретают статус «публичного ученого» (по аналогии с «публичным политиком» и т.п.), так как их количественные характеристики вычисляются на основе онлайн-статистики без их прямого участия и результаты доступны в Интернете для всех желающих. Некоторые аспекты этого феномена обсуждаются в Заключении.

3.4. Показатели активности

Показатели активности являются динамическими и в первую очередь характеризуют интенсивность, с которой ученые, а также организации/подразделения выкладывают (а значит, прежде и создают) новые статьи/материалы в открытый доступ. С другой стороны, данная активность также проявляется в изменении коллекций, архивов и информационного пространства Соционет в целом. Применительно к последним объектам активность, по сути, означает интенсивность их пополнения новыми статьями/материалами в единицу времени.

Рассчитываются следующие показатели:

а) *ac_series* — показатель интенсивности пополнения коллекции статей/материалов — характеризует темп роста коллекции в единицу времени: сколько новых статей/материалов появилось в ней за заданный промежуток времени (прирост за месяц или год);

б) *ac_archive* — показатель интенсивности пополнения архива, объединяющего коллекции, принадлежащие одной организации, — рассчитывается суммированием абсолютных приростов размеров всех коллекций архива за заданный промежуток времени (прирост за месяц или год);

в) *ac_person* — показатель активности ученого — учитывает рост количества его статей/материалов, выложенных в открытый доступ за заданный промежуток времени (прирост за 1 год), подсчитывается количество новых связей со статьями и материалами, установленными из профиля ученого;

г) *ac_institution* — показатель активности института/подразделения — фиксирует изменение количества статей/материалов, принадлежащих сотрудникам, а также размера коллекций института/подразделения (без дубликатов и повторов), подсчитывается на основе связей, установленных из профиля организации/подразделения;

д) *ac_base* — показатель активности всего профессионального сообщества, учитываются статьи/материалы, которые выкладываются в открытый доступ на сервер Соционет, считается как общее количество новых статей/материалов, поступивших в информационное пространство Соционет на заданном сервере (*base*) за заданный промежуток времени (день, неделя, месяц, год).

Показатель *ac_base* — интенсивность пополнения информационного пространства в целом — уже рассчитывается в системе Соционет и доступен пользователям на странице <http://socionet.ru/portal.html> в виде графика динамики суточных порций новых поступлений за последние 30 дней, а также недельных порций за последние 30 недель. См. ниже реальные примеры этих графиков.



В дальнейшем предполагается разработать показатели, характеризующие другой вид активности — активность пользователей ip (заданные и(или) среднестатистические) в просмотре статей и материалов. Этот показатель характеризует скорость изменения усредненного количества запросов в расчете на одного пользователя за заданный промежуток времени (день, неделя, месяц, год). Планируется также проводить анализ распределения (структуры) «интересов» пользователей по отношению к множеству объектов информационного пространства Соционет. Это позволит оценить среднюю структуру запросов с одного и того же исходящего адреса (ip), а также скорость изменения структуры интересов со временем.

3.5. Показатели влияния

К середине 2007 г. в информационном пространстве Соционет существовало не более 10 статей/материалов, в которых были установлены связи с другими материалами типа электронного цитирования. Для практического вычисления показателей данного вида пока нет достаточных данных. Популярный подход, когда данные о связях цитирования формируются на основе автоматических процедур (в результате анализа списков литературы в полных текстах статей, представленных в он-лайн⁴⁷), на наш взгляд, является неточным. В том числе и потому, что упоминание статьи в списке литературы не всегда означает ее реальное использование в соответствующем исследовании.

Оформление научного результата в классической форме статьи не является, с точки зрения современных потребностей, удобным средством для отражения факта использования автором результатов исследований других ученых. Аккуратная фиксация факта использования результатов чужих исследований требует определения некоторых дополнительных характеристик этого использования (например, фиксация конкретного значения на шкале от «полностью принимаю» до «полностью отрицаю» и т.п.). Кроме этого раздел «Литература» обычных научных статей не предназначен для точного компьютерного учета связей цитирования.

