

Особенности формирования интегрированной системы научных коммуникаций в социальных науках¹

Т. И. Жукова, В. И. Тищенко

Стремительная эволюция и распространение Интернета на все сферы деятельности человека модифицировали процесс генерирования профессиональной общности в научной среде, создав новые модели коммуникативного пространства. Качество и темпы информационного обмена в научном сообществе имеют важнейшее значение, и переизбыток информации в определенном смысле сродни ее недостатку: количество электронных ресурсов, на которых можно найти профессионально важную информацию, постоянно возрастает. Кроме очевидного положительного характера этого естественного процесса развития, можно заметить и ряд отрицательных моментов. С точки зрения ученого-пользователя, потребность в анализе содержания лавинообразно увеличивающихся ресурсов в Интернете создает ряд технических проблем. Когда их количество становится слишком большим, пользователи теряют контроль над информационными потоками и не могут извлекать выгоду из публикуемой новой информации, хотя потенциально она им доступна. Все это приводит к потерям времени и сил для того, чтобы быть в курсе профессиональных новостей. Разработчики информационных ресурсов, в свою очередь, из-за обособленности сайтов и отсутствия общих правил не могут извлекать выгоду от специализации и разделения труда между ними. Это проявляется в создании практически дублирующих друг друга наборов сервисов поиска, навигации и т. п. Поддержание информационной «целостности» электронных ресурсов приводит в этой ситуации к созданию дублирующих друг друга коллекций материалов и т. п.

В настоящее время уже очевидной стала необходимость формирования единого информационного пространства, интегрированного источника научной информации, которое предусматривало бы объединение сведений о разнородных исследовательских информационных ресурсах, обеспечение актуальности этих сведений и широких возможностей для достаточно

¹ Работа выполнена при поддержке РГНФ, грант № 05-03-12313.

точного поиска необходимой информации, поддержку средств научной коммуникации, сервисов, связанных с возможностью оперативного информирования пользователей о нужных им ресурсах и т. п. Решение данной задачи возможно при создании пространственно-образующего механизма, который предполагает наличие двух взаимосвязанных факторов:

- разработчики электронных информационных ресурсов должны сформировать стандартизованные описания своих коллекций и создать техническую возможность для свободного доступа (на чтение) к этим данным через Интернет;
- необходимо, чтобы хотя бы на одном интернет-сервере, была запущена программа, которая по списку интернет-адресов, где располагаются стандартизованные описания ресурсов, собирает и переносит в единую базу данных все добавления и изменения в этих коллекциях.

При одновременной реализации этих двух факторов в открытом доступе в сети Интернет возникает и постоянно актуализируется единая база данных стандартизованных описаний информационных ресурсов, которые фактически принадлежат разным людям или организациям и полностью управляются ими.

История создания открытых пространственно-образующих механизмов для профессиональных сообществ занимает около 10 лет. На данный момент известны два международных проекта, которые определили базовый формат стандартизации и принципы работы механизмов виртуальной интеграции стандартизованных ресурсов в единое информационное пространство научно-образовательного сообщества. Первопроходцем является проект RePEc (Research Papers in Economics), начавший популяризацию и продвижение данных идей в международном масштабе с 1997 г. (только в рамках экономической науки) (<http://repec.org/>). Начиная с 1999 г. в разработку этой проблематики на базе подходов RePEc включился проект Соционет — (<http://www.cir.ru/docs/ips/collections/Socionet.jsp>). В 2000 г. стартовал проект Open Archives Initiative (<http://www.openarchives.org/>), который, основываясь на опыте RePEc и других близких проектов, предложил универсальные и междисциплинарные решения для интеграции информационных ресурсов. В настоящее время в научно-образовательном сообществе действует около сотни самостоятельных пространственно-образующих механизмов, большая часть которых связана с проектом RePEc [1–3].

Эти события свидетельствуют о зарождении нового феномена в жизни традиционно сложившегося научного сообщества, а именно о становлении виртуальных сетевых профессиональных сообществ, которые действуют и взаимодействуют во вновь сложившемся особом коммуникационном пространстве. Вхождение в это пространство в действительности означает формирование не просто совместной, групповой деятельности.

Речь, по сути дела, идет о становлении единой деятельности всех участников этого процесса — процесса становления инновации.

