

Цифровое хранение и глубокая инфраструктура*

Стюарт ГРЕНГЕР
(Stewart GRANGER)

Университет г. Лидс, США

Делается попытка объяснить понятие "глубокая инфраструктура" в применении к цифровому хранению. Объясняются четыре основных аспекта глубокой инфраструктуры: организационный, правовой, культурный и технический. Для создания глубокой инфраструктуры предлагается использовать двойной подход: прагматический и стратегический.

ВВЕДЕНИЕ: ГЛУБОКАЯ ИНФРАСТРУКТУРА В ОТЧЕТЕ СПЕЦИАЛЬНОЙ КОМИССИИ

В отчете Специальной комиссии по цифровому хранению [1] словосочетание "глубокая инфраструктура" используется несколько раз. Это словосочетание сразу запомнилось мне, так как я уверен, что оно так удачно и так чрезвычайно внушительно, как ни одно другое. Цель этой статьи — попытаться развить идею, которая может помочь объяснить понятие "глубокая инфраструктура" [2].

Несмотря на то, что этот отчет создает ощущение того, насколько трудно определить и сделать осязаемым характер проблем, которые надо решить, я думаю, уместно сказать, что отчет Специальной комиссии ясно определил только два основных аспекта, которые эта комиссия подразумевает под глубокой инфраструктурой: необходимость сертифицированных архивов (и предполагаемый процесс сертификации) и понятие "ошибкоустойчивого механизма". Что касается первого, то мы недавно получили превосходный отчет RLG/OCLC [3]. Однако последний аспект — ошибкоустойчивый механизм — это один из аспектов отчета, который меня беспокоит [4]. Картина глубокой инфраструктуры представлена на рис. 1.

Диаграмма на рис. 1 требует некоторого пояснения. Очевидно, что она предполагает наличие картины, которая желательна, но которой в настоящее время не существует. Далее в этой статье я объясню, что я имею в виду под четырьмя основными аспектами глубокой инфраструктуры, обозначенными на рис. 1. Стрелки предназначены для представления идеи относительно того, что технические решения для проблем цифрового хранения должны удовлетворять требованиям пользователей, а именно тех, которые нуждаются в хранении цифрового содержания или имеют долговременный доступ к нему, т. е. всех нас. Это может быть самой привлекательной частью программирования — определяй требования, а затем проектируй и создавай систему, удовлетворяющую этим требованиям (привлекательной, но часто не легкой, чтобы ей следовать!). Однако я хочу доказать то, что ситуация, с которой мы постоянно сталкиваемся, не является реальной. В настоящее время технологии часто проектируются и разрабатываются больше для выгоды поставщиков, чем для пользователей, а люди, связанные с цифровым хранением, как предполагается, должны преодолеть тяжелые испытания, предлагаемые этими технологиями. Этот спорный взгляд будет встречаться позже в этой статье.

Я надеюсь, что диаграмма на рис. 1 окажется полезной, но, как и все такие попытки, она несомненно имеет свои недостатки. Нигде на этой диаграмме нет слов "экономика" или "стоимость", но, конечно, поскольку она раскрывает многие из центральных понятий, они должны, очевидно, иметь многие стоимостные и экономические значения. Более того, очевидно, что эти понятия предполагают десятилетия работы, прежде чем они полностью реализуются. Одно последнее предостережение: в этой статье я попытаюсь описать "большую картину", но я не утверждаю, что эта картина полная.

Глубокая инфраструктура

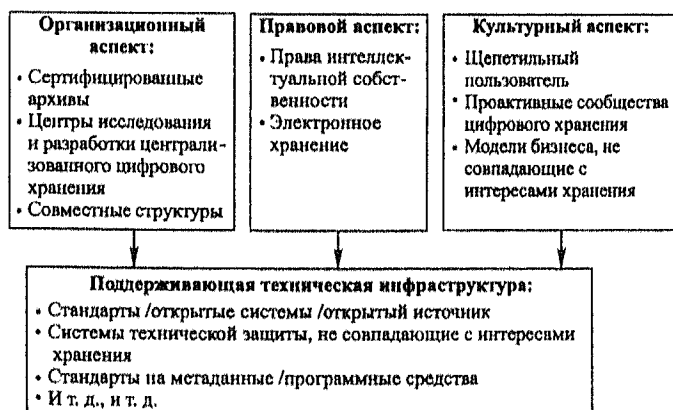


Рис. 1. Представление глубокой инфраструктуры

ОРГАНИЗАЦИОННЫЕ АСПЕКТЫ ИНФРАСТРУКТУРЫ ЦИФРОВОГО ХРАНЕНИЯ

Сертифицированные архивы

В 1994 г. Комиссия по хранению и доступу (Commission on Preservation and Access — CPA) и Специальная комиссия Группы научных библиотек (Research Libraries Group —