На наш взгляд, существует потребность в разработке новой модели фиксации научного результата, включая установление связей с использованными исследованиями. Ныне действующая модель «научной статьи» создана в доинтернетовскую эпоху и должна быть адаптирована к техническим возможностям онлайн-научной инфраструктуры (research e-infrastructure) и организационным потребностям e-Science.

Мы предлагаем считать, что установление связей цитирования (что фактически означает определенный режим использования чужих иссле-

⁴⁷ Подобный подход используется, например, на <http://citec.repec.org/>

дований) является частью процесса формирования профессиональной социальной сети ученого. В Соционет уже созданы начальные средства для формирования профессиональных социальных сетей. В частности, для целей электронного цитирования при создании описаний статей (метаданных) рекомендуется использовать поле «Использованная литература». Развитие данных средств для построения более адекватной современным условиям онлайн-модели фиксации научного результата обсуждается в Заключении.

Влияние результатов исследований, измеряемое обычно в виде индекса, в литературе определяется следующим образом:

«Индекс влияния отдельного исследования означает степень его полезности для других ученых и [других его] пользователей в генерировании новых исследований и разработок: как часто данная работа была прочитана, использована, процитирована в других исследованиях, а также в образовательных, технологических, культурных, социальных и практических приложениях» [8].

Исследователи онлайн-индексов влияния предлагают следующую концепцию для данных измерений (цит. по [8]):

«Первым приближением к измерению индекса влияния результатов исследования является сама публикация (как факт признания результата достойным к опубликованию). Исследование, которое не дало результата в виде любого опубликованного результата, не имеет влияния.

Вторым приближением к измерению индекса влияния результата исследований является место публикации. Чтобы быть принятой к публикации, статья исследователя должна быть прорецензирована, т.е. оценена квалифицированными специалистами с точки зрения ее соответствия стандартам качества журнала. В большинстве научных областей существует иерархия журналов, вершина которой использует самый жесткий отбор по самому высокому стандарту качества. Поэтому второе приближение к измерению индекса влияния для исследовательских статей соответствует месту журнала, который принял статью к публикации, в данной иерархии. Однако даже опубликованная в журнале высокого качества статья не имеет влияния, если никто не собирается ее читать.

Третье приближение к индексу влияния связано с уровнем использования статьи. Определить его для печатных изданий довольно трудно, но в онлайн-эру загрузки текста статьи могут быть легко посчитаны. И даже если статья загружалась и читалась, она может остаться невостребованной для новых исследований и приложений.

Четвертое измерение, которое в настоящее время является самой точной оценкой индекса влияния научной статьи, связано с индексом цитирования статьи, который указывает на использование ее содер-

жания (другими учеными, не авторами данной статьи) в новых исследованиях».

Таким образом, онлайн-индекс влияния статьи, выложенной в электронном виде в открытый доступ в онлайн-репозиторий исследовательской организации, может представлять собой функцию от следующих параметров:

- количество просмотров описания статьи и загрузок ее полного текста, исключая повторные попытки и загрузки сетевыми роботами от поисковых сервисов типа Яндекс, Google и т.п.;
- количество цитирований статьи, т.е. ее упоминаний в других выложенных в открытый доступ статьях в разделе «Литература» или «Использованные материалы», исключая случаи самоцитирования;
- показатель влияния авторов, которые цитируют исходную статью в своих исследованиях;
- рейтинг (импакт-фактор) издания, в котором статья была опубликована или депонирована (для изданий без рецензирования).

Индексы влияния могут быть рассчитаны для:

- статьи, на основе вышеописанных параметров;
- ученого, как агрегирование индексов влияния его статей;
- подразделения, как агрегирование индексов влияния его сотрудников;
- исследовательской организации, как агрегирование индексов влияния ее подразделений;
- научного направления, как агрегирование индексов влияния исследователей, занимающихся данным направлением;
- научной дисциплины, как агрегирование индексов влияния статей, принадлежащих данной дисциплине, на другие дисциплины. Это позволяет измерить уровень междисциплинарности науки и его изменение со временем. Кроме того, представляет интерес исследование разбиения корпуса статей заданной дисциплины на кластеры, в которых статьи связаны друг с другом цитированием, а также изменения весов связей между кластерами со временем.