Если же рассмотреть этот процесс с точки зрения единичных членов такого сетевого сообщества, то возникает проблема включения в процесс научной коммуникации в рамках совершенно новой инновационной модели, требующей разработки специального инструмента, который обеспечил бы диалог с хранилищами информационных массивов и служил некоторой технологической средой, позволяющей повысить собственные возможности.

Следует заметить, что, несмотря на глобальность замысла и его реализации, все же каждая предметная область, обладая своей спецификой, накладывает определенные требования на идеологию разработки подобных информационных сред взаимодействия и вносит существенные изменения в их структуру. В данном случае, речь идет о социологических исследованиях.

Одно из важнейших направлений развития современных компьютерных технологий связано с процессом сбора, накопления, хранения и обработки статистических данных, используемых в социологических исследованиях. В социологии, как и в ряде других наук, существует расхождение между объемом собранной научной информации и объемом статистических данных, фактически использованных учеными в социологических исследованиях. Вообще, экспериментальное исследование в социальных науках имеет весьма ограниченные возможности в силу уникальности социальных процессов, невозможности многократного повторения эксперимента, исключительной трудности проверки и сопоставления экспериментальных результатов. В последнее время его проведение вызывает еще больше затруднений в связи с отсутствием финансовых и иных ресурсов. Поэтому существенно выросло значение уже проведенных социологических обследований, наличной информации; имеется в виду возможность повторного использования информации, проведения анализа в контекстах, которые не были включены в первоначальные исследовательские программы. Это направление сегодня составляет отдельную отрасль социологии — вторичные исследования [4].

Все это ставит на повестку дня проблему разработки современных компьютерных процедур централизованного хранения и активного использования существующих массивов данных, а также накопление результатов собственно социологических исследований. Проблема эта на первый взгляд кажется легко решаемой, однако в процессе ее решения возникает целый ряд новых задач, решить которые можно только с применением новых информационных технологий.

В нашей стране процесс формирования единого информационного научного ресурса развивается преимущественно в направлении создания

коммуникационной среды взаимосвязанных распределённых гетерогенных систем в рамках т. н. цифровых или электронных библиотек, которые представляют собой информационные системы нового класса [5].

Во многих странах мира активный интерес к электронным библиотекам как к информационным системам нового класса пробудила объявленная в США осенью 1993 г. национальная программа исследований и разработок электронных библиотек (Digital Libraries Initiative — DLI). Электронная библиотека рассматривалась участниками программы DLI как система, обеспечивающая сообществу пользователей доступ понятным для них образом к большим репозиториям мультимедийной информации и знаний, организованным при отсутствии каких-либо сведений о способах их применения. По мнению директора программы DLI С. Гриффина, электронная библиотека — это «средство обеспечения интеллектуального доступа к распределённым хранилищам информации на основе создания информационных сред, которые возвышают доступ от возможности получения необработанных данных — битов — до восприятия более полных знаний и смысла, содержащихся в цифровых коллекциях». На симпозиуме в Санта Фе (1997), где были сформулированы основные направления исследований для второй фазы программы DLI (DLI-2), подчеркивалось, что не следует отождествлять электронные библиотеки с совокупностью оцифрованных коллекций и инструментария управления ими, а нужно понимать их более широко — как среду, объединяющую коллекции, сервисы и людей для поддержки полного жизненного цикла создания, распространения, использования и сохранения данных, информации и знаний.

Функциональные возможности электронных библиотек варьируются в довольно широком диапазоне. Вообще говоря, предполагается, что они предоставляют пользователю глобальный доступ с помощью различного рода сервисов в среде Веб к коллекциям цифровых информационных ресурсов, которые могут быть распределёнными и в различных аспектах неоднородными. При этом могут обеспечиваться разнообразные возможности их интеграции на техническом и/или на семантическом уровне, развитие средства каталогизации и индексирования. В некоторых электронных библиотеках предусматриваются пользовательские интерфейсы с повышенным уровнем семантики, например, с возможностями семантического поиска требуемых информационных ресурсов, мультязыкового доступа и визуализации данных, средства персонализации пользовательских интерфейсов и т. д. Разнообразный характер могут иметь и их информационные ресурсы — от традиционных библиотечных электронных каталогов до сложных интегрированных коллекций информационных ресурсов, включающих полнотекстовые документы, числовые данные, графические, аудио и видеоресурсы, геоданные, снабженные разнообразными метаданными.

