

Developing the Digital World of Government Information and Official Publications: A View from the United States

Thomas F. Lahr (Presenter)

Robin Haun-Mohamed & Eleanor G. Frierson

**WORLD LIBRARY AND INFORMATION CONGRESS: 71ST IFLA
GENERAL CONFERENCE AND COUNCIL**

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

<http://www.ifla.org/IV/ifla71/Programme.htm>

Развитие цифрового мира правительственной информации и официальных публикаций: взгляд из Соединенных Штатов

Резюме:

Американские правительственные организации создают новые цифровые правительственные информационные службы, многие через сотрудничество и партнерские отношения как внутри, так и вне правительства. Крупные инициативы по оцифровыванию и цифровому сохранению предложены или находятся в разработке. Поддержка на самых высоких уровнях создает возможности для инноваций и стимул для изменения в методах, которыми создаются и становятся доступными правительственная информация и официальные документы. Инновационные системы и программы правительства США обеспечивают доступ к информации правительства Соединенных Штатов, которого раньше не было вообще, либо поиск информации занимал много времени.

National Biological Information Infrastructure (NBII) – программа Геологического обозрения США, которая обеспечивает расширенный доступ к данным и информации по вопросу национальных биологических ресурсов.

Science.gov. – межведомственный (включая GPO и USGS/NBII) межсетевой (шлюзовый?) портал, который обеспечивает дополнительный доступ к отобранной авторитетной правительственной информации по науке и технологиям.

В моем выступлении будут приведены подробности работы американских правительственных организаций по созданию новых служб по цифровой правительственной информации, часто с применением сотрудничества и партнерства внутри и вне правительства. Я хочу обсудить проекты, касающиеся конкретно Бюро по печати правительства США (GPO), Национальной биологической информационной структуры (NBI) и межорганизационной информационной службы –Science.gov.

Бюро по печати правительства США (GPO), - организация правительства США, существующая с 1861 года, и, являясь партнером Science.gov, создает и распространяет материальные и электронные документы правительства США.

У GPO три основные миссии:

- Обеспечивать все организации правительства США издательскими и печатными службами на базе возмещения расходов
- Обеспечивать постоянный доступ к публикациям и другой информации правительства США. Эта задача выполнена в сотрудничестве приблизительно с 1250 американскими библиотеками в рамках Федеральной программы депозитарных библиотек (FDLP).
- Обеспечивать доступ широкой публики к копиям печатных и электронных документов и другой информации правительства США.

Федеральная программа депозитарных библиотек

FDLP обеспечивает бесплатный постоянный публичный доступ к документам всех трех ветвей американского правительства. Общественный доступ к публикациям американского правительства начинается с 1813 года, но FDLP официально функционирует с 1895 года, когда она (FDLP) была передана в Бюро по печати правительства США. В FDLP входит более 1250 депозитарных библиотек по всем Соединенным Штатам и их территориям.

- 53 являются региональными (100% отбор)
- Прочие – избирательны (отбирают публикации, больше всего отвечающие информационным требованиям данного сообщества).

Проект Бюро по печати правительства США по оцифровыванию унаследованных материалов

GPO планирует оцифровывать полное собрание материальных документов правительства США.

GPO также определяет и приобретает цифровые публикации через печатный процесс и через Веб-сайты.

Вместе обе инициативы создадут обширную коллекцию настоящих, прошлых и будущих публикаций правительства США.

С самого начала действия Федеральной программы депозитарных библиотек американское общество имело бесплатный доступ ко многим информационным документам федерального правительства. До начала 1990-х эта информация имелась, главным образом, в печатном виде, в виде микрофишей, карт и в других материальных форматах. Поскольку GPO перешло, преимущественно на электронную версию FDLР, были предприняты шаги для идентификации, каталогизации и архивирования он-лайнных электронных ресурсов.

Легкость, с которой можно получить доступ к он-лайнным электронным ресурсам через всемирную паутину, привела к желанию получить дополнительные ресурсы в электронном формате. Эти просьбы подтолкнули многих в депозитариях и других информационных сообществах начать оцифровывание материалов унаследованных собраний.