Автоматический расчет индексов влияния, их визуализация вместе с другими онлайн-показателями Соционет, а также их использование при принятии решений об аттестации ученых, установлении персональных научных надбавок и т.п., создает условия для формирования новой онлайн-поведенческой модели ученых.

Авторы получают дополнительную мотивацию к популяризации и продвижению в профессиональном сообществе результатов своих исследований для их более частого онлайн-цитирования и, как следствие, повышения своего индекса влияния. В электронной среде существенно

легче и быстрее происходит информирование коллег о появлении новых статей/материалов. Средства персонального информационного мониторинга ученых позволяют им в автоматическом режиме отслеживать появление материалов, отвечающих их профилю интересов⁴⁸. Проведенные исследования показали, что выкладывание статей в открытый доступ (в форме электронного депонирования) сразу после их готовности повышает количество их цитирований в 2 раза [4].

Кроме того, электронные версии статей, выложенные в открытый доступ в Соционет их авторами, могут редактироваться в любое время, включая изменение раздела «Использованная литература». Это означает, что авторы могут развивать текст статьи и добавлять в него цитирование других источников по мере их обнаружения. Уже выложенные в он-лайн статьи могут служить источником для постоянного повышения индекса влияния других статей.

Большой вес имеет цитирование статей авторами, которые сами обладают высоким уровнем влияния. Поэтому продуктивные ученые становятся объектом конкуренции других авторов за право доказать полезность своих результатов для использования в работах мэтра.

Все имеющиеся в информационном пространстве цитирования являются публичными, так как сведения о связях между статьями, а также исходные данные и алгоритмы расчета показателей влияния, открыты для проверки и воспроизведения. Эти данные открыты для анализа и общественного контроля и могут быть публично оспорены при их использовании для принятия решений в конкретных проблемных случаях.

4. Наукометрическая сигнальная система – использование онлайн-показателей

Комплекс онлайн-показателей, состоящий из показателей востребованности, активности и влияния, может быть использован:

- исследовательскими организациями (институтами и подразделениями) для оценки научно-исследовательской деятельности ученых;
- редколлегиями журналов и издателями для принятия решений об отборе материалов для публикации;
- органами управления РАН для наукометрического анализа функционирования сферы научных исследований и разработок;

⁴⁸ Аналогичную функцию выполняет сервис «Персональный информационный робот» Соционет (<http://spz.socionet.ru/robot.htm>).

- финансирующими организациями (фондами) для принятия решений о поддержке отдельных ученых, направлений исследований и т.п.;
- исследователями-науковедами для проведения различных наукометрических исследований.

В публикации [8, с. 10] приводится перечень вопросов, на которые можно получить ответ на основе онлайн-статистики:

- Какие исследования наиболее часто используются? Кем?
- Какие исследовательские направления развиваются наиболее активно? Под чьим влиянием?
- Какие исследования доказывают немедленную полезность, какие долгосрочную и какие имеют устойчивое долгосрочное влияние?
- Существуют ли работы, чье значение вновь актуализируется по прошествии времени? Можем ли мы идентифицировать частоту и природу такого «медленного горения»?
- Какие исследования и исследователи наиболее авторитетны? Чьи исследования больше всего используют эти авторитетные исследования, и наоборот – какие исследования использованы в авторитетных?

Начиная с 2007 г. правительства Великобритании и Австралии реализуют программы оценки научно-технической деятельности, включающие в качестве одной из компонент онлайн-показатели результатов деятельности исследовательских организаций. Данные программы должны начать действовать в полном объеме в обеих странах в 2008 г.

Британская программа Research Assessment Exercise⁴⁹ (RAE) предусматривает оценивание эффективности университетов Великобритании с 4-летней периодичностью. Финансирование университетам выделяется пропорционально эффективности, рассчитанной на основе индекса RAE [4].

