

# Технология поддержки электронных научных публикаций как «живых» документов \*

© С.И. Паринов

Центральный экономико-математический институт РАН  
sparinov@gmail.com

© М.Р. Коголовский

Институт проблем рынка РАН  
kogalov@cemi.rssi.ru

## Аннотация

Практика размещения самими учеными результатов своих исследований в форме научных статей и материалов в открытом доступе в сети Интернет постепенно получает организационную поддержку. Все большее распространение в научной среде получают идеи открытых архивов, свободного доступа к результатам исследований, самоархивирования в форме препринтов или постпринтов, а также требования к ученым от гранто- и работодателей по обязательному электронному депонированию всех законченных результатов исследований. Научные статьи и материалы, депонируемые их авторами в электронном репозитории своей организации, являются частью профессиональной информационной среды. Они цитируются наряду с "полноценными" публикациями в рецензируемых журналах. При этом онлайн-средства для электронного депонирования являются общедоступными и достаточно просты в использовании. Как следствие, авторы научных статей и материалов могут вносить в них изменения в общем случае в течение всей своей профессиональной жизни. При массовом использовании подобной практики электронные научные статьи и материалы получают статус "живого" документа (в зарубежной литературе "liquid publication"- «текущая (или неустойчивая) публикация»). Появление научных статей с подобным статусом имеет как положительные, так и

отрицательные последствия. В работе сформулированы условия, при которых происходит ослабление негативных и усиление позитивных аспектов превращения научных статей в "живые" документы. Предложенные решения практически реализуются в системе Соционет. Описаны особенности реализации данных решений в виде инфраструктурных онлайн-сервисов, что открывает научному сообществу достаточно разнообразные возможности для их использования.

## 1 Введение

В последние годы у ученых появилась возможность самостоятельно выкладывать (депонировать) свои статьи и материалы (в виде пре- и постпринтов) в открытый доступ благодаря многочисленным электронным библиотекам, онлайн-репозиториям и открытым электронным архивам. Все более распространенной особенностью электронных репозиториях является неограниченная возможность авторов электронных статей вносить в них изменения на протяжении всей своей профессиональной жизни, что превращает эти ресурсы в "живые" документы.

Появление электронных научных документов с подобным статусом имеет как позитивные, так и негативные последствия.

Данная проблема становится актуальной, и ее исследование обладает достаточной новизной как в российской науке, так и в международной. Первая известная нам статья, в которой данная проблема рассматривается в терминах "текущей публикации" ("liquid publication") была опубликована итальянскими учеными на английском языке в 2007 г. [1, 2]. На русском языке, по-видимому, первое обсуждение данной проблематики в терминах "живого" документа появилось также в 2007 г. в

---

Труды 11<sup>й</sup> Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2009, Петрозаводск, Россия, 2009.

журнальной публикации [3] одного из авторов данной статьи.

Отражением актуальности данной темы является выделение гранта 7-й Рамочной Программы ЕС проекту LiquidPub (<http://project.liquidpub.org/>), а также все более частое упоминание проблематики "living publication" в различных крупных международных проектах по открытому доступу к результатам исследований (что, например, было замечено авторами этой статьи в презентациях на OAI6 - CERN workshop on Innovations in Scholarly Communication, <http://indico.cern.ch/conferenceDisplay.py?confId=48321>).

В предлагаемой статье содержится анализ данной проблемы, сформулированы условия, при которых, по мнению авторов, происходит ослабление негативных и усиление позитивных аспектов превращения электронных научных материалов в "живые" документы. Предложены практические решения для программной реализации данных инноваций в системе Соционет (<http://socionet.ru/>). Описаны особенности реализации данных решений в виде инфраструктурных онлайн-сервисов, что открывает научному сообществу достаточно разнообразные возможности их использования.