GPO, в сотрудничестве с депозитарными библиотеками и другими заинтересованными организациями информационного сообщества, разрабатывает план по оцифровыванию материала унаследованного собрания. Поскольку правительство США обязано обеспечить постоянный публичный доступ к информационному продукту, как к настоящему, так и к историческому, то целью плана является сделать так, чтобы цифровая коллекция была в наличии в общественной сфере при бесплатном постоянном публичном доступе через FDLР.

Сканированные образы и ассоциативные метаданные должны производиться на качественном уровне, соответствующем поддержанию сохранности, равно как и будущие (повторения) производной продукции. Задачей проекта является оцифровать, по меньшей мере, 70% исторической коллекции к декабрю 2007 года и GPO рассматривает способы для успешного достижения этой цели.

Сфера проекта: Что следует оцифровать и распространить электронным способом?

- Публикации правительства США – информационные материалы, которые публикуются как отдельные документы за счет правительства, или, как положено по закону (44 U.S.C.).
- Национальная библиография публикаций правительства США
- Национальное собрание публикаций правительства США

GPO оцифровывает неклассифицированные публикации правительства США. Правительственная публикация, в соответствии с определением в законе (44U.S. Code), -

это «информационный материал, который публикуется как отдельный документ за счет правительства, либо как положено по закону».

В качестве выполнения своей миссии по обеспечению доступа к публикациям правительства США, GPO в течение 100 лет осуществляло библиографический контроль над публикациями правительства США.

Все публикации в Национальной библиографии являются частью Национального собрания публикаций правительства США. Национальная коллекция, находящаяся сейчас в разработке, будет состоять из множества коллекций реальных и цифровых публикаций в закрытых и открытых архивах (in light and dark archives), размещенных на многочисленных сайтах и работающих с разными партнерами внутри и вне правительства США.

Сфера проекта: настоящие и унаследованные публикации

В сферу федеральных публикаций для оцифровывания входят:

- Унаследованные, или исторические, публикации
- Настоящие бумажные публикации, которых нет в электронном формате
- Прочие реальные материалы, включая мультимедийные форматы

Унаследованные, или исторические, публикации

GPO в сотрудничестве с депозитарными библиотеками и другими заинтересованными организациями в информационном сообществе разрабатывает план по оцифровыванию материалов унаследованной коллекции. Приоритетом при оцифровывании для этой коллекции является обеспечение населения онлайн-версиями наиболее важных публикаций правительства США, включая Свод законов США, отчеты Конгресса, Федеральный реестр, Свод федеральных постановлений, законопроекты Конгресса и т.д., с тем, чтобы увеличить ретроспективный охват этих популярных баз данных в GPO Access.

Другие публикации, которые представляют особый интерес для населения или подвергаются опасности из-за хрупкости бумаги, также подлежат первоочередному оцифровыванию. Депозитарные библиотеки обследовались в 2004 году, в результате чего был составлен список конкретных наименований, которые библиотеки хотели бы иметь в оцифрованном виде в первую очередь.

Настоящие публикации

Большая часть публикаций, которые GPO собирается сканировать, будет представлять собой часть исторической коллекции; однако, GPO может также оцифровать некоторые настоящие публикации, когда нет возможности получить электронный формат от выпускающей федеральной организации.

Службы GPO по цифровой перекодировке, подразделение, ответственное за цифровые службы, будет также заниматься конвертированием широкого спектра других материалов, включая микроформаты, карты, мультимедийные и реальные электронные публикации. GPO продолжает работать с информационным сообществом в направлении разработки деталей, необходимых для конверсии этих специализированных форматов.

Аутентификация (Установление подлинности)

Аутентичное содержание: Содержание, которое подтверждено GPO, должно быть полным и неизменным по сравнению с версией, одобренной или опубликованной издающей организацией.

Официальное содержание: Содержание, попадающее в рамки Национального собрания публикаций правительства США и одобренное официальным источником, или полученное из официальных источников в соответствии с принятой программной политикой и процедурами.