Правительство Австралии реализует программу Research Quality Framework⁵⁰ (RQF). Она основывается на полудецентрализованной информационной модели. Децентрализованная часть: полные тексты с результатами исследований будут храниться в информационных репозиториях научной организации (университет), где они выполнены. Централизованная часть: сотрудники службы RQF будут отвечать за центральный репозиторий, собирающий метаданные публикаций из всех университетских репозиториях, и будут обновлять подсчет цитирования независимо от университетов (совместно с организациями Thomson Scientific и, возможно,

⁴⁹ Research Assessment Exercise. RAE 2008 // <http://www.rae.ac.uk/>

⁵⁰ Research Quality Framework: Assessing the quality and impact of research in Australia RQF Development 2006 // http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/research_quality_framework/rqf_development_2006.htm

с EndNote Web). Австралийское правительство финансирует университеты для создания ими своих онлайн-репозиторий и их полный ввод в действие к концу 2007 г.

В статье [8] анализируются возможные следствия существующих в настоящее время тенденций по расширению открытого доступа и применению различных онлайн-показателей:

- авторы, статьи, журналы, институты и направления исследований будут иметь коэффициент эндогамии: как часто они цитируют себя? цитируются в рамках отдельной научной области? среди различных областей? в разных дисциплинах?
- авторы, статьи, журналы, институты и направления исследований будут также иметь коэффициент латентности и долговечности как с точки зрения цитирования, так и скачиваний полных текстов: как быстро увеличивается цитирование и скачивание? как быстро наступает пик? как долго они востребованы?
- компьютерное сравнение текстов и семантический анализ будут распознавать случаи плагиата, нераспознанного параллелизма и потенциальной конвергенции;
- метрика степени близости может быть рассчитана между двумя людьми: статья, авторами, группами и научными направлениями.

5. Заключение

Для того чтобы представить себе суть и содержание онлайн-будущего науки (e-Science) в тех аспектах, которым посвящена данная статья, достаточно продлить в будущее процессы и тенденции, существенная часть которых описана выше. В качестве резюме отметим основные для нас моменты формирования e-Science: международное научное сообщество как глобальная сеть ученых, связанных профессиональным разделением труда, функционирует главным образом на основе Интернет-технологий; большая часть профессиональных взаимодействий осуществляется учеными в он-лайн, где и находится в свободном доступе корпус результатов научной деятельности; профессиональное информационное пространство совместно с социальными сетями визуализирует текущую структуру научного сообщества (кто чем занимается, какие результаты достигнуты и т.п.); онлайн-сигнальная система визуализирует структуру текущих профессиональных интересов (какие темы и ученые наиболее востребованы, какие результаты наиболее используемы и т.п.); ученые принимают решение о темах для исследования на основе данных, получаемых от наукометрической сигнальной системы.

Онлайновая наукометрическая сигнальная система как часть e-Science в идеале должна представлять в открытый доступ текущие индикаторы функционирования научного сообщества, обновляемые в реальном времени. С ее помощью научное сообщество должно получать своевременную и точную информацию о работе системы профессионального разделения труда. Например, что происходит с научно-исследовательской деятельностью как в целом, так и по отдельным дисциплинам и разделам науки? какое место в этом процессе занимает заданный ученый или исследовательская организация? какое влияние они оказывают на совокупную профессиональную деятельность? и т.п.

