

База данных RePEc и ее российский партнер система Соционет

Thomas Krichel
Palmer School,
Long Island University, USA

Сергей Паринов
ИЭОПП СО РАН, Новосибирск

Аннотация

Онлайновая экономическая библиотека RePEc.org занимает второе место в мире (после arXiv.org) по количеству бесплатно предлагаемых научных материалов. У RePEc совсем другая модель организации пополнения базы данных по сравнению с arXiv и совсем другое содержание предоставляемой информации. Предлагаемая статья посвящена описанию этих особенностей.

Модель организации RePEc имеет архитектуру открытого типа. База данных открыта в двух смыслах: 1) для пополнения (организации могут помещать в нее свои материалы), а так же для развития (разработчики могут создаваться различные сервисы для пользователей). Библиотеки традиционного типа – включая многие электронные библиотеки – закрыты в обоих направлениях. В данной статье также обсуждается особенность функционального соединения с RePEc с системой Соционет.

Что касается содержания предоставляемой информации, база данных RePEc нацелена на создание реляционного набора данных о научных ресурсах, а также связанных с ними сведений. Это должно включать данные обо всех авторах, статьях и организаций, имеющих прямое отношение к исследованиям по экономике. Подобный амбициозный проект может быть реализован только в случае, если затраты на сбор данных будут децентрализованы и малы, а также если выгоды от использования данной информации будут достаточно велики.

1. Введение

Это обычное явление для Интернета – предоставлять средства для организации сотрудничества разработчиков контента по проектам сбора информации. Особенно такое сотрудничество вызывает интерес, когда речь идет об академической информации. Разработчики контента и провайдеры информационных ресурсов очень заинтересованы в распространении своих материалов. Интернет предлагает техническую инфраструктуру, позволяющую функционировать бесплатной системе распространения научных материалов.

Однако этого не достаточно, т.к. требуются социальные инфраструктуры, которые создают условия для использования технической инфраструктуры с минимальным внешним субсидированием.

В период написания данной статьи, существуют две широкомасштабные тематические системы, предназначенные для дальнейшего распространения научной информации в сети Интернет. Самым продвинутым и широко известным является архив arXiv.org, содержащий информацию по физике, математике и компьютерным технологиям. На втором месте располагается экономическая электронная библиотека RePEc.org. Данная статья имеет отношение к последнему. RePEc предоставляет пользователям «Открытую Библиотеку». Обычно, открытая библиотека является коллективной формируемой системой, предоставляющей и использующей текстовую информацию. Действуя на таких же принципах, RePEc собрал свыше 50,000 изданий в электронном виде. Большинство из них доступны всем – совершенно бесплатно. Полные наборы данных также включают ссылочную информацию на внешние публикации, программное обеспечение, исследовательские организации и контактную информацию об авторах. Эти коллекции содержат более 150 тысяч записей к моменту написания данной статьи. Полный набор данных доступен пользователям Интернет через порядка 10 различных интерфейсов.

Во втором разделе статьи, мы представляем RePEc как коллекцию научной информации. В разделе 3 мы продолжаем развивать идею RePEc. Мы обсуждаем возможность расширения RePEc, которые позволят описать научную дисциплину, в большей степени, чем просто документы, созданные специалистами по данной дисциплине. В разделе 4 мы обсуждаем последние достижения в деле дополнительного стимулирования поставщиков информации. Пятый раздел посвящен системе обслуживания пользователей RePEc. Заключение по всей статье представлено в разделе 6.

Создание RePEc началось в 1992. Мы намеренно опускаем описание истории развития, чтобы уделить все внимание настоящему состоянию данной системы. Только поэтому в данной статье нет должного внимания тем людям, которые стояли у истоков RePEc и сделали огромный вклад в развитие сервиса. Ознакомиться с историей развития (включая предшественника - проект NetEc) вы можете в [Krichel \(1997\)](#).

2. Наборы данных RePEc

2.1. Происхождение и мотивация RePEc

Научная система обмена информацией объединяет производителей и потребителей научных материалов. За большую часть таких материалов создатели не получают никакого денежного вознаграждения. Их компенсация за эту работу заключается в широком распространении материалов и общественном признании. Распространение и признание пользователей являются двумя основными функциями научного обмена данными.

Научная система обмена экономической информацией в существенной степени

основана на научных журналах. Рецензирование статей в журналах играет наиважнейшую роль, но является дорогим удовольствием по времени. Согласно [Trivedi \(1993\)](#), обычно требуется свыше 3 лет от момента подачи научной работы до ее размещения в журнале, не говоря уже об отказе в размещении. По неофициальным данным за последние несколько лет постепенный рост времени задержки публикаций приостановился и стабилизировался, благодаря усилиям редакторов журналов, долго боровшихся с данным негативным явлением.