## 2 "Живые" документы и связанные с ними проблемы

Прежде всего, уточним смысл понятия "живой" документ. В ряде случаев авторы представленных в цифровом виде научных статей, экспериментальных данных или других электронных научных документов, публикуемых в электронных библиотеках или в других свободно доступных онлайн-репозиториях, осуществляют обновление контента этих научных документов на протяжении их жизненного цикла в соответствующей информационной среде. Такие обновляемые электронные научные документы, поддерживаемые в онлайн-информационной среде, будем называть «живыми» документами (living document).

Конечно, не каждый из публикуемых в онлайн-информационных средах электронных научных документов может стать «живым» документом. Не станут такими документами, например, авторефераты или полные тексты уже защищенных диссертационных работ. Многие научные статьи также не попадут в эту категорию. Вместе с тем, имеется целый ряд типов научных документов, которые могут стать «живыми» документами. Приведем лишь несколько примеров:

- обзорная статья в некоторой области науки, которую автор может пожелать актуализировать по прошествии некоторого времени с момента ее публикации в связи с появлением новых значимых результатов в охватываемой данным обзором области

- пополняемая научная библиография по какой-либо тематике исследований
- описание пополняемой музейной коллекции в систематизирующих областях науки, например, ботанического гербария, коллекции насекомых, минералов и т.д.
- отчет, представляющий результаты многоэтапного научного проекта
- опись архивных документов
- математическая статья, автор которой нашел более изящное доказательство теорем по сравнению с приведенными в данной статье.

Поддержка «живых» научных документов, как уже отмечалось, порождает ряд проблем, которые подробно обсуждаются вместе с предлагаемыми подходами к их решению в дальнейших разделах статьи. Здесь мы лишь заметим, что сложность поддержки «живых» документов связана с динамичностью структуры информационной среды, элементами которой они являются. Эта структура образуется явно представленными ссылками между документами, обладающими различными свойствами. Однако семантическая адекватность этих ссылок зависит от состояния документов, на которые направлены эти ссылки. Обычно ссылка на какой-либо документ из другого документа относится к некоторому фрагменту целевого документа. Если этот фрагмент в «живом» целевом документе исчезает при его редактировании автором или его содержание изменяется, ссылка на него становится или может стать семантически неадекватной.

Таким образом, проблема поддержки «живых» документов связана именно со структурными аспектами информационной среды, в которой они существуют. Поддержка этой структуры требует наличия как специальных сервисов в соответствующем онлайн-информационном репозитории, так и в онлайн-научной инфраструктуре, создающей единую научную информационную среду из репозитивов отдельных организаций.

Функциональные требования к таким сервисам и другие технологические вопросы поддержки «живых» документов подробно рассматриваются далее.

## 3 Проблемы поддержки «живых» документов

Массовое использование информационных и интернет-технологий в научно-исследовательской деятельности, кроме прямых и очевидных выгод, иногда порождает ситуации, которые, на первый взгляд, являются проблемными и мешают нормальному функционированию созданных технологий. Подобные проблемы иногда сигнализируют, что данные технологии имеют скрытый потенциал, который также может быть использован, если будут созданы необходимые для этого условия. При этом важно определить когда

эффект от использования подобного потенциала информационных технологий превосходит затраты на создание необходимых дополнительных условий. Следовательно, усилия по созданию необходимых условий имеют смысл и будут оправданы.

Примером подобной ситуации является появившаяся в последние годы возможность для ученых самостоятельно выкладывать свои статьи и материалы (в виде пре- и постпринтов) в открытый доступ через многочисленные онлайн-репозитории и открытые электронные архивы. Явным позитивным следствием этого является ускорение кругооборота научных знаний. Статистически зафиксировано [4], что подобная практика увеличивает количество цитирований работ ученых примерно в два раза. Обратная сторона - у ученых появляется неограниченная возможность вносить изменения в электронные версии своих статей на протяжении всей своей профессиональной жизни, что превращает такие научные статьи в «живые» документы.