Построение сигнальной системы для научного сообщества на базе онлайн-наукометрических показателей как новое направление исследований по состоянию на июль 2007 г. находится в самом начале своего развития. На наш взгляд, прогресс в этом направлении в первую очередь зависит от проведения дальнейших исследований по следующим темам:

- 1) онлайн-метрики и системы наукометрических показателей, наиболее точно и полно отражающих процессы, происходящие в профессиональном научном сообществе (востребованность и влияние результатов исследований, состояние науки и тенденции ее развития);
- 2) система правильных стимулов и поведенческие модели для ученых по распространению влияния своих исследований средствами Интернет-технологий, что должно обеспечить значимость и достоверность онлайн-наукометрической статистики и построенных на ее основе показателей;
- 3) онлайн-модели информационного поведения исследовательских организаций и ученых, адаптированные к новым возможностям онлайн-исследовательской инфраструктуры в целях их максимально полного профессионального самовыражения, что в свою очередь должно повысить объективность онлайн-наукометрической статистики;
- 4) механизмы достижения общественного согласия профессионального сообщества ученых по набору онлайн-наукометрических показателей, алгоритмам их подсчета и их институционализации (правилам их применения), которые должны обеспечить доверие к данным показателям в целях их успешного практического внедрения в процедуры оценки и стимулирования деятельности ученых и исследовательских организаций.

В заключение предлагаем два подраздела, относящиеся к реализации тем 2 и 3 из этого списка:

а) объективные процессы развития, формирующие e-Science, приводят к изменению некоторых сторон статуса ученого, в частности, усилению его публичности;

б) некоторые «священные» атрибуты традиционной деятельности ученого, например, написание статей и их публикация в научных журналах, теряют часть своего значения, которое переходит к новым формам деятельности, лучше адаптированным к возможностям онлайн-технологий и потребностям e-Science.

5.1. Статус публичного ученого

Процесс создания и использования новых знаний, а также результативность соответствующих исследований, осуществляемых на средства налогоплательщиков, могут и должны стать прозрачными и понятными для общества. Новые Интернет-технологии и опыт западного научного сообщества в реализации международных инициатив, связанных с открытым доступом к результатам исследований, дают для этого необходимые инструменты и примеры. Фактически это означает повышение уровня публичности как для проводимых исследований, так и для самих исследователей.

Последовательная реализация описанных ранее направлений формирования e-Science модернизирует взаимоотношения между отдельным ученым и научным сообществом, а в его лице и с обществом в целом. Вот важные, на наш взгляд, пункты изменений, возникающие исходно на стороне ученого:

1. Ученый размещает в открытом онлайн-доступе 100% своих статей с результатами законченных исследований. Это, с одной стороны, является актом немедленного электронного депонирования полученных результатов и фиксирует приоритет их автора, а с другой — открывает возможность их раннего распространения и использования в научном сообществе для получения нового научного знания. Отчуждение результатов исследований от их авторов будет полным, если электронные депоненты статей не остаются на частных сайтах, а попадают в профессиональное информационное пространство, создаваемое интеграцией метаданных институтских ОА в единую национальную или международную информационную систему.

2. Ученый имеет техническую возможность на протяжении многих лет при необходимости вносить изменения в размещенные им электронные депоненты статей, устанавливая различные виды связей между своими статьями и другими объектами информационного пространства, например, материалами других авторов (цитирование). Материалы такого типа имеют статус «живого» документа, они видоизменяются по мере развития

научных представлений их автора. Подобного ранее не существовало и поэтому сейчас необходимо определить статус таких материалов.

3. Ученый может предлагать свои депонированные в ОА статьи в редколлегии научных журналов с целью получения для них статуса публикации. В этих условиях факт опубликования электронного депонента в научном журнале означает способность данного материала пройти через научное рецензирование и стать частью избранной коллекции качественных научных публикаций, что гарантировано статусом соответствующего журнала. При этом научные журналы теряют одну из своих прежних монопольных функций: распространение научных знаний, так как полный текст электронного депонента исходно уже находится в открытом доступе. За что тогда будут платить деньги подписчики журналов?

В ответ на стороне научного сообщества возникают свои изменения:

4. Научное сообщество проявляет интерес к электронным депонентам и публикациям ученых в открытом доступе, что измеряется в количествах их просмотров и загрузок полных текстов. Это фиксируется онлайн-статистикой и доводится до всех, кому это интересно, сигнальной системой в виде показателей востребованности результатов исследований. Кроме того, размещенные в открытом доступе материалы автора цитируются в электронных депонентах и публикациях других ученых. Индексы онлайн-цитирования электронных материалов позволяют построить показатели влияния исследований заданного автора на процессы создания нового знания в научном сообществе в целом. Данные показатели рассчитываются автоматически и обновляются в реальном времени для *всех* ученых, чьи материалы представлены в открытом доступе.