Характерными свойствами электронных библиотек как нового класса информационных систем являются: глобальный доступ к системе в среде Интернета; поддержка метаданных для системы и пользователей; неоднородные и распределенные информационные ресурсы; материализованные и виртуальные коллекции; интеграция информационных ресурсов на различных уровнях; развитые средства каталогизации информационных ресурсов; пользовательские интерфейсы с повышенным уровнем семантики, с возможностями их персонализации; мультязыковые текстовые информационные ресурсы и интерфейсы системы; встроенные сервисы (поиск, визуализация данных, статистический аппарат и др.) [6, 7].

Несмотря на неоспоримые преимущества этого класса информационных систем, им не удастся учесть специфику коммуникаций в сфере социологических исследований, связанной с профессиональными запросами социологов.

Особенность информационного обеспечения в этой науке связана, главным образом, со спецификой социологической информации и используемых исследовательских методов. До недавнего времени социологи пользовались традиционными источниками информации. В основном, это общие универсальные статистические издания, чаще всего материалы соответствующих статистических организаций, а также данные эмпирических исследований, результаты проведения массовых обследований населения, периодические издания по конкретным социальным проблемам. В последнее время вследствие бурного развития информационных технологий появились нетрадиционные источники получения социальной информации, представляющие принципиально новые возможности для прикладной социологии и практически полностью видоизменившие ее облик. Это — информация из компьютерных сетей общего пользования, главным образом, из Всемирной компьютерной сети Internet. Возможность непосредственного интерактивного доступа к накопленным интеллектуальным ресурсам всего развитого научного сообщества ознаменовала собой гигантский шаг, по сути, революцию, в области информационного обеспечения социологических исследований.

Для ученых уже давно стала очевидной непреходящая ценность и уникальность экспериментальных исследований в социальных науках, невозможность многократного повторения эксперимента, исключительная трудность проверки и сопоставления экспериментальных результатов. Кроме того, специфика информационных массивов, полученных в результате эмпирических социологических исследований, состоит еще и в том, что ценность полученных данных не только не убывает по мере их использования другими учеными, но даже возрастает, поскольку эти данные обогащаются новыми обобщениями и выводами, углубляющими характеристики изучаемого объекта.

Поэтому необходимое требование, предъявляемое к информационным ресурсам в социологии — это возможность находить и получать информацию не только о результатах исследований в виде различного рода публикаций, отчетов, тезисов докладов и пр., но и самое главное — иметь доступ к самим массивам данных, созданным исследователями в процессе их работы.

В эмпирической социологии преобладающая часть результатов, их фактическая составляющая добывается усилиями не отдельных исследователей, а усилиями исследовательских коллективов. Соответственно, первоначальной формой материальной фиксации знания является массивы данных исследователей, которые являются генератором гипотез, всей смысловой фактуры анализа, что, естественно выражается во всем документальном материале. Эти массивы могут повторно оказаться объектом Вторичного анализа, имеющего совершенно иные цели, методы и гипотезы, чем первоначальные варианты. Интегрированная информационная система в социологии должна обеспечить возможность наряду с хранящимися в ней массивами социологических данных и справочно-информационными материалами, накапливать сведения, полученные в итоге Вторичного анализа первичных данных, результаты исследований конкретного коллектива ученых. Ценность первичной информации от этого только увеличивается, поскольку одновременно можно ознакомиться и с ее интерпретацией, а в содержание архива привносится субъективная характеристика этого коллектива.

Таким образом, наряду с общепринятыми структурными элементами, составляющими остов информационных систем, в социологии необходимо иметь возможность получать доступ к массивам данных, которые использовались в первичных исследованиях, или же сами в свою очередь были результатом специальных эмпирических исследований. Таким образом, наряду с представлением анализа результатов и выводов исследований в форме публикаций, содержание информационной системы становится формой непосредственного обмена данными и знаниями в социологическом сообществе. Сотрудничество, корпоративная связь могут осуществляться по каналам доступа к информации этих систем, основываясь не только на наличии электронной вычислительной базы, но и на единстве профессионального сообщества в представлениях о методологии и методах эмпирического исследования, на единстве методов формирования и использования массивов данных. Такое единство создает принципиально иной, более развитый тип коммуникации внутри виртуального сообщества, нежели простой обмен информацией, и дает право относиться к подобной инфраструктуре как к интегрированной информационной системе научных коммуникаций в социологии.