Исследователи в передовых областях науки не могут полагаться только на журналы, так как должны всегда иметь информацию о самых последних достижениях. Общепринятым является предпубликация материалов в препринтах и рабочих документах, а также в материалах конференций. Доступ к подобным материалам часто ограничен узким кругом пользователей, т.к. их распространение основано на энтузиазме и доброй воле самих исследователей.

Временной промежуток между неофициальным распространением и официальным размещением публикаций в академических журналах может быть искоренен только через реструктурированием процесса контроля за качеством (рецензирование) материалов. Неудобство от задержек публикаций, конечно, может быть исправлено путем повышения эффективности неофициального обмена информацией. Это является текущей мотивацией участия создателей информационных ресурсов в проекте RePEc. В RePEc попадают публикации, которые не прошли через рецензирование, поэтому данная система по существу является просто системой распространения научной информации в сети. Она является независимой и самостоятельной по отношению к процессу оценки качества научных материалов.

2.2. На пути к системе распространения научной информации через Интернет

Самый важный объект для распространения научной информации, это - исследовательские работы. Один из путей реализации данного процесса через Интернет был воплощен в [arXiv.org](#). препринт -архиве, организованным в 1991 Полом Гинспаргом в государственной лаборатории Лос Аламоса. Его первоначальной тематикой были исследования по физике частиц. Авторы статей используют этот архив для того, чтобы загрузить свои работы на удаленный компьютер, которые там и хранятся. ArXiv.org содержит сейчас около 170,000 работ, покрывающих широкий круг разработок в различных областях математики, физики, компьютерных технологий. При этом основное внимание уделяется первоначальной области специализации данного архива.

Были предприняты попытки превзойти архив arXiv.org в области экономики, используя "[Economics Working Paper Archive](#)" (EconWPA) поддерживаемый в Washington University in St. Louis. Успеха этой инициативы был ограничен несколькими причинами:

- Экономисты не публикуют свои работы от своего имени, а чаще от экономических факультетов и исследовательских центров.
- Экономисты используют более широкий ассортимент инструментария для

форматирования текста, чем физики. Это снижает функциональность архивное хранение онлайн и усложняет процесс создания хорошего архива.

- Как правило, экономисты не обладают блестящим знанием компьютерной грамотности, и как результат, они в большей степени встречаются с трудностями в процессе загрузки своих работ на удаленный компьютер.
- Существует определенная путаница в восприятии сетевых пред-публикаций на централизованных и хорошо известных системах с точки зрения публикаций в журналах.

Исследования по экономике проводятся не только в университетах и в исследовательских центрах. Существует много государственных учреждений - центральных банков, статистических институтов и других учреждений - которые осуществляют существенный вклад в исследовательскую деятельность в области экономики. Структура этих организаций, благодаря своим размерам, является очень жесткой, что делает более сложным процесс электронных публикаций в централизованных архивах.

Идеальная система должна способствовать гармоничному объединению децентрализованной природы сети Интернет и хорошо организованной системы arXiv.org в абсолютно бесплатный для пользования электронный архив научных материалов. Чтобы объяснить, почему доступ к информационным источникам должен быть бесплатным, нужно рассмотреть отличия коммерческих и эзотерических авторов, которые описаны в [Harnad \(1995\)](#). Авторы научных работ относятся больше к классу эзотерических авторов, чем к коммерческим. Они не рассчитывают на денежное вознаграждение за свои работы; наоборот, они, главным образом, заинтересованы во внимании других подобных авторов, и в меньшей степени, в привлечении внимания широкой общественности. Поэтому авторы данного класса заинтересованы в широком распространении своих материалов. Если доступ к таким информационным источникам облагается определенной платой, тогда сама по себе система не дотягивает до идеальных стандартов.

Способ создания такой идеальной системы - создание сетевых архивов, в которых любое учреждение может хранить информацию о своих публикациях. Такая сетевая структура является открытой, что означает, что любой человек или учреждение может предоставить туда информацию о своих работах. Она называется открытой также потому, что на ее основе может создаваться служба сервиса для пользователей. Подобная двойственность открытости способствует положительному эффекту обратной связи. Чем больше источником пользуются, тем он более эффективен как инструмент распространения информации, а значит, большее количество авторов и учреждений могут стать участниками нашего обмена информацией. Чем больше база данных, тем она полезней для исследователей. А это опять приводит к более частому использованию источника.