5. Каждый исследователь (в общем случае — каждая исследовательская организация), *хочет он этого или нет*, получает публичный индекс (или набор индексов), характеризующий его деятельность. Значения этих индексов служат основой для публичного сравнения ученых между собой, на этой основе создаются и обновляются их рейтинги. Ученый получает статус публичной фигуры как «плату» за использование средств госбюджета. Необходимы правила, которые бы регламентировали права и обязанности ученых в этом новом контексте (например, какой ценой ученый может отказаться от публичного статуса).

6. Научное сообщество имеет возможность использовать публичные показатели *востребованности* и влияния результатов исследований ученых для материального и морального стимулирования их работы. Аттестационные комиссии в институтах могут использовать ранжирование ученых по данным показателям наряду с традиционными способами. Научные фонды и другие организации, финансирующие научные исследования, могут на основе данных показателей принимать решение о поддержке определен-

ных ученых и(или) определенных направлений исследований. Подобные возможности также требуют тщательного изучения.

5.2. Онлайн-модель фиксации результата исследования

Эффективность наукометрической сигнальной системы в существенной степени зависит от точного учета качественных и количественных характеристик использования результатов исследований. Традиционно это измеряется на основе данных о цитировании, извлекаемых из раздела «Литература» статей. Такой подход дает удовлетворительные результаты⁵¹ в количественных измерениях, но при этом достаточно слабо измеряются качественные характеристики. Главная причина этих проблем заключается в несоответствии формы цитирования (в общем случае – использования) чужих результатов исследования, допустимой в рамках жанра традиционной научной статьи, современным потребностям создания наукометрической сигнальной системы. Необходима новая форма фиксации научного результата автора, которая, в том числе, позволит разнообразить способы использования результатов других авторов. Она должна быть адаптирована к новым возможностям, предлагаемым исследовательской онлайн-инфраструктурой, и лучше соответствовать духу и требованиям e-Science.

Не отказывая в праве на существование традиционным научным статьям, мы хотим предложить к обсуждению идеи, которые могли бы привести к разработке дополнительной специализированной формы фиксации научного результата:

1. Необходимо разделить функции, которые сейчас объединяет в себе традиционная научная статья, между разными специализированными средствами фиксации научного результата. В первую очередь это касается различных наукометрических измерений вкладов и влияний авторов. В целях повышения точности измерения научного вклада и влияния автора нужно дать ему возможность представить научный результат в максимально «сухой» и формализованной форме. Например, в виде заполнения специально сконструированной онлайн-формы. Результат ее заполнения некоторым ученым предстает для научного сообщества, как минимум, в двух видах: а) появление новых сущностей или узлов в глобальной сети понятий научного знания (собственно новый научный результат); б) появление в этой сети новых разнокачественных связей между существующими узлами (связи использования и т.п.). Эти добавления к сети научного знания являются авторскими и доступны всем другим ученым для различных вариантов использования, специфицированных сценариями новой формы фиксации научного результата.

2. Реализация подобных идей означает, что ученый будет иметь два специализированных инструмента для фиксации результатов своих исследований:

1) онлайн-интерфейс, специально сконструированный для авторского добавления/развития узлов и(или) связей между ними в глобальной сети научного знания;

2) написание научной статьи, жанр которой специализирован на популяризации и продвижении результатов исследований ученого в научном сообществе, и ее публикация в научных журналах (таким образом в e-Science уточняется ниша для журналов, см. п. 3 в подразделе «Статус публичного ученого»).