Появление у научных статей и материалов статуса "живых" документов имеет для научного сообщества, как минимум, следующие важные последствия:

- ученый в ответ на критику или по собственной инициативе может вместо создания новой статьи доработать свой исходный текст и внести исправления в текущую версию сделанного им электронного депонента данной статьи; в результате данная статья, формально представляя собой информационный объект с тем же идентификатором (URI, URL или "handle" в Соционет) в электронном информационном пространстве, содержательно будет существенно отличаться от предыдущей версии;

- как результат, все связи цитирования, уже установленные с этой статьей из других научных материалов могут потерять свою актуальность и должны быть пересмотрены их авторами как содержательно (может измениться, или исчезнуть объект цитирования), так и технически (объект цитирования может исчезнуть, или переместиться в другую часть статьи).

Очевидное административное решение данной проблемы путем запрещения изменений в уже депонированных электронных научных документах (т.е. их новые версии заведомо получают статус нового электронного документа), фактически, воспроизводит в электронной среде традиционные принципы бумажных научных публикаций. Таким образом, остаются не исследованными и не использованными потенциальные выгоды от превращения научных статей в "живые" документы.

Сохранение в открытом доступе предыдущих версий научных статей и использование механизма поддержки многоверсионности документов, которым располагают некоторые широко распространенные инструментальные средства для разработки электронных библиотек, является, безусловно, полезным в данном контексте. Однако

поддержка связей между версиями документов является, на наш взгляд, недостаточной, т.к. для использования потенциальных выгод от превращения научных статей в "живые" документы требуется более тонкий механизм, позволяющий создавать, отслеживать и поддерживать сети разнокачественных связей между фрагментами исходных и цитируемых документов. В первую очередь здесь представляют интерес сети связей между результатами исследований различных ученых, формирующиеся вследствие взаимного научного цитирования.

Предполагается, что создаваемый в среде системы Соционет механизм поддержки "живых" документов, позволит сделать более эффективным процесс использования результатов исследований научным сообществом и кругооборота научных знаний.

Использование научным сообществом положительных моментов, связанных с "живыми" документами, означает попытку усовершенствовать профессиональные взаимодействия между учеными, которые в данном конкретном случае облечены в форму написания и редактирования статей, а также их цитирования в других статьях, которые в свою очередь тоже могут изменяться.

Наиболее важные, на наш взгляд, потенциальные выгоды для научного сообщества от полного и эффективного использования феномена "живого" документа связаны с созданием качественно новых условий для функционирования сети профессиональных связей между учеными. Подобные связи возникают, динамически меняются или устойчиво существуют между учеными в процессе создания ими нового научного знания за счет специализации ученых, определенного разделения труда между ними и использования ими результатов друг друга. Переход в научной практике от традиционной статьи к "живому" документу означает в этом контексте существенное повышение эффективности сети профессиональных связей, выполняющей важную роль в процессе современного кругооборота научных знаний.

Поддержка статуса «живых» документов для научных статей требует определенных доработок и развития применяемых сейчас в исследовательских организациях электронных библиотек, онлайн-репозиториях и открытых архивов, включая встроенные в них алгоритмы электронного депонирования и цитирования статей и материалов. Кроме этого, требуется создание новых и развитие некоторых из уже существующих онлайн-сервисов во внешней по отношению к исследовательским организациям среде, которые в данном проекте рассматриваются как элементы научной онлайн-инфраструктуры. Система Соционет [7] является хорошим объектом для реализации этого, т.к. содержит как средства электронного депонирования и создания институтских репозиториях, так и представляет

собой пример реализации онлайн-научной инфраструктуры.

Важная и сложная проблема в данном случае – определение возможного комплексного эффекта от превращения научных статей в «живые» документы и сопоставление его с затратами на создание программно-технических условий для поддержки «живых» документов в практической деятельности научного сообщества.

#### **4 Условия эффективного использования «живых» документов**

Одна из задач – спецификация набора информационных технологий, при которых научное сообщество может получить максимально возможную выгоду от превращения научных статей в «живые» документы. И в последующем – создание пилотных образцов данных технологий. Проведение их опытной эксплуатации авторы планируют в рамках программы «Открытый доступ к результатам исследований» институтов Отделения общественных наук РАН.