Привести систему к такому виду - нелегкая задача. Человек - заложник привычки. Системы научных коммуникаций эволюционируют. Академическая карьера напрямую зависит от результата научных коммуникаций. Поэтому изменения в этой области происходят медленно, т.к. зависят от важных жизненных параметров

тех, от кого этот прогресс зависит. Система научных коммуникаций в Интернете будет более успешной, если она улучшает текущую практику, не заменяя ее полностью. Распространение не рецензируемых научных материалов в прошлом основывалось на активности организаций, которые их создавали. Публикации распространялись на основе обмена. RePEc – это способ чтобы перенести этот процесс в Интернет.

2.3. Архитектура RePEc

RePEc может быть описан как децентрализованная система, дающая возможность распространения научной информации по экономике. RePEc позволяет всем исследовательским организациям включать свои материалы в децентрализованную архивную структуру, что делает эти материалы доступными в Интернете. Отдельные исследователи тоже могут участвовать в этом процессе, но им лучше использовать возможности [EconWPA](http://econwpa.wustl.edu/wpawelcome.html) (<http://econwpa.wustl.edu/wpawelcome.html>).

Любой участник данного процесса обмена должен поддерживать свои собственные коллекции данных, используя набор стандартных форм (темплейтов). Такая коллекция темплейтов, заполненных данными конкретной организации, называется «архив». Архив поддерживается анонимным ftp сервером или веб сервером, которые контролируются провайдером архива. Каждый такой провайдер имеет полный контроль над содержанием своего архива. Им не нужно передавать эти данные в какое-либо централизованное место. Администратор архива имеет полное право вносить поправки и удалять из него документы.

2.3.1. Пример архива

Давайте рассмотрим примерный образ архива. Архив OECD находится по адресу - <http://www.oecd.org/eco/RePEc/oed/>. В указанном каталоге находятся два файла. Первый из них - [oedarch.rdf](#):

Template-Type: ReDIF-Archive 1.0
Handle: RePEc:oed
Name: OECD Economics Department
Maintainer-Email: eco.contact@oecd.org
Description: The working papers of the Economics Department of the OECD
URL: <http://www.oecd.org/eco/RePEc/oed>

Данный файл содержит общую характеристику архива. Оглавление отдельных коллекций, входящих в архив, находится во втором файле - [oedseri.rdf](#):

Template-type: ReDIF-Series 1.0
Name: OECD Economics Department working papers
Type: ReDIF-Paper
Provider-Name: OECD Economics Department
Provider-Homepage:

<http://www.oecd.org/eco/eco/>
Maintainer-Email: eco.contact@oecd.org
Handle: RePEc:oed:oececd

Этот файл предоставляет оглавление архива. Формат обоих файлов - ReDIF. Более подробную информацию об этом формате смотрите в [Krichel \(2000\)](#) .

Публикации, входящие в коллекции, представлены в этом же формате ReDIF. Их описания можно найти в подкаталогах с именем соответствующей коллекции (см. например <http://www.oecd.org/eco/RePEc/oed/oececd>). Эта директория содержит описания публикаций (в формате ReDIF). В ней также может храниться и сам файл с текстом данной публикации. Администратор архива сам определяет, где хранятся полные тексты документов – в самом архиве или вне его. Если документ доступен в онлайн, то в описание публикации включается ссылка (URL) на местонахождение соответствующего файла документа для его загрузки. Файл документа может представлять собой не только текстовый формат, но и быть например, набором данных, компьютерной программой и т.п.

Членство в RePEc не предполагает для администраторов его коллекций, что документы должны быть свободно доступны всем пользователям. Некоторые коммерческие академические журналы предоставляют описание своего контента также в формате RePEc. Если пользователь имеет подписку или необходимые договоренности с издателями коммерческих коллекций (например, JSTOR, старые выпуски журнала *Econometrica* или *Journal of Applied Econometrics*), то хранящиеся в RePEc ссылки позволят просмотреть документы и в этих случаях.

2.4. Документы доступные в RePEc

Более 160 архивов в 25 странах на данный момент составляют коллекции RePEc, принадлежащие различным организациям. Более 100 университетов представляют здесь свои научные доклады, включая например такие американские университеты, как Berkeley, Boston College, Brown, Maryland, MIT, Iowa, Iowa State, Ohio State, UCLA, и Virginia. Коллекции RePEc также содержат информацию о всех научных докладах NBER, исследовательские работы CEPR, содержание базы данных Fed in Print of the US Federal Reserve, коллекции полных документов от IMF, World Bank и OECD. Также материалы многих других исследовательских центров по всему миру. Важной частью RePEc являются коллекции EconWPA. В целом, на время написания данной статьи (сентябрь 2001) из RePEc можно загрузить свыше 50,000 документов.