3. Одной из уже существующих систем, которая могла бы стать образцом онлайн-интерфейса для фиксации научного результата, на наш взгляд, является платформа Вики (<http://ru.wikipedia.org/>). Еще лучше для этого подходит ее новая реализация «Семантическая Вики» (http://ru.wikipedia.org/wiki/Semantic_Wiki). С учетом необходимого развития новую систему можно было бы назвать «Научная Вики». Стилистика добавления нового узла в глобальную сеть научного знания может быть прямо позаимствована из Вики: ученый в традиционном для Вики стиле создает страницу, которая посвящена описанию результата его исследований. Первое отличие от обычной Вики – все узлы должны быть авторскими (имя автора должно быть в явной форме представлено на странице и должно быть связано с его персональным профилем).

Если в своем тексте ученому необходимо сделать ссылки на результаты других ученых, то Научная Вики должна позволять ему получить список соответствующих узлов сети научного знания, конкурирующих между собой за право быть процитированным ученым. Если ни один из существующих узлов не устроил ученого, то он должен иметь возможность присвоить ссылке статус заказа на необходимый ему для цитирования в своей работе научный результат, сопроводив этот заказ комментариями, почему имеющиеся результаты ему не подходят. Если кто-либо из ученых заявит по прошествии времени, что данный заказ выполнен, то автор исходной статьи должен получить об этом уведомление.

Механизм использования чужого результата исследований, которое обычно имеет вид гиперссылки на соответствующий узел/материал, также требует развития. Научная Вики должна давать ученым достаточно широкие возможности для спецификации качественных параметров использова-

⁵¹ См., например, <http://citec.repec.org/>, а также обзор литературы по индексированию на http://elibrary.ru/projects/citation/cit_literature.asp

ния чужих результатов. Автор должен иметь формализованный интерфейс для спецификации своего отношения к цитируемому материалу, фиксации роли данного цитирования в своем материале, и т.д. и т.п.

4. Если представить, что Научная Вики реализована, то наукометрические задачи оценки вкладов и влияний ученых получают гораздо более благоприятные условия для решения. В условиях конкуренции научных результатов разных авторов за право быть процитированными только наиболее значимые из них получают максимальное количество ссылок с материалов, находящихся в зоне сбора онлайн-статистики. Все это может быть аккуратным образом зафиксировано и количественно оценено. Поскольку ссылки цитирования имеют в Научной Вики развитый набор формализованных атрибутов, то программным образом можно оценивать и качество цитирования.

В результате данных мероприятий отдельный ученый получает эффективную систему распространения и продвижения созданных им новых знаний, а также надежную сигнальную систему, информирующую его о реакции общества на результаты его исследований и об общих тенденциях в науке.

Исследовательская организация получает механизм оперативного доведения результатов коллективных исследований своих сотрудников до возможных потребителей, достаточно объективные показатели результативности вклада и влияния своих исследователей, а также показатели востребованности реализуемых ими научных направлений.

Научное сообщество получает дополнительную систему мотиваций, которая способствует более эффективному выполнению учеными их общественных функций по созданию новых знаний, и статистическую базу для разносторонних наукометрических исследований текущего функционирования и тенденций развития науки в целом.

Логика развития Интернет-ресурсов и онлайн-сервисов для научных исследований и разработок подвела нас к рубежу, когда не только стали возможны, но и очевидно необходимы меры по модернизации форм и методов научной деятельности. Мы рассчитываем, что данный материал внесет свой вклад в определение наиболее перспективных направлений реорганизации российской науки.