Ситуация превращения научной статьи в «живой» документ может быть представлена следующим образом:

- научная статья, обозначим ее «СТ1», выложена в электронном виде в открытый доступ ее автором «А1» (или другим лицом, являющимся представителем автора);
- научная статья «СТ2», выложенная в открытый доступ автором «А2», цитирует содержание «СТ1» (в общем случае цитат может быть несколько, и цитироваться может любое количество статей).

Превращение научных статей в «живые» документы не приведет к нарушению процесса научных исследований при наличии комплекса онлайн-сервисов, обеспечивающих следующие возможности:

- автор А1 может получить список всех статей, в которых цитируется его статья СТ1, а также может включить уведомления о появлении новых цитирований ее или исключении ранее созданных;
- автор А1 при изменении электронной версии статьи СТ1 получает автоматическое уведомление о списке существующих на текущий момент связей цитирования с его статьей СТ1, которые могут быть нарушены в результате его действий по изменению СТ1;
- автор А2 получает автоматическое уведомление о необходимости проверки в его статье СТ2 связи цитирования со статьей СТ1 ввиду того, что автор А1 внес изменения в СТ1 (информирование о возможно необходимом обновлении цитат);
- автор А2 может получить список текущих изменений в СТ1 по сравнению с предыдущей версией СТ1, для которой А2 создал в своей статье СТ2 связь цитирования со статьей СТ1;

- читатели статьи СТ1 должны видеть список всех связей цитирования, которые установлены из других статей со СТ1;

- читатели СТ1 должны видеть для каких связей цитирования с СТ1 не было выполнено обновление цитат после изменения СТ1;

- читатели статьи СТ2 должны видеть предупреждение, если автором А2 не было выполнено обновление цитат на статью СТ1 после изменения СТ1.

Подобные сервисы создадут условия для оперативной координации между авторами статей, связанных друг с другом ссылками цитирования, что обеспечит непрерывную цепную реакцию согласований и обновлений в содержании статей, представляющих собой «живые» документы.

Читатели статей, имеющих статус «живого» документа, также будут знать, какие ссылки цитирования в них гарантированно соответствуют их текущему содержанию, а какие возможно устарели и могут быть проигнорированы.

В системе Соционет существуют все необходимые предпосылки для реализации данного набора сервисов, включая средства для электронного депонирования (размещение статей в открытом доступе), а также инструменты цитирования (создание связей цитирования между статьями, размещенными в электронных библиотеках, открытых архивах или электронных репозиториях).

#### **5 Реализация поддержки "живых" документов в системе Соционет**

Для практической проверки описанного подхода осуществляется разработка пилотной версии необходимого программного обеспечения, которое может быть свободно использовано научным сообществом для извлечения потенциальных выгод от превращения статей в "живые" документы.

Создаваемое программное обеспечение входит в состав системы Соционет (<http://socionet.ru/>) и представляет собой развитие уже существующих в Соционет средств электронного депонирования научных статей и материалов, а также совершенствование уже существующих процедур формирования и контроля связей электронного цитирования.

На данный момент система Соционет в целом представляет собой комплекс следующих подсистем:

1. *Информационный хаб.* Эта подсистема интегрирует в стандартизованную базу данных метаданные открытых архивов и тематических коллекций, принадлежащих различным научным организациям, а также предоставляет по нескольким популярным протоколам все собранные метаданные в стандартизованном виде для их внешнего использования в целях обеспечения возможности внешним разработчикам создавать на их основе новые тематические ресурсы и сервисы.

2. *Профессиональное информационное пространство*, которое предлагает пользователям средства навигации и поиска по содержанию всех метаданных, собранных информационным хабом Соционет.