Библиографические формы и темплейты, предназначенные для описания публикаций, на данный момент включают "документ", "статья" и "компьютерная программа". Темплейт "статья" предоставляет набор полей для полного описания публикации в журнале. Это в данное время используются журналами *Canadian Journal of Economics*, *Econometrica*, *the Federal Reserve Bulletin*, and *IMF Staff Papers*, *the Journal of Applied Econometrics*, *the RAND Journal of Economics*. Этот список отражает только небольшую часть журналов, участвующих в RePEc. Такое участие, естественно, не подразумевает свободный доступ к статьям.

Коллекции RePEc также включают несколько сотен ссылок на темплейты "компьютерная программа", которые включают описания функций, процедур или кодовые фрагменты на языках Stata, Mathematica, MATLAB, Octave, GAUSS, Ox, RATS, также как код на FORTRAN, C and Perl. Возможность заносить в каталог и описывать компоненты программного обеспечения позволяет пользователям, использующим перечисленные языки, искать код, подходящий к их проблеме, даже если он написан на другом алгоритмическом языке. Архивы программного обеспечения, ограниченные только одним алгоритмическим языком, например те, которые поддерживаются частными производителями программного обеспечения, не обладают такой широтой возможностей. Поскольку большинство программ, написанных на языках высокого уровня, могут свободно переводиться, скажем, из GAUSS в MATLAB, то широта возможностей RePEc может быть очень полезна для пользователя.

3. Метаданные ReDIF

Из предоставленной в предыдущем разделе информации, можно нарисовать простую модель RePEc:

Много архивов ==> Одна база данных ==> Много сервисов

Термин "RePEc" первоначально является сокращением от "Research Papers in Economics" (Исследовательские работы по экономике). На самом деле, этот термин приобрел уже более широкое значение, так как RePEc больше чем просто описание ресурсов. Правильнее будет сказать, что RePEc – это реляционная база данных, посвященная экономике, как дисциплине.

В публикации [Karlsson and Krichel \(1999\)](#) предоставлено одно из возможных интерпретаций термина «дисциплина». Ее авторы рассматривают модель дисциплины, как состоящей, главным образом, из 4-х элементов, организованных в таблицу:

ресурс ----- коллекция

человек ---- организация

Понять эту таблицу поможет следующее краткое объяснение. «Ресурс» - это по существу любая форма научной деятельности, включая исследовательский документ, набор данных, компьютерная программа - любой документ, автором которого является научный деятель. «Коллекция» - логическая организация ресурсов в группы. Например, любой обзор может называться ресурсом, если он включен в коллекцию. «Человек» - просто индивид или организация, которые действуют как «человек» в контексте RePEc.

Все вместе это представляет собой реляционную базу данных, которая предоставляет описание не только документов, но и их авторов, а также "организации", где авторы работают, и т.д. Вся предоставляемая в RePEc информация имеет ReDIF формата метаданных. В подразделах 3.2 и 3.3 это иллюстрируется применительно к данным типа "организация" и "человек",

соответственно.

3.1. Более подробно о контенте

Для лучшего понимания основных моментов ReDIF необходимо начать объяснение с примера. Вот - специальным образом подобранный пример ReDIF данных из реальной коллекции, находящийся по адресу -

<ftp://www.econ.surrey.ac.uk/pub/RePEc/sur/surrec/surrec9601.rdf>:

Template-Type: ReDIF-Paper 1.0
Title: Dynamic Aspect of Growth and Fiscal Policy
Author-Name: Thomas Krichel
Author-Person:
RePEc:per:1965-06-05:thomas_krichel
Author-Email: T.Krichel@surrey.ac.uk
Author-Name: Paul Levine
Author-Email: P.Levine@surrey.ac.uk
Author-WorkPlace-Name: University of Surrey
Classification-JEL: C61; E21; E23; E62; O41
File-URL: <ftp://www.econ.surrey.ac.uk/pub/RePEc/sur/surrec/surrec9601.pdf>
File-Format: application/pdf
Creation-Date: 199603
Revision-Date: 199711
Handle: RePEc:sur:surrec:9601

Когда мы смотрим на этот пример, ReDIF данные напоминают стандартный библиографический формат, с авторами, заголовком и т.д.. Единственная вещь, которая кажется немного таинственной - поле "Author-Person". Это поле содержит некий код, который известен системе RePEc. Он ведет к другим данным, доступным по адресу -

ftp://netec.mcc.ac.uk/pub/RePEc/per/pers/RePEc_per_1965-06-05_THOMAS_KRICHEL.rdf

Ниже приводится фрагмент этих данных (мы не показываем некоторые поля для экономии места):