6. Литература

1. Берлинская Декларация об открытом доступе к научному и гуманитарному знанию (пер. Э.М. Мирского) // <http://informika.ru/text/magaz/newpaper/messedu/2003/cour0311/200.htm>
2. Berlin 3 Open Access: Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. 28 February – 1 March, 2005, University of Southampton, UK // <http://www.eprints.org/events/berlin3/outcomes.html>
3. *Harnad St.* 2005. The Implementation of the Berlin Declaration on Open Access. Report on the Berlin 3 Meeting Held 28 February – 1 March 2005, Southampton, UK // D-Lib Magazine. 2005. March. Vol. 11. N 3 // <http://www.dlib.org/dlib/march05/harnad/03harnad.html>
4. *Харнад С.* Максимизация научного эффекта через институциональные и национальные обязательства самоархивирования для открытого доступа // Труды конференции «Current Research Information Systems'2006», Берген, Норвегия, 11–13 мая 2006 г. (пер. С.Ю. Князевой) // http://socionet.ru/docs/harnad-crisrev_ru.doc
5. Research Quality Framework: Assessing the quality and impact of research in Australia RQF Development 2006 // http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/research_quality_framework/rqf_development_2006.htm
6. Research Assessment Exercise. RAE 2008 // <http://www.rae.ac.uk/>
7. *Harnad St.* 2006. Online, Continuous, Metrics-Based Research Assessment. Future UK Research Assessment Exercise (RAE) to be Metrics-Based. 2006 // <http://eprints.ecs.soton.ac.uk/12130/01/rae-metric.html>
8. *Shadbolt N., Brody T., Carr L., Harnad S.* (2006). The Open Research Web: A Preview of the Optimal and the Inevitable. 2006 // <http://eprints.ecs.soton.ac.uk/12453/>
9. *Harnad St.* Open Access Scientometrics and the UK Research Assessment Exercise. Preprint of Invited Keynote Address to 11th Annual Meeting of the International Society for Scientometrics and Informetrics. Madrid, Spain, 25–27 June 2007 // <http://eprints.ecs.soton.ac.uk/13804/03/oa-scientometrics.pdf>
10. *Паринов С.И.* Информационные хабы. Соционет, электронный депонент. 2006 // <http://socionet.ru/publication.xml?h=репер:rus:mqijxk:9&type=paper>
11. *Паринов С.И.* Программа «Открытый доступ к результатам исследований» в Отделении общественных наук РАН. Соционет, электрон-

ный депонент. 2006 // <http://socionet.ru/publication.xml?h=repec:rus:mqjxx:12&type=paper>

12. *Бездушный А.А., Нестеренко А.К., Сысоев Т.М., Бездушный А.Н., Себряков В.А.* Возможности технологий ИСИР в поддержке Единого научного информационного пространства РАН // Электронная библиотека. 2004. Т. 7. Вып. 6. // <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2004/part6/bnsbs>

Parinov S. E-Science: scientometric signal system: Working paper WP2/2007/01. Moscow: State University – Higher School of Economics, 2007. – 52 p. (in Russian).

Moving of research activity into online environment resulted by: a) arising of complexity of current research information systems (CRIS); b) integration of CRISs under national research e-infrastructure; c) forming of e-Science social shapes as more effective organizational and individual behavior models. Succession in the last issue depends on proper work of scientometric signal system which puts to public different research performance and impact indicators. In this article we analyzed existed trends to e-Science and described a Socionet based scientometric signal system for research institutes of Social Sciences Section of RAS, including online statistics accumulating and its using for making scientometric indicators. We also discuss research online metrics and algorithms of scientometric indicators counting.

JEL: C81, C88.

Keywords: research e-infrastructure, online statistics for scientometrics, research institutes of Social Sciences Section of RAS, Socionet.

Sergey Parinov,
main researcher of CEMI RAS,
professor SU-HSE,
head of Socionet workgroup

Препринт WP2/2007/01
Серия WP2
Количественный анализ в экономике

Паринов Сергей Иванович

**Онлайновое будущее науки:
наукOMETрическая сигнальная система**

Зав. редакцией оперативного выпуска *А.В. Заиченко*

Корректор *Е.Е. Андреева*

Технический редактор *О.А. Быстрова*

ЛР № 020832 от 15 октября 1993 г.
Формат 60×84¹/₁₆. Бумага офсетная. Печать трафаретная.
Тираж 150 экз. Уч.-изд. л. 3,5. Усл. печ. л. 3,1.
Заказ № . Изд. № 690

ГУ ВШЭ. 125319, Москва, Кочновский проезд, 3
Типография ГУ ВШЭ. 125319, Москва, Кочновский проезд, 3