По мнению специалистов, метод Вторичного анализа социологических данных на базе такой системы демонстрирует новый тип коммуникации



Рис. 1. Схема прохождения информационных потоков через интегрированную информационную систему

в неограниченном по масштабу интернет-сообществе, который стал возможен именно с применением информационно-коммуникационных технологий. И именно на базе этого типа коммуникации появляется одна из специфических черт нового виртуального научного сообщества. Здесь может и не быть прямой полемики исследователей, но она может разворачиваться за счет различного использования первичных данных, использования разных моделей, разных методологий. Рефлексия этих расхождений принадлежит уже не отдельным участникам дискуссии и не сводится к их частным отношениям. Она есть факт сообщества в целом. В рамках этой коммуникации, связующим элементом которой является единый информационный ресурс, могут возникать отдельные напрямую коммуницирующие сообщества, которых объединяет не только тематическое единство, а и вопросы парадигматического единства и расхождения. Все эти сообщества являются открытыми и предельно демократическими. Но при всем том, в них сохраняется высокий профессиональный уровень, поддерживаемый не только авторитетом лидеров — которые могут в этой ситуации меняться достаточно быстро, и следованием принятой технологии деятельности. Сообщество в целом оказывается уже не столько иерархическим, сколько сетевым [8].

В терминах вышеизложенных оснований схема интегрированного информационного ресурса научных коммуникаций в социологии может быть представлена в виде схемы, изображенной на рис. 1.

Как показано на схеме, система разработана таким образом, чтобы адекватно вписаться в технологическое окружение входных и выходных

потоков данных: используя входные потоки из целого ряда источников, она связывается с многочисленными услугами поддержки, а затем с более широкой средой управления данными. Эта интеграция может быть достигнута благодаря специальному значению, которое придается процессам стандартизации в системе, и благодаря важной роли такого информационного потока, как *метаданные*.

Метаданные — основное средство коммуникации в системе

В сложившейся ситуации, когда сведения о ресурсах чаще всего представлены в виде слабоструктурированного текста, когда поисковые системы осуществляют полнотекстовый поиск нужных данных по запросам в свободной форме, пользователь получает огромное количество «шумовой» информации, среди которой очень трудно выбрать действительно полезные знания. Учитывая это обстоятельство, для представления сведений о ресурсах стали использовать структурное представление, выделять понятие *метаданных*, описывающих содержимое ресурса в виде набора именованных значений, в том числе связей с другими ресурсами. Метаданные используются для автоматизированного анализа содержимого ресурса, построения поисковых индексов и позволяют обеспечить достаточно высокую точность и эффективность поиска разнородной информации.

Очевидно, что глубина структуризации метаданных является одной из важнейших характеристик любой информационной системы. Слишком большая глубина усложняет процессы подготовки метаданных, слишком малая — ограничивает возможности системы. Глубина структуризации метаданных во многом определяется задачами конкретной системы. Во многих случаях при интеграции разнородных информационных ресурсов высокая степень структуризации не требуется, и даже усложняет процессы формирования, поиска и представления данных. В связи с этим одной из первоочередных задач при разработке интегрированной системы представления научных данных является определение минимальной глубины структуризации метаинформации о ресурсах. Предполагая постепенное наращивание функциональности системы, необходимо выработать методику развития схем метаданных — увеличения глубины описания той или иной предметной области, детализации представления информации.

В профессиональных информационных системах используют достаточно большую степень детализации и структуризации данных, что обусловлено необходимостью проведения специальных исследований и поддержки

соответствующих процессов обработки информации. Становясь частью информационной базы, метаданные, обогащая хранящиеся в архиве данные, защищена во временном отношении и остается доступной для последующих исследователей.

Очевидно, что без хороших метаданных эмпирические исследования, основанные на информационных массивах, собранных другими авторами, становятся «игрой предположений». Информационные ресурсы будут представляться конечным пользователям как более или менее бессмысленные собрания статистических конструкций. Обращение к массиву метаданных как бы встраивает пользователя в процесс исследования, делая его непосредственным участником всех этапов и позволяя ему заново формулировать гипотезы и проверять альтернативные возможности на тех же самых данных.