Template-Type: ReDIF-Person 1.0
Name-Full: KRICHEL, THOMAS
Name-First: THOMAS
Name-Last: KRICHEL
Postal: 1 Martyr Court
10 Martyr Road
Guildford GU1 4LF
England
Email: t.krichel@surrey.ac.uk
Homepage: <http://openlib.org/home/krichel>
Workplace-Institution: RePEc:edi:desuruk
Author-Paper: RePEc:sur:surrec:9801

Author-Paper: RePEc:sur:surrec:9702
Author-Paper: RePEc:sur:surrec:9601
Author-Paper: RePEc:rpc:rdfdoc:concepts
Author-Paper: RePEc:rpc:rdfdoc:ReDIF
Handle: RePEc:per:1965-06-05:THOMAS_KRICHEL

В этой записи присутствуют коды (дескрипторы) документов, которые написал данный человек. В таком виде эти данных позволят специальному сервису системы RePEc показать список всех работ данного автора. Это полезно, когда нам необходимо найти все работы одного автора. Также важно иметь централизованную запись о контактных сведениях автора, что позволяет не делать обновление элементов реляционной базы данных при записи в нее каждого документа. Действительно, в этом случае запись в базе данных, соответствующая некоторой публикации, является правильной в момент написания публикации и ее помещения в базу, однако контактная информация о ее авторе будет правильной все время.

В примере заполнения темплейта "человек", представленном выше, мы видим также другой идентификатор в поле "Workplace-Institution". Он также указывает на другую запись в базе данных, которая расположена на <ftp://crefe.dse.ugam.ca/pub/RePEc/edi/inst/desuruk.rdf> и описывает организацию, где работает данный человек.

Template-Type: ReDIF-Institution 1.0
Primary-Name: University of Surrey
Primary-Location: Guildford
Secondary-Name: Department of Economics
Secondary-Phone: (01483) 259380
Secondary-Email: economics@surrey.ac.uk
Secondary-Fax: (01483) 259548
Secondary-Postal: Guildford, Surrey GU2 5XH
Secondary-Homepage:
<http://www.econ.surrey.ac.uk/>
Handle: RePEc:edi:desuruk

Нет необходимости комментировать эти данные в деталях. Более интересно, вероятно, узнать, откуда эти данные поступают в RePEc.

3.2. Регистрация организаций

Регистрация организаций выполняется через проект EDIRC. Это сокращение означает "Economics Departments, Institutions and Research Centers" (Экономические Факультеты, Учреждения и Исследовательские Центры). Этот сервис и набор данных был создан К. Цимерманом ([Christian Zimmermann](#)), профессором экономических наук в университете Квебек, Монреаль ([Universite du Quebec a Montreal](#)) для исследовательских организаций по экономике. Первоначальное намерение заключалось в том, чтобы создать каталог, вмещающую всю информацию об экономических подразделениях в международном масштабе, представленных в Интернете. Таких организаций сейчас много, поэтому на

данный момент уже зарегистрированы около 6,000. Все эти записи включены в RePEc. В базе данных содержится информация о веб-страницах учреждений, их контактная информация. Даже есть информация о специализации некоторых учреждений. Таким образом, всегда можно найти список учреждений, в которых, например, специализацией является экономика труда и т.п. На данный момент содержание EDIRC связывается с остальным содержанием RePEc через HoPEc - персональный регистрационный сервис (см. ниже). Другие способы соединения разнокачественных данных возможны, но редко используются.

3.3. Персональная регистрация

[HoPEc](#) - персональный регистрационный сервис - имеет структуру, отличную от EDIRC. Одному ученому практически невозможно завести данные для всех исследователей в мире, которые занимаются экономикой. Один из возможных подходов - просить администраторов архивов зарегистрировать всех представленных в нем авторов. Это усложнило бы работу администраторам, поддерживающим архивы, но общие затраты были бы небольшими, после того, как все авторы были бы зарегистрированы. Однако материалы одних и тех же авторов участвуют в разных архивах одновременно. Кроме этого авторы могут "мигрировать" из одного архива в другой. Это является проблемой, которая может быть решена регистрацией авторов через одну централизованную систему.

Исследователи, которые регистрируются в HoPEc, идентифицируются с помощью некоторого кода, который обычно включает его имя, а также некую календарную дату. HoPEc предлагает в качестве календарной даты взять дату рождения пользователя. Однако предоставление любой другой даты является приемлемым. Главное, чтобы пользователь мог сразу о ней вспомнить, когда это потребуется. При регистрации пользователь сначала предоставляет некоторую информацию о себе. Запрашиваемая информация обычно включает его имя, URL домашней странички регистрирующегося и адрес электронной почты. Он также может внести информацию о своих академических интересах, используя схему классификации журнала экономической литературы ([Journal of Economic Literature](#)).