*Перевод Granger S. Digital preservation and deep infrastructure // D-Lib Magazine. — 2002. — Vol. 8, No. 2.

RLG) [2] начали вместе описывать и изучать характер надежного хранилища для цифровых материалов. Главные результаты отчета CPA/RLG включали следующие пункты:

- Долговременное хранение цифровой информации в масштабе, соответствующем запросам будущих исследователей и ученых, потребует глубокой инфраструктуры, способной поддерживать распределенную систему цифровых архивов.
- Решающий компонент инфраструктуры цифрового архивирования — это существование достаточного числа надежных организаций, способных осуществлять хранение, перемещение и обеспечение доступа к цифровым массивам.
- Процесс сертификации цифровых архивов необходим для создания общего климата надежности относительно перспектив хранения цифровой информации.

Как упоминалось выше, сейчас мы имеем отчет RLG/OCLC [3] по этому аспекту инфраструктуры цифрового хранения. Здесь я только перечислю, что содержит отчет. Отчет:

— предполагает определение надежного цифрового хранилища:

— указывает первичные признаки надежного цифрового хранилища,

— представляет подход к разработке программы сертификации,

— определяет обязанности услужливого цифрового хранилища в виде открытой системы архивной информации (OAIS),

— информирует сообщества RLG/OCLC о других разработках, необходимых для реализации надежного хранилища,

— дает формальные рекомендации для дальнейшей работы.

Помимо сертифицированных архивов, я думаю, необходимы также два других вида организаций, обсуждаемых ниже.

Совместные структуры

Даже поверхностный взгляд на проблемы цифрового хранения указывает на позитивную необходимость сотрудничества среди заинтересованных сторон и учреждений. Очевидно, что такое сотрудничество, вероятно, сократит стоимостные расходы либо путем экономии масштаба, либо путем других средств. Я думаю, что это и правильно, и важно, но, как я полагаю, не передает масштаба стоящей перед нами проблемы. Пример может помочь пояснить, что я имею в виду. Предположим, я убежден, что будущие исследователи могут счесть полезным иметь некоторые хранимые данные, которые собираются супермаркетами [5]. Однако, это небольшое слово "некоторые" ставит большие вопросы. Во-первых, кто отбирает информацию для хранения? Очевидный ответ, — вероятно, пользователи данных — только ставит еще больше вопросов. На первый взгляд, такой потенциальный интерес должны иметь социологи и историки, и это предполагает, что должен быть некий совместный механизм, посредством которого заинтересованные стороны могут отбирать такие данные. В принципе, это прекрасно, но таких совместных механизмов в настоящее время нет.

Пример второй крупнейшей проблемы показывает, что создатели данных представляют собой различное множество людей из числа потенциальных пользователей данных. Будут ли создатели данных иметь какое-либо стремление к сотрудничеству с пользователями? Некоторые могут утверждать, что это происходит там, где возрастает необходимость в

"ошибкоустойчивом" механизме, но прежде чем продолжить говорить об этом, давайте попытаемся рассмотреть эту проблему в более широком контексте. Более широкий контекст должен позволить начать с отражения широкого разнообразия существующих интересов. Один путь (предположительно не прямой) рассуждения о цифровом мире может быть представлен как диаграмма Венна (рис. 2).



Рис. 2. Представление цифровой области

Кто-то может представлять этот мир разделенным на три категории — библиотеки и архивы, научные сообщества и коммерческие круги, — каждая из которых имеет подмножества, работающие в первую очередь в цифровой области.