3. *Онлайновое рабочее место ученого* (в терминах системы – "личная зона Соционет"), которое предоставляет пользователям разнообразные возможности для электронного депонирования результатов исследований, создания коллекций электронных материалов, создания связей между материалами, управления открытыми архивами организаций и т.п.

4. *Персональный информационный робот*, входящий в онлайновое рабочее место, который позволяет ученому в автоматическом режиме отслеживать появление в информационном пространстве Соционет интересных для него материалов.

5. *Профессиональная социальная сеть*, которая визуализирует различные типы взаимозависимостей между электронными материалами, заданные их авторами.

6. *Подсистема представления информационных ресурсов научных организаций* в виде Открытых Архивов (ОА) по протоколу OAI-PMH, который стал стандартом де-факто для распространения материалов в международном научном сообществе.

7. *Фрагмент наукометрической сигнальной системы*, аккумулирующей статистику просмотров метаданных материалов, загрузок их полных текстов и развития сети связей между материалами для расчета и визуализации показателей востребованности, использования и активности по отношению к авторам материалов, научным организациям и другим информационным объектам [6], [8].

Создаваемые в рамках перечисленных выше подсистем средства поддержки статуса «живых» документов должны информировать:

- автора изменяемой статьи, которая цитируется в других статьях, какие связи и цитаты установлены с его статьей и что из них он нарушает в момент ее изменения;

- авторов статей с цитатами из изменяемой статьи о сделанных изменениях в цитируемой статье, для принятия решений о корректировке соответствующих связей цитирования;

- читателей электронных статей о том, что определенные цитаты в них могут быть нарушены.

В связи с этим для создания пилотного образца предлагаемых онлайн-сервисов осуществляется следующая модернизация системы Соционет:

1. Расширение набора атрибутов связей цитирования, включая временные метки создания и обновлений связей цитирования. Развитие модели и алгоритмов электронного цитирования в целом, что включает в себя внедрение в научную практику использования учеными качественных характеристик связей цитирования, а также систему мониторинга связей цитирования и оповещений об

их нарушении или об изменении свойств связей. Подробнее о развитии модели электронного цитирования см. в [5].

2. Модификация структуры базы данных, содержащей сведения о связях между информационными объектами системы.

3. Создание сервисов автоматического мониторинга событий, связанных с изменениями в научных статьях, включая изменение набора и атрибутов связей цитирования.

4. Создание системы генерации отчетов для читателей научных статей и автоматических уведомлений для авторов статей, связанных через цитирование.

Организационной поддержкой данных работ является комплекс следующих мероприятий, проводимых в Отделении общественных наук (ООН) РАН в рамках программы "Открытый доступ к результатам исследований":

- все исследовательские организации ООН РАН создают онлайн-открытые архивы, интегрированные в информационное пространство Соционет и совместимые с международными системами распространения научных материалов (75% институтов ООН РАН на апрель 2009 г. сформировали содержание своих ОА);

- исследовательские организации обязывают своих научных сотрудников депонировать в институтских ОА результаты всех открытых исследований (один из институтов – ЦЭМИ РАН – в апреле 2007 г. ввел в действие данное положение приказом директора);

- использование онлайн-наукометрии при определении персональных научных надбавок исследователей, первая версия наукометрической сигнальной системы функционирует с 01.01.2007, она собирает данные о востребованности статей и материалов из институтских ОА, рассчитывает для публичного просмотра ряд наукометрических показателей, характеризующих статьи, их авторов и научные организации.

## 6 Заключение

Анализ рассматриваемой проблемы, на наш взгляд, показал реальность создания программных средств, способных ослабить (если не исключить полностью) негативные моменты от превращения научных статей в "живые" документы, а также усилить возможные позитивные последствия данного объективного процесса развития научно-исследовательской среды.

Превращение научных статей в "живые" документы даст заведомо положительный результат для научного сообщества при сочетании двух факторов: удачного программно-технического воплощения данной системы и ее массового применения исследователями. Затраты на реализацию этих факторов, на наш взгляд, являются оправданными, т.к. ожидаемая отдача от этих усилий достаточно весома.