Закончив ввод запрашиваемой информации, он приступает ко второму шагу: к созданию ассоциации между информацией пользователя и базой данных в RePEc. Самая распространенная ассоциация - авторство в определенных документах, размещенных в RePEc. Однако, другие варианты тоже возможны. Например, указание того, что пользователь является редактором определенных коллекций. Система регистрации ищет имя пользователя в базе данных RePEc и показывает ему предложения для установления ассоциаций. Пользователь выбирает, какие из предлагаемых ассоциаций верны. Методы проверки достоверности этих ассоциаций - не самая сильная сторона. HoPEc полагается на честность пользователя.

Существует ряд серьезных проблем, с которыми сталкиваются сервисы подобные HoPEc. Во-первых, по причине отсутствия исторически накопленного опыта работы таких сервисов, не так-то просто донести до пользователя цель и значение данного сервиса. Некоторые люди думают, что им нужно регистрироваться, чтобы получить доступ к сервисам RePEc. Благодаря этому мы получаем информацию о

тех, кто заинтересован в использовании RePEc, или более точно – о тех, кто просто не может догадаться, что использование нашей системы не требует регистрации. С другой стороны это, конечно, засоряет нашу базу бесполезными записями. Необходимо также упомянуть, что персональный регистрационный сервис также затрагивает определенные конфиденциальные моменты.

Подводя итоги, надо сказать, что HoPEc предоставляет информацию о пользователях, об их месте работы и исследовательских интересах, а также связывает эту информацию с описанием ресурса в RePEc. Это позволяет лучше идентифицировать "человека" в системе RePEc и обновлять персональные данные своевременно. Предоставленная информация может быть так же успешно использована для других целей, например, поддержание базы данных членства в научных сообществах, для создания списка людей, участвующих в конференциях и т.п. Есть надежда, что данные HoPEc будут использованы в научном сообществе как общий источник персональной информации. После 15 месяцев работы регистрационный сервис оказался вполне успешным. Около 10% всех документов в RePEc сейчас имеют как минимум одного зарегистрированного автора.

4. Стимулы к аккумуляции данных

Существует множество хороших проектов по созданию научных цифровых библиотек, но только arXiv и RePEc выделяются пользователями, благодаря ценности содержащейся в них информации. Аккумуляция информации остается наиважнейшей социальной проблемой, которую не в состоянии решить не одна компьютерная технология сама по себе.

RePEc можно считать успешным проектом в деле привлечения данных. Это особенно заметно, если учесть низкий уровень суммарного финансирования этого проекта. В начале 1992 не было вообще свободно доступных электронных документов. Коллекция была начата путем открытия доступа (через gopher и WAIS) к набору научных материалов по экономике, собранные Fethy Mili, начальником информационного центра факультета экономики университета Монреально (Universite de Montreal). Коллекция электронных документов пополнялась с помощью друзей, сетевых знакомств и через создания ссылок на удаленные документы, например, создавая для них описания в формате, используемом в проекте RePEc. Очень немного организаций предоставили описания своих публикаций для преобразования и хранения нашим проектом. Один из примеров - федеральные резервные банки США. Вторая стадия развития данного проекта характеризовалась тем, что организации начали открывать свои архивы в RePEc. На третьей стадии, которая еще не началась, авторы будут иметь возможность выкладывать документы непосредственно в RePEc через сервис HoPEc. К настоящему времени авторы не были достаточно вовлечены в сбор данных о своих публикациях. Для их вовлечения требуется хороший стимул. Важным компонентом такого стимула является демонстрация авторам, что документы, представленные RePEc, имеют хорошую экспозицию и доступность. Эта демонстрация должна сопровождаться убедительными статистическими данными, поскольку мы имеем дело с учеными, многие из которых привыкли полагаться на статистические показатели.

Первым примером сбора статистики об обращении пользователей к содержанию RePEc является система [LogEc](#). Этот сервис является результатом работы [Sune Karlsson](#). LogEc основан на программе, которая анализирует данные трафика по обращению к описаниям и полным текстам документов RePEc. Эти данные трафика берутся из лог-файла сервера, на котором работает система RePEc. Поскольку в проекте RePEc существует несколько независимых серверов, то на каждом из них настраивается своя программа анализа лог-файла. Когда такая программа отработала, она создает отчет о результатах анализа трафика и сохраняет его на сайте LogEc. Таким образом, LogEc сервис собирает лог-файлы с различных серверов для создания перекрестной статистики пользования ресурсом системы. После этого можно определить наиболее популярные коллекции документов и любой другой информации, хранящейся в RePEc. Это также позволяет оценить популярность авторов, которые зарегистрировали свои публикации в службе HoPEc. Для зарегистрированных авторов мы можем предоставлять им все данные о доступе к их публикациям из лог-файлов для более детального анализа. Мы можем ранжировать авторов по количеству обращений к их публикациям. Это - очень эффективная стратегия привлечь внимание авторов. Созданная Christian Zimmermann система оповещения через электронную почту авторов, которые зарегистрированы в HoPEc, рассылает им, сгенерированные в LogEc рейтинги, которые учитывают самую разнообразную информацию и стимулируют авторов включать в RePEc их новые публикации.