Аутентификация

Как часть своей основной задачи по обеспечению постоянного публичного доступа к официальным и аутентичным публикациям правительства США, GPO сейчас внедряет электронные подписи в Инфраструктуре открытого ключа (PKI) с тем, чтобы обеспечить аутентичность электронного содержания, включая конвертированное содержание.

Официальное содержание – это содержание, попадающее под рамки Национального собрания публикаций правительства США, и одобренное, или полученное из официальных источников в соответствии с принятой программной политикой и процедурами.

Аутентичное содержание – содержание, подтвержденное GPO, должно быть полным и неизменным относительно версии, которая одобрена или опубликована издающей организацией.

Инициативы GPO по аутентификации дают им возможность удостоверить документы на аутентичность и официальный статус документа и давать возможность пользователям быть уверенными в отсутствии изменений в файле, поскольку GPO подтвердила подлинность этих документов. В процессе сертификации создается знак сохранности (integrity mark), который используется при передаче аутентификационной информации пользователям.

Работа с версиями

Работа с версиями – это приобретение, библиографический контроль, доступ, сохранение и аутентификация всех «уникальных особенностей» публикаций. Цифровое

содержание может изменяться на самой обычной базе и очень легко изменить содержание без ведома пользователя.

Работа с версиями может быть завершена оценкой различной атрибутики документов (а именно, структура, содержание и формат), созданием метаданных об этих атрибутах, их мониторингом, модернизацией метаданных.

GPO сотрудничает с другими для того, чтобы определить необходимые механизмы приобретения и сохранения новых версий различных категорий документов. Например, версию постоянно модернизируемой цифровой карты можно получать на периодической основе.

Доступ к цифровым публикациям правительства США

Доступ к цифровым публикациям будет осуществляться через GPO Access и сеть федеральных депозитарных библиотек.

GPO Access

<http://www.gpoaccess.gov>

Федеральная программа депозитарных библиотек

<http://www.gpoaccess.gov/fdlp.html>

Копии, сохраняемые в цифровом виде и, связанные с ними, метаданные будут сохраняться в электронном архиве GPO. Бесплатный публичный доступ к содержанию будет возможен через производные файлы, имеющиеся на GPO Access, что обеспечивает свободный электронный доступ к важному информационному продукту правительства США.

Федеральные депозитарные библиотеки предоставляют доступ к GPO Access для тех, кто без Интернета и без экспертной справочной помощи находят и используют эти публикации.

Администрация Национального архива США (NARA) признает GPO официальным филиалом архива для материалов электронного содержания на GPO Access. NARA признает легальное хранение публикаций правительства США на GPO Access в качестве официальных архивов Соединенных Штатов, так, как GPO осуществляет физическое хранение и отвечает за постоянный публичный доступ и сохранность записей.

Настоящие (текущие) проекты GPO в связи с планами оцифровывания

- Список проектов по оцифровыванию публикаций правительства США
- Сетевой сбор (Web Harvesting)
- Будущая цифровая система

Для того, чтобы увеличить доступ к цифровым ресурсам и предоставить информацию о GPO и других цифровых проектах, находящихся в разработке, GPO создает всесторонний обширный список цифровых проектов по публикациям правительства США.

GPO продолжает работу с пилотным проектом по автоматическому сетевому сбору (?) на сайте по защите окружающей среды, чтобы определить ускользнувшие электронные публикации.

GPO планирует получать и содержать цифровые материалы, которые уже возникли в цифровом виде, в одной системе. Ожидается, что эта система, известная как Цифровая система будущего, в настоящее время находящаяся в разработке, основана на модели Открытой архивной информационной системы. Вся эта система при включении в себя более мелких взаимодействующих между собой, в том числе и новой интегрированной библиотечной системы GPO, будет полностью запущена в 2007 году.

Другая программа правительства США, Национальная информационная инфраструктура по биологии (NBII) www.nbio.gov была создана с целью обеспечения базы, которая давала бы возможность самой разнообразной аудитории получать готовый доступ и эффективно использовать информацию по всему биологическому разнообразию и экосистемам. Программа NBII – это обширное предприятие, созданное с целью обеспечить более широкий доступ к данным и информации о национальных биологических ресурсах. В сотрудничестве с Геологическим обозрением США (USGS) эта система предоставляет данные по естественным наукам и информацию, которая надежна в научном отношении. NBII делает возможным собирать и систематизировать информацию по биологии у партнеров, а затем предложить этот продукт и услуги этим же партнерам и другим исследователям.