Адекватные метаданные становятся обязательным фактором и начальным условием для расположения статистического массива в информационной базе данных (тем более, если имеется в виду интегрированная система), так как только они могут стать основой для структурирования характеристик исследования. Без них идеалистическая цель — возможность использовать в исследованиях некоторые комбинации из различных наборов данных, так и останется идеалистической.

Сообщество российских исследователей придает самый высокий приоритет содержательным разработкам, связанным с развитием инструментов и математического обеспечения для поддержки массивов метаданных [8–10]. С содержательной точки зрения именно эти массивы являются связующим звеном между пользователями научной информации, разработчиками системы и производителями информации, поскольку:

- пользователи данных очень редко принимают участие в создании массивов данных;
- информация часто используется для других исследовательских целей, чем было предназначено ее производителями;
- данные часто используются много лет спустя после того, как они были созданы;
- академические пользователи часто сравнивают и комбинируют данные из широкого диапазона источников — и по времени, и по месту (пространству).

Из этих четырех причин следует, что расстояние между конечными пользователями статистического материала и его создателями бывает очень значительным, и незадокументированный и неформальный информационный массив может увести вторичного аналитика далеко от целей исследования, не давая гарантии достоверности полученных результатов. Поэтому единственное, на что может положиться исследователь при выборе

необходимого ему информационного поля, это исчерпывающий, легкодоступный и структурированный массив метаданных.

Для социальной науки справедливо даже более сильное утверждение, что информация может стать доступной и достоверной только благодаря метаданным. Они обеспечивают «мосты» между производителями данных и их пользователями и передают информацию, которая является существенной для вторичных аналитиков.

Метаданные также являются отправной точкой для любой системы обнаружения ресурса. Академические пользователи часто ищут массивы данных, соответствующие их задаче, которые могли бы использоваться, чтобы пролить свет на тему их исследования, при доказательстве теории или проверке гипотезы. И хотя основная структура электронных каталогов устроена так, что дает представление об общем содержании источников данных, необходима дополнительная, более точная и научная информация для проведения процесса обнаружения данных, тем более что объем массивов объединенных каталогов данных умопомрачительно велик, и, осуществляя поиск по нескольким общим параметрам, можно получить лавину информации, необъятную с точки зрения выбора одного единственного адекватного массива.

Существует еще один фактор, который нельзя не упомянуть, это потенциальная роль метаданных в качестве некоего моста между данными, пользователями и интеллектуальной продукцией пользователей. С помощью ссылок и гиперсвязей, отсылающих к сообщениям и научным исследованиям, написанным на основе данных, а также с помощью отдельных механизмов, связывающих пользователей с конкретными организациями, ответственными за научный результат, метаданные становятся важным средством коммуникации и узлом связи в процессе накопления знания.

Слишком часто набор данных или публикация таблицы становится конечной точкой в процессе производства и анализа статистической информации. Поэтому включение вторичных исследований в этот процесс, ставший возможным благодаря более тщательной разработке массивов метаданных, кажется весьма значительным и разумным достижением.

Технологии и стандарты метаданных в социологии

Для того чтобы в машинном виде воплотить все разумные замыслы содержательных разработчиков, для достижения статистической и исследовательской общности, а также для гарантирования единообразия во всех элементах системы метаданных, были применены следующие стандарты и технологии.

DDI — стандарт метаданных

На протяжении нескольких последних лет было предпринято много инициатив в направлении создания унифицированных стандартов метаданных. Однако ни одна из них не достигла того уровня приемлемости, который необходим, чтобы стать единственным эталоном стандарта. Большинство социальных информационных систем документировало свои информационные запасы согласно стандартному описанию исследования, принятому еще в середине 1970-х гг. К сожалению, с тех пор развилось много локальных «диалектов» этого стандарта, приведших к тому, что запасы метаданных были адаптированы к требованиям различных пользователей и поисковых систем. Следствием этого немедленно стало существенное снижение уровня стандартизации и потому невозможность осуществления интегративных процессов.

Чтобы улучшить эту ситуацию, в 1995 г. под эгидой Межуниверситетского консорциума по политическим и социальным исследованиям был создан новый международный комитет (Инициатива документации данных — the Data Documentation Initiative (DDI)), членами которого стали специалисты социальных и гуманитарных библиотек США, Канады, Европы, а также из большинства организаций — производителей и поставщиков информации в этих странах, таких как Бюро переписи США, Бюро трудовой статистики США, Комитет по статистике Канады и т. п.