Существует один общий вывод из этого. Чтобы действительно активизировать участие отдельных ученых в этом процессе, необходимо брать в расчет особенности мышления и действий, свойственные ученым. Хотя ученые достаточно искушены в создании "продуктов" (т.е. статей, книг, и т. п.), которые они производят, однако их поведение в процессе оценки качества "продуктов" является достаточно простым. Они очень обеспокоены вниманием других ученых к их статьям (есть ли признаки того, что люди читали их статьи), или информацией, что некоторые рецензенты нашли достойными упоминания о них в своих сборниках, и т.п. Успешно развивающаяся академическая электронная библиотека должна эксплуатировать это поведение для эффективного пополнения своего содержания.

5. Система Соционет

[Система Соционет](#) представляет ряд сервисов, использующих и развивающих идеи RePEc, и предназначена в первую очередь для русскоязычных пользователей. Соционет содержит полную базу данных RePEc (так же, как [IDEAS](#), EconPapers и другие зеркала RePEc), а также включает множество коллекций документов на русском языке (расположенных в архиве RePEc:rus). Сервисы Соционет дают возможность визуализации содержания новых пополнений базы RePEc/Socionet и помогают пользователям справляться с большими количествами новых данных, добавляющихся в RePEc ежедневно. Для этого действуют следующие сервисы и подсистемы:

1. Построение списка коллекций, в которых за последние 24 часа появились новые записи (публикации) и/или были изменения в содержании. Кроме этого ежедневно строится график новых поступлений за последние 30 дней, отражающий количества поступлений, со ссылками к спискам новых поступлений в соответствующие дни.
2. Все оглавления (для архивов, серий, и документов) включают маркировку цветом тех элементов списков, которые изменились сегодня (красный цвет) или в течение последних тридцати дней (розовый цвет).
3. Пользователь может настроить своего личного информационного робота (и-робот) для автоматизированного контроля содержания входных потоков данных.

Последний сервис (и-робот) позволяет делать персонализацию содержания базы данных, исключая из нее неинтересные для пользователя архивы и серии, а также фильтрацию входного потока новых поступлений. Персонализация позволяет пользователю отставить на экране компьютера только небольшое количество коллекций, что упрощает визуальный контроль за появлением в них новых поступлений. Фильтрация использует профиль интересов пользователя и выбирает с заданной регулярностью (например, анализирую новые поступления за последние сутки, неделю, месяц и т.д.) из всего множества новых поступлений только те данные, которые соответствуют интересам пользователя. И-робот, включающий на данный момент времени только эти два сервиса, создает отчеты о результатах работы персонализации и фильтрации. С указанной регулярностью и-робот отправляет отчеты пользователю по электронной почте и/или создает отчеты в виде статических веб-страниц.

В качестве инструментов для решения упомянутой выше проблемы стимулирования пользователей к размещению своих публикаций в RePEс система Соционет также предлагает:

- Открытый Сетевой Архив (<http://socionet.ru/archives/>) для публикации исследовательских материалов в электронной форме по всем основным дисциплинам общественных наук (близкий по возможностям к EconWPA, но имеющий разделы по всем основным разделам общественных наук, а также возможности загрузки документов в архив через электронную почту);
- Личный менеджер ReDIF архива и серий (<http://socionet.ru/socioarm/>) для создания и управления (прямо на компьютере пользователя под операционной системой MS Windows) коллекциями электронных публикаций совместимыми с RePEс. Эта программа позволяет облегчить

создание коллекций, т.к. она отвечает за соблюдение формата ReDIF при введении данных о публикациях в соответствующие формы.

Сервисы системы Соционет являются развитием общей идеи и базовой концепции RePEc. Главный вклад Соционет – перевод системы RePEc от состояния "глобальный электронный каталог" (или "цифровая библиотека") к новому более комплексному качеству, которое можно охарактеризовать как "единое информационное пространство" для сообщества исследователей по общественным наукам. В этом новом случае инфраструктура RePEc должна позволять интегрировать в единую базу данных существенно более широкий набор различных данных, а в идеале – все информационные потоки, которые возникают или могут возникать внутри данного профессионального сообщества. Как средство для преодоления неизбежного в этом случае информационного переполнения участников такого информационного пространства, в Соционет особое внимание уделяется технологиям автоматической обработки (персонализация и фильтрация) потока новых поступлений.