Конечные пользователи NBII – это те, кто работает с биологическими ресурсами, изучает их, или пользуется ими. Обычно, эти пользователи, как из государственного, так и из частного сектора – ученые, плановики, ответственные работники из промышленности, из федеральных, государственных и местных правительственных организаций, межправительственных организаций, преподаватели и студенты, а также другие граждане. Конечные пользователи NBII обращаются в Программу с самыми разными информационными запросами. Девиз Программы – «Создание знаний через партнерство» – отражает её цель - объединение интеллектуального капитала частного сектора с обязанностью правительства удовлетворять информационные потребности руководящих работников из области естественных ресурсов страны. Данные и информация, предоставляемые этим группам

включает высококачественные базы данных по биологии, информационный продукт и услуги, а также разнообразие аналитических инструментов. Партнеры NBII работают также с новыми стандартами, инструментами и технологиями, что облегчает пользователям NBII находить, объединять и использовать информацию о биологических ресурсах. Партнерская структура NBII охватывает ресурсы более 200 организаций.

Одним из ключевых компонентов NBII является структура на базе «сетевого узла» (node), разработанная для того, чтобы обеспечить широкое партнерство и поступление информации из всех слоев общества. Установка этих узлов позволяет NBII предоставлять пользователям быстрый доступ к данным и информации о биологических ресурсах с национальным, и, более того, с международным охватом вопросов по биоразнообразию и окружающей среде. В узлах NBII концентрируются ключевые элементы обеспечения информацией и услугами, предусмотренными NBII.

Узлы NBII бывают трех типов:

Региональные – Имеют географическую ориентацию. При таком подходе местные данные, коллекторы данных и их владельцы - все включены в этот процесс. Это также позволяет заинтересованным партнерским группам создавать активно работающие коалиции при обращении к проблемам по биологии в пределах своих регионов. Например:

- Предложение информации по оценке стратегии и практики, связанных с попыткой остановить сокращение поголовья лосося в северо-западной части Тихого океана, равно, как и по управлению региональной лесной экосистемой.
- Поддержка информационных систем, касающихся биоразнообразия и оценки водоразделов в Калифорнии, на тихоокеанском побережье, а также экосистем пустынь юго-запада.

Тематические – Работа по отдельным биологическим проблемам (например, сохранение определенных видов птиц) путем обеспечения поддержки и инфраструктуры по данной тематике. Такие проблемы часто затрагивают различные географические территории и имеют национальное значение. Например:

- Агрессивные виды – это проблема номер один двадцать первого века, она представляет собой существенную экономическую проблему для нашей нации. Этот узел предоставляет доступ через единый портал к огромному спектру информации по агрессивным видам по всей стране.

Инфраструктура – Акцент делается на разнообразные области инфраструктуры, такие, как сети и их безопасность, стандарты, словари, аппаратно- программное обеспечение, инструментарий, геопространственное применение и инструменты для

поиска – получения и анализа данных. Деятельность NВII направлена на то, чтобы мы лучше понимали и сохраняли жизненные ресурсы нашей нации, пользуясь обширной информацией с описанием этих ресурсов. Кроме того, NВII обязана гарантировать, что данные по биоразнообразию будут рассортированы и доступны для тех, кто принимает решения об использовании естественных ресурсов и их сохранности. По мере изменения технологий, NВII стремится решить эту проблему. Но, на данной презентации представлены лишь некоторые успехи NВII, сама же Программа нацелена в будущее. Мы приглашаем все заинтересованные стороны присоединиться к нам с тем, чтобы помочь NВII стать еще более эффективным инструментом для облегчения принятия решений по проблемам окружающей среды.

Теперь я хотел бы обратиться к обзору межорганизационных усилий Соединенных Штатов, которые объединяют данные и информацию из предыдущих примеров с другой научной информацией, Science.gov.