Диаграмма показывает существование изобилия различных интересов и культур: общественных и частных, коммерческих и бесприбыльных, национальных культур и подкультур. Эти различные интересы и культуры будут, естественно, влиять на организационный способ действия, если он относится к цифровому миру. На этом он не останавливается. В зависимости от присущих им различных интересов и культур цифровые сообщества, вероятно, имеют и различные приоритеты, например в отношении доступа, управления правами, взаимодействия или хранения и т. д. И здесь он также не остановится до тех пор, пока, как сказано выше, роли в цифровом процессе (создатель данных — пользователь данных — хранитель данных) не приблизятся к различным направлениям и интересам. Несомненно, некоторые из этих интересов будут конфликтовать, но многие будут просто различными.

Это противоречит той основе, что совместные структуры должны быть построены. Возможно, эти совместные структуры могут привести к конфликтному решению, но Оптимистично, по меньшей мере, предположить, что такое решение будет всегда возможным. Тем не менее, наверно, есть сфера для большого объема совместных усилий в подходе к тем решениям, которые не может осуществить одна группа или сообщество собственными силами. Таким образом, в отличие от того, что предложено в следующем разделе, "совместные структуры" представляют собой не столько сами организации, сколько процессы коммуникации между существующими (или и новыми) организациями. Возможно, такие инициативы, как коалиция цифрового хранения, могут рассматриваться в качестве первых шагов в направлении таких структур [6].

Учитывая, что вышесказанное исходит из конституциональной перспективы цифрового хранения, также стоит отметить, что проблемы цифрового хранения всё в большей степени становятся важными и для отдельных лиц. В возрастающей степени отдельные лица делают большие вклады

в свое использование цифровых объектов. Например, рассмотрим использование ими персональных цифровых ассистентов (PDA — Personal Digital Assistants). Подобно другим технологиям, PDA могут неожиданно стать устаревшими в виртуальном плане, даже когда они производят впечатление хорошо созданных. Хотя это важно с точки зрения отдельного лица, возникают также две точки зрения относительно более широких интересов сообщества. Клиффорд Линч [7] указывает, что в отношении физических объектов скорее отдельные лица, нежели учреждения, часто являлись теми, кто хранил объекты, позже признанные важными. Поэтому мы должны быть осторожными с решениями, относящимися к проблемам цифрового хранения, которые должны принять форму разрешения специальной обработки для учреждений (например, сертифицированные архивы), но не для отдельных лиц.

Центры исследования и разработки централизованного цифрового хранения

Учреждение должно предпринять множество действий для хранения цифровых объектов, таких, которые необходимо провести самому учреждению, или, по меньшей мере, третьей стороне, осуществляющей действия от имени учреждения. Однако существуют и другие действия, действия не такого характера — действия, которые может предпринять незначительное число людей от имени многих лиц. Для таких действий учреждения должны осознать необходимость сотрудничества в целях поддержки централизованных исследовательских центров даже только по стоимостным причинам. Такие действия могут включать:

- Разработку и поддержку эмуляторов
- Разработку средств метаданных
- Обеспечение услуг по восстановлению данных (коммерческие службы восстановления данных существуют, но они очень дорогие).

Конечно, создание таких центров, особенно и предпочтительно на международной основе, должно, очевидно, стать большим предприятием; тем не менее, совершенно ясно, что это необходимо.

ПРАВОВЫЕ АСПЕКТЫ

Правовая среда, в которой происходит цифровое хранение, наиболее очевидно поднимает две трудные темы: права интеллектуальной собственности и электронное хранение.

Права интеллектуальной собственности

В разумном мире должен существовать баланс между интересами, с одной стороны, обладателей авторского права, и, с другой стороны — потребителей информации. В мире книг разумный баланс был достигнут. Авторы имеют право на авторский гонорар и защищены законом авторского права. Этот баланс зависит от практических факторов континента, таких как факт, что фотокопирование целой книги не является привлекательным выбором с точки зрения требуемого усилия и затрат. Интересы читателей защищены полисами справедливого использования и первой продажи [8]. Первая продажа, в частности, делает возможным существование публичных библиотек с выдачей книг на дом. Как сообщается «Комитетом по правам интеллектуальной собственности и

возникновению информационной инфраструктуры» [9], «цифровое нашествие» угрожает этому балансу. С точки зрения поставщика содержания, угрозы возникают от легкости, с которой могут копироваться цифровые объекты без потери качества, и из-за того, что объекты могут с легкостью распространяться через сети. Наоборот, потребителю информации могут угрожать навязывание драконовских условий использования и новые модели бизнеса (например, лицензирование, плата за просмотр и т. д.), которые приведут к исчезновению «первой продажи». В отношении первого недавно один продавец электронной книги включил условия использования с абсурдным запретом на чтение работы вслух!