В Соционет созданы достаточно простые и гибкие сервисы, которые позволяют модифицировать структуру базы данных в соответствии с естественным развитием потребностей членов данного сообщества и содержания информационных потоков. Как результат этих работ на начало 2002 г. Соционет включает материалы по 6 дисциплинам общественных наук (экономика, социология, политические науки, демография, закон, и психология). По сравнению с другими зеркалами RePEc (например, IDEAS) Соционет имеет некоторые дополнительные типы темплейтов. В настоящее время это темплейт "схема", с помощью которого создаются коллекции тематических классификаторов и рубрикаторов для научных дисциплин. В ближайшее время будут добавлены темплейты "новости" и "рецензия" и др. Этот процесс расширения спектра используемых темплейтов отражают общую способность системы Соционет к интеграции различных типов информационной деятельности членов сообщества в общую базу данных.

Второй набор идей, который определяет развитие сервисов Соционет, может называться "замыкание циркулирования информации" внутри информационного пространства. Система Соционет должна помочь пользователям не только найти публикации, в которых они нуждаются для их исследований; она должна также стимулировать их делать ссылки на уже существующие в Соционет публикации и помещать результаты своих исследований обратно в данное информационное пространство. Текущее воплощение этих идей включает функционирование на сайте Соционет Открытого Сетевого Архива для помещения в информационное пространство отдельных публикаций. А также инструмент помещения в Соционет целых коллекций публикаций: Персональный Менеджер ReDIF архива и серий (SocioARM). Эти две подсистемы имеют хорошую интеграцию с другими сервисами Соционет, что позволяет автоматически добавлять в их меню новые темплейты, дисциплины и коды классификаторов, как только они появляются в центральной базе данных Соционет.

Третья идея, определяющее развитие сервисов Соционет – "более глубокое использование информационных материалов". Например, исследователи и преподаватели, которые являются одной из главных групп пользователей,

должны иметь больше разнообразных возможностей для использования материалов базы данных Соционет. Например, они должны иметь возможность создавать из них тематические подборки или коллекции публикаций. Такие вторичные коллекции могут быть использованы как учебный материал для студентов, или могут выступать в качестве научного обзора. Вторичные коллекции, собираемые на регулярной основе специалистами по определенной тематике, могут также представлять интерес для других исследователей как полный, пополняемый и возможно аннотированный список литературы. Публикации, размещенные в Соционет, могут служить объектом для рецензий, а также становится основой для научных дискуссий.

Все эти возможности уже создаются и, как ожидается, в скором времени войдут в число сервисов системы Соционет.

6. Выводы

На данный момент система RePEc почти полностью поддерживается за счет усилий добровольцев. Это - значительное организационное достижение. К сожалению, поскольку база данных требует постоянного внимания, то работа для добровольцев ни когда не кончается. Скоро мы переступим десятилетний рубеж, с тех пор как началось развитие базы данных RePEc. Хочется надеяться, много поколений исследователей будут и дальше ее пользователями.

Работа людей, поддерживающих базу данных, влияет на весь мир. Их деятельность по развитию электронных публикаций важна для стран со средним и низким уровнем дохода, где большинство исследовательских организаций слишком бедны, чтобы позволить себе получить доступ к научным коммерческим журналам. Многие публикации, хранящиеся в базе данных RePEc, могут загружаться абсолютно бесплатно. Поскольку вся база данных RePEc является бесплатной, то есть надежда, что этот проект пропагандирует идеальную систему бесплатных потоков информации. RePEc уничтожает различия между информационно-бедными и информационно-богатыми.

Примечания:

а) Обсуждаемые в этой статье результаты получены при финансовой поддержке программы [Electronic Library Programme](#) при Joint Information Systems Committee of the UK Higher Education Funding Councils Education, а также программы Media and Culture Program от московского представительства Фонда Форда ([Ford Foundation](#)).

Об авторах

[Thomas Krichel](#)
[Palmer School](#)
[Long Island University](#)
720, Northern Boulevard

[Sergei I. Parinov](#)
[Institute of Economics and Industrial Engineering](#)
[Siberian Branch of the Russian Academy of Sciences](#)
17, Lavrentiev Prospekt

Brookville, NY 11548-1300

USA

krichel@openlib.org

630090 Novosibirsk

Russia

parinov@ieie.nsc.ru

© Т. Крихель, С. Паринов, 2002.

© Перевод с английского Р. Пузырев, 2002