Теперь вам известно, как две американских правительственных организации создают новые службы в области цифровой правительственной информации. Обе эти организации участвуют также в Science.gov, которая предоставляет более ценный и широкий доступ к информации, получение которой ранее занимало много времени, либо её трудно было найти. Этот портал дает возможность поиска по 30 базам данных и более, чем 1700 научным Веб-страницам с доступом более, чем к 47 млн. страниц правительственной научной информации. Science.gov была запущена в декабре 2002 года, обеспечивая, в первое время, широкий доступ и единый поиск в огромном количестве правительственных документов по научной и технической информации, как это указано в Science.gov. Здесь есть ключевые варианты для любых пользователей, при этом поиск будет происходить одновременно в веб-сайтах и базах данных. Science.gov имеет более 1700 веб-сайтов, отобранных и каталогизированных руководителями организаций. Сайты сформированы для поиска по одной и более из 12 тем и 175 подтем. Свободный поиск отобранных веб-сайтов базируется на текстах данного веб-сайта и связанных с ним сайтов (до одного уровня). Science .gov дает также пользователю возможность работать с отобранными сайтами по теме или по каким-либо отдельным признакам. Там также, по мере появления, высвечиваются какие-то новые особенности, такие, например, как новый “Alerts”. Science.gov позволяет пользователям производить поиск как поверхностный, так и глубокий, когда обычный поиск ничего не дает. Эта программа дает возможность производить одновременный поиск по 30 базам данных, что вообще недоступно для поисковых возможностей Веб. Этот специальный невидимый веб-поиск был разработан для Science.gov. Существует единый поисковый интерфейс для веб-сайтов и баз данных.

Доступ ко всей информации свободный, никакой регистрации не требуется. Альянс Science.gov – это группа американских правительственных организаций, созданная на добровольной основе с научными задачами. Эти организации участвуют в разработке, эксплуатации и укреплении Science.gov. В Альянс входят важнейшие американские научные организации и министерства. Административная поддержка и координация осуществляется группой CENDI (www.cendi.gov), межведомственной рабочей группой старших менеджеров по научной и технической информации правительства США. Организации содействуют в разработке программного обеспечения, работе с серверами, оказании реальных услуг, проверке использования денежных средств фондов и занимаются каталогизацией. Каждая организация отбирает самую качественную научную информацию для включения в Science.gov. Альянс сопредседательствует в качестве представителя в First.Gov, www.firstgov.

Мы определили ряд проблем и многие из них естественны для данного предприятия – широта федеральных исследований и разработок, масштаб аудитории, заинтересованной в этой информации. Эти проблемы обозначены в решениях по поводу того, что должен отбирать Унифицированный указатель ресурсов (URL) и каким образом должна быть выражена таксономия. Мы продолжаем работать над тем, чтобы предоставить информацию еще более широкой аудитории. Разумеется, главной задачей было соединить ресурсы из разных ведомств в один целостный продукт для потребителей.

К концу 2005 года Версия 3 программы Science.gov будет готова, что обеспечит увеличение точности поиска и MetaRank – улучшенную систему ранжирования с использованием метаданных и логического (булевского) поиска. Версия 4, которая планируется на 2006 год и далее, увеличивает соответствие адресного ранжирования (Deer Rank) и ранжирование по соответствию полного текста для точного поиска. Так как мы занимаемся техническими усовершенствованиями, то нам необходимо производить оценку, насколько удовлетворяются запросы потребителя и продолжать совершенствовать и улучшать всю схему и поисковые возможности. Со времени Версии 1, 2002 года, мы в обычном порядке добавляли содержание и базы данных и проверяли таксономию и связывали все это с другими порталами First.gov, такими, как Kids.gov. Мы будем работать над увеличением числа членов Альянса и развивать поддержку и соответствующие ресурсы. Очень важно, что мы будем продолжать то, что мы делаем лучше всего – сотрудничество и (разработка) инноваций. Мы услышали сегодня о трех инновационных системах и программах правительства США, которые обеспечивают более ценный дополнительный доступ к правительственной информации, которая прежде была недоступна или её поиск занимал много времени.