Первоначально основная цель DDI состояла в том, чтобы заменить устаревший формат описания OSIRIS на более современный формат. Однако, по мере осуществления проекта, амбиции разработчиков стали намного большими. Стало возможным рассматривать подобный формат как потенциально структурирующий массив метаданных и тем самым организующий весь процесс накопления, распределения и анализа данных внутри социальных и поведенческих наук. Дополнив этот процесс разработкой специального языка программирования XML DDI, авторы обеспечили своего рода некоторый «клей», который позволяет объединить различные стадии этого процесса.

В 1999 г. согласованный проект разработанного DDI-стандарта был подвергнут интенсивной экспертизе тринадцатью организациями по обе стороны Атлантики. Эта деятельность подразумевала производство структурированных массивов метаданных для разных типов информации, разработку программного обеспечения для введения новых стандартов для уже существующих источников, сравнение DDI-стандарта с другими стандартами, функционирующими в системе. В конечном итоге, этот стандарт был одобрен участниками комитета и принят к использованию совместно с его подробным описанием в руководстве для пользователей.

Стандарт DDI декларирует обязательные элементы, из которых должен состоять документ, сопровождающий информацию в системе, харак-

теристики этих элементов, а также их функциональные взаимосвязи между собой.

Иерархически элементы организованы в виде структуры логического дерева. Основными разделами стандарта являются следующие:

- 1) **описание документа**, которое характеризует не только сам документ с метаданными, но и источники, на основании которых метаданные были разработаны (этот раздел может таким образом рассматриваться как своего рода метаданные для метаданных);
- 2) **описание исследования**, которое содержит подробную информацию относительно исследования и сбора данных (содержание, методы сбора информации, обработка данных, источники, условия доступа и т. д.);
- 3) **описание файла**, которое описывает каждый отдельный файл данных (форматы, размерности, пакеты обработки информации, информация о пропущенных данных и т. п.);
- 4) **описание переменных**, которое описывает каждую отдельную переменную в файле данных (формат, название переменной и ее метки, определения и т. п.);
- 5) другие **связанные материалы**, которые могут содержать ссылки на сообщения, публикации и другую машиночитаемую документацию, которая может пригодиться пользователям данных.

С более подробной информацией об этом стандарте можно ознакомиться в нескольких источниках [11, 12].

XML — стандарт синтаксиса

Наиболее значительным развитием семантики Всемирной сети WWW, сети, которая обеспечивает специализированные структуры и информацию для различных сообществ, является создание XML-синтаксиса (eXtensible Markup Language). Этот язык дает возможность разработчику создавать свои собственные шаблоны и формализованные структуры, что в свою очередь позволяет ученым работать только в содержательном контексте, не беспокоясь относительно стандартизации и возможности взаимодействия. За несколько последних лет потенциал качественного совершенствования существующих веб-технологий заметно исчерпался из-за сдерживающего влияния, которое стали оказывать слабые стороны языка HTML — основного выразительного и структурообразующего средства гипермедийных информационных ресурсов Web. Эти слабые стороны заключаются, прежде всего, в закрытом характере языка HTML, в невозможности расширения его функциональности самим пользователем, а также в отсутствии поддержки метаданных, которые могли бы описывать структурные и семантические свойства документов.

Магистральное направление развития веб-технологий в последнее время связано с новым расширяемым языком разметки XML.

В 1996 г. рабочая группа под эгидой Консорциума Всемирной паутины выпустила первый рабочий проект языка XML, который был предназначен стать господствующей технологией для включения исследователей в область широких функциональных возможностей, которые предоставляет Интернет. Язык XML представляет собой удобное для реализации подмножество стандартизованного языка SGM, который в отличие от своего предшественника — языка HTML, является «открытым» языком. В нем не фиксируется множество ярлыков, используемых для разметки документов. Создатель XML-документа или набора документов может сам конструировать необходимое ему множество, определяя его с помощью метасредств языка XML — набора синтаксических конструкций, называемых декларациями разметки (Markup Declaration). Язык XML позволяет описывать логическую структуру документов в терминах составляющих их элементов; такие описания — Document Type Definition (DTD) могут поддерживаться в глобальной среде независимо от экземпляров XML-документов и могут унифицироваться в интересах различных сообществ пользователей. Благодаря этому можно создавать новые языки разметки, адекватные потребностям различных сфер применения.