Для исследователей это -  
- получение оперативных сигналов о новых цитированиях материалов ученого, удалений уже существующих цитирований, а также об изменении цитируемых результатов исследований;

- возможность быть в курсе развития/изменения научных результатов, которые используются данным ученым в своей работе, а также автоматически оповещать ученых, которые используют результаты твоих исследований, о развитии/улучшении данных результатов;

- наличие комплексной картины, включая историю того, кто, когда и зачем (при внедрении модели цитирования с качественными атрибутами [5]) цитировал материалы ученого и т.п.

Для научного сообщества –

- более высокий уровень информированности ученых о появлении новых результатов исследований;

- новые стимулы и лучшие условия для развития собственных результатов исследований как следствие новых результатов у других ученых;

- улучшение среднего уровня использования результатов исследований в научном сообществе;

- повышение степени профессиональной связанности ученых, использующих результаты друг друга, и, как следствие, ускорение процессов создания нового научного знания.

При массовом использовании технологий поддержки "живых" документов просматриваются перспективы превращения научной статьи в элемент профессиональной социальной сети ученого, в которой связи цитирования между статьями становятся поводом для устойчивых профессиональных взаимодействий между исследователями. Как следствие, возможно превращение корпуса научных результатов во множество взаимосвязанных "живых" документов, когда изменения в одном документе могут породить цепную реакцию изменений в связях цитирования и изменение содержания других "живых" документов.

## Литература

- [1] Fabio Casati, Fausto Giunchiglia, Maurizio Marchese. Publish and perish: why the current publication and review model is killing research and wasting your money, ACM Ubiquity 8 (3), Feb 2007. [http://www.acm.org/ubiquity/views/v8i03\\_fabio.html](http://www.acm.org/ubiquity/views/v8i03_fabio.html)
- [2] Fabio Casati, Fausto Giunchiglia, Maurizio Marchese. Liquid Publications: Scientific Publications meet the Web, Version 2.3, October 1, 2007, <http://liquidpub.org/attachment/wiki/WikiStart/LiquidPub%20paper-latest.pdf>
- [3] Паринов С.И. e-Science - онлайнное будущее науки. // Информационные технологии. - 2007. - № 9. Приложение.

- [4] Стивен Харнад. Максимизация научного эффекта через институциональные и национальные обязательства самоархивирования для открытого доступа. Постпринт, 2006. <http://socionet.ru/publication.xml?h=repes:rus:mqjixk:10>
- [5] Паринов С.И. Новый подход к оценке результатов научно-исследовательской деятельности. Соционет: электронный депонент, 2008, <http://socionet.ru/publication.xml?h=repes:rus:mqjixk:20>
- [6] Когаловский М.Р., Паринов С.И. Информационные ресурсы, наукометрические показатели и показатели качества метаданных системы Соционет. Труды девятой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции, Переславль, 2007». – Переславль-Залесский: Изд. «Университет города Переславля», 2007.
- [7] Паринов С.И., Ляпунов В.М., Пузырев Р.Л. Система Соционет как платформа для разработки научных информационных ресурсов и онлайнных сервисов //Электронные библиотеки. – 2003. - Том 6. - Выпуск 1.
- [8] Когаловский М.Р., Паринов С.И. Метрики онлайнных информационных пространств. Журнал Экономика и математические методы, 2008, т. 44, №2, с. 108-120

## An approach to support electronic research publications as «living» documents

Sergey Parinov and Mikhail Kogalovsky

Currently popular researchers' practice of research results self-archiving in institutional repositories results in researchers' ability to modify text of their publications during a long time. In this case a research publication becomes a living document (or "liquid publication" [1]). In the article we analyze possible benefits of it for a research community and necessary technical conditions to get the benefits. We discuss technical solutions as development of the Socionet system (<http://socionet.ru/>).

---

\* Данная работа поддержана грантами РФФИ проект 09-07-00378 и РГНФ проект 09-02-12117-в