Я особенно обеспокоен тем, что введение систем технической защиты (TPSs — Technical Protection Systems) грозит сделать более трудными (и поэтому более дорогими), невозможными или противоправными действия, необходимые для цифрового хранения. Вывод в отчете «The Digital Dilemma» («Цифровая дилемма») [9] состоит в том, что «... существует больше правовых причин обойти контроль доступа, чем в настоящее время признается в Законе об авторском праве цифрового тысячелетия» (Digital Millennium Copyright Act — DMCA).

Упоминание систем технической защиты указывает на правильно сформулированный вывод, сделанный в «Цифровой дилемме», который состоит в том, что достижение разумного баланса между интересами потребителей и интересами поставщиков содержания — это не только правовой вопрос; скорее, это достижение баланса между правовой системой, существующей технологией и моделью бизнеса. С технологией в ее современном состоянии непрерывного изменения невозможно определенно разрешить все эти проблемы в настоящее время. Действительно, рекомендация из «Цифровой дилеммы» гласит о том, что «Законодатели не должны рассматривать кардинальные изменения закона об интеллектуальной собственности и публичной политике в это время, чтобы позволить эволюционному процессу завершиться» [9, с. 16]. Что сейчас важно, так это чтобы сообщество цифрового хранения было проактивным в целях гарантии развития этого процесса таким образом, чтобы не нанести вреда интересам хранения.

Электронное хранение

Упомянутый выше эволюционный процесс долгожданного закона об интеллектуальной собственности и публичной политике должен включать разработку систем электронного хранения, сходных с такими же системами, используемыми для книг. «Цифровая дилемма» рекомендует специальной комиссии изучить проблему и в интересном примечании предусматривает один важный выход, который гласит:

«... один сценарий может требовать добровольного (или обязательного) хранения цифровых работ, которые защищены авторским правом США, и тех, которые предлагаются для продажи по лицензии или, будучи распространяемыми свободно и бесплатно, обеспечиваются системами технической защиты. Такие обязательные экземпляры не могут быть доступными общественности через депозитарий, поскольку они все еще предлагаются общественности через обладателей авторских прав, за исключением просмотра в самой библиотеке (как в случае с печатными работами). Все депозитные копии должны быть «вне подозрения» (т. е. без шифров и других лимитирующих доступ механизмов)... Цель здесь состоит



Рис. 3. Возможные конфликты между поставщиками и цифровыми сообществами

в том, чтобы распространить на цифровой мир традиционный, сбалансированный закон об интеллектуальной собственности — обеспечение достаточного контроля за работой, чтобы предложить стимул для творчества, а также гарантию того, что со временем все работы станут частью публичной, интеллектуальной записи на благо всего общества. Обеспечение хранения материалов «вне подозрения» может помочь в рассмотрении проблем доступа, которые возникают из механизмов технической защиты, а также проблем, поднятых архивированием» [9, с. 208].

КУЛЬТУРА

«Кажется, что некоторые поставщики содержания имеют амбиции, более подходящие для некоторой оруэлловской антиутопии». Клиффорд Линч [7].

Под «культурой» в этом контексте я понимаю множества мнений, отношений и ценностей, связанных с встречающимися проблемами, касающимися цифрового хранения. Прежде чем посмотреть на некоторые вызывающие беспокойство новые разработки, стоит поразмышлять над культурой информационной технологии в недавнем прошлом. Как потребителю, мне кажется, что культура была достойной сожаления, особенно на самом большом рынке информационной технологии, а именно в мире персонального компьютера. Я упомяну только два аспекта этой управляемой поставщиком культуры — надежность и моральный износ. Общеизвестно, что почти более десяти лет действующие системы наиболее популярных платформ были ненадежными и склонными к исчезновению. Современные версии этих операционных систем рекламируются как более надежные, хотя надежность не рассматривалась в качестве основного, стоящего на первом месте, требования. В то же время потребители были постоянно виртуально вынуждены повышать качество как технического, так и программного обеспечения, несмотря на реальные нужды потребителей, поэтому после относительно короткого срока неудача в улучшении станет в результате тем, что они будут иметь неподдерживаемую систему.