Предполагается, что с помощью языка XML и стандартной структуры метаданных DDI представятся новые возможности осуществления сравнительного и другого научного исследования, основанного на привлечении многочисленных наборов данных, что является возрастающей тенденцией в социальных и поведенческих науках. Пользователи смогут приспособлять сложные наборы данных к статистическим пакетам SPSS, SAS, STATA через DDI-кодировку. Информация, доступная программному обеспечению через DDI, будет намного более богатая и полная, чем когда она обеспечивалась специальными средствами этих статистических пакетов, к тому же сам процесс импортирования данных в диалоговое сетевое использование будет намного ускорен.

Таким образом, становится очевидным, что разработка XML-языка и DDI-стандарта является полезным приложением не только производителям данных и организаторам информационных систем, но также и пользователям данных, которые получают в распоряжение новые инструменты, чтобы более успешно и эффективно осуществлять свою деятельность.

Все вместе вышеописанные технологии и стандарты дают возможность разработчикам не заботиться о технических характеристиках в своей работе, а сосредоточиться на воплощении в жизнь требований как пользователей данных, так и их распространителей.

Осознание возможности создания интегрированного информационного ресурса научных коммуникаций как некоторого действительного явления,

как реального феномена социологической науки, оценка продуктивности перспектив его массового использования представляется вполне актуальной задачей. Однако сейчас, на современном этапе, кажется очень важным разработать и принять единую стандартную методику разработки таких систем и объединить ее с содержанием, методологическими основами и методическими приемами проведения конкретного социологического исследования.

Литература

1. Жукова Т. И., Сазонов Б. В., Тищенко В. И. Подходы к созданию единой сетевой инфраструктуры научного сообщества // Социальные коммуникации в новой информационной среде. Труды ИСА РАН. Т. 26. Ч. 2. М.: Издательство ЛКИ/URSS, 2007.
2. Паринов С. И., Ляпунов В. М., Пузырев П. Л. Система Соционет как платформа для разработки научных информационных ресурсов и онлайн-сервисов // Электронные библиотеки. 2003. Т. 6. Вып. 1.
3. Krichel T., Паринов С. И. База данных RePec и ее российский партнер система Соционет // Электронные библиотеки. 2002. Т. 5. Вып. 2. <http://www.elbib.ru/journal/2002/200202/KP/KP.ru.html>
4. Социальная информатика. Основания, методы, перспективы. М.: URSS, 2003. Гл. 8. Архивы формализованных данных социальных наук.
5. Коголовский М. Р., Новиков Б. А. Электронные библиотеки — новый класс информационных систем // Программирование. Российская академия наук. 2000. № 3.
6. Еришова Т. В., Хохлов Ю. Е. Межведомственная программа «Российские электронные библиотеки»: подходы и перспективы // Электронные библиотеки. 1999. Т. 2. Вып. 2. <http://www.elbib>
7. Сютюренко О. В., Хохлов Ю. Е. Распределенные библиотечные сети и электронные библиотеки в России // Электронные библиотеки. 2000. Т. 3. Вып. 5. <http://www.elbib.ru/journal/2000/200005/HS/hs.ru.html>
8. Алексеев А. Н., Созыкин А. В., Масич Г. Ф., Бездушный А. Н. Подсистема проведения конференций и ее метаданные // Электронный журнал, посвященный созданию и использованию электронных библиотек. Т. 7. Вып. 2. М.: Ин-т развития информационного общества, 2004.
9. Бездушный А. Н., Вежневцев А. А., Серебряков В. А., Шкотин А. В. Метаданные: определение и использование. Четвертая ежегодная международная конференция «Электронные изображения и визуальные искусства». М., 2001.
10. Хохлов Ю. Е., Арнаутов С. А. Обзор форматов метаданных. Ин-т развития информационного общества. http://www.elbib.ru/index.phtml?page=elbib/rus/methodology/md_rev
11. Жукова Т. И. Опыт разработки информационных систем для социологических исследований // Вестник РГНФ. 1998. № 2.
12. Bethlehem J., Kent J., Willeboordse A. On the use of metadata in Statistical data processing // Working Paper. UN/ECE Work Session on Statistical Metadata, Geneva, Switzerland, 22–24 September 1999. № 23.