На этом фоне совсем неутешительно видеть, как возможности новой информационной технологии становятся доступными, что может угрожать потребителям в общем и цифровому хранению, в частности. Рассмотрим следующий ряд возможных конфликтов, приведенных на рис. 3.

Это факт, что устаревание может быть в интересах некоторых людей, по крайней мере, если эти интересы понимаются узко в терминах коммерческой выгоды.

Перед лицом этих угроз существует утешающая мысль,

что, если поставщики попытаются протолкнуть свои продукты слишком далеко, они столкнутся с сопротивлением покупателя. Например, если производители электронных книг для чтения предпримут попытку навязать драконовские условия использования, они убьют гусыню, которая, как они надеются, снесет золотые яйца. Однако в настоящее время неясно, насколько это утешительно, поскольку те же самые поставщики могут быть в состоянии протолкнуть продукты достаточно далеко, чтобы создать проблемы для цифрового хранения, прежде чем отпугнуть потенциальных покупателей. Некоторые из нас полагают, что определенные методы извлечения прибыли не приемлемы, включая действия, которые ставят под угрозу наше настоящее или будущее цифровое наследие.

Угрозы, указанные на диаграмме рис. 3, представляют пессимистический сценарий. Однако существует альтернативный оптимистический сценарий. Оптимистический сценарий заключается в том, что хотя современное цифровое хранение является трудным и дорогостоящим, поскольку производители удовлетворяют требования проактивного сообщества цифрового хранения, цифровое хранение можно сделать более легким и дешевым. Разработка и адаптация стандартов внесут свой вклад в то, чтобы сделать этот оптимистический сценарий реальностью. Кроме того, то, что я подразумеваю под проактивным сообществом цифрового хранения, — это такое сообщество, которое не примет фатально вносимую правовую культуру. Рассмотрим это положение:

«... стандартизация сеет семена своего собственного уничтожения, поощряя продавцов внедрять нестандартные признаки в целях сохранения рыночной доли» [11].

Я считаю это положение поразительным — это, скорее, походит на криминальное высказывание: «если у вас нет законов, мне не придется их нарушать».

Возможно, часть скептицизма относительно стандартизации происходит от предположения некоторых лиц, что когда кто-то говорит о стандартах, он, должно быть, говорит о МОС (Международная организация по стандартизации) или схожих процессах стандартизации. Общеизвестно, что стандартизация МОС медленная, — случается, что стандарт представленный вовремя, уже устарел, — или, по крайней мере, таково ощущение. Я хочу сказать по этому поводу две вещи: первое — некоторые стандарты МОС действительно доказывают свою важность. Второе — модель МОС не является единственной моделью стандартизации, Техническая рабочая группа Интернета (IETF — Internet Engineering Task Force) работает по совершенно другой модели. Протоколы сети (т. е., стандарты) являются обязательным условием Интернета. Интернет работает, и было бы трудно утверждать, что Интернет быстро не изменяется и не развивается [12].

Ссылка на Интернет должна нам напомнить, что действительно стандарты повсеместны. В работе “Стандарты для системы хранения электронных публикаций” (“Standards for DSEP — Deposit System for Electronic Publication”) [13] даются указания примерно на 40 соответствующих стандартов (за исключением каких бы то ни было, касающихся форматов данных [14]). Старая шутка гласит, что “мы любим стандарты — вот почему у нас их так много”. Действительно, избыток стандартов иногда составляет проблему, и нигде проблема не кажется более острой, чем тогда, когда она относится к метаданным. Иногда кажется, что сообщество пользователей имеет свою 30-страничную спецификацию для метаданных, но только метаданных для специальной цели. Мне известны две спецификации, цифрового хранения. Как мы реально можем иметь метаданные, которые обеспечат все нужды архива (обнаружение, управление правами, взаимодействие и хранение), чтобы назвать самые очевидные? Тем не менее, существуют области, где новые стандарты будут необходимы. Когда обещанная земля обетованная сертифицированного архива наступит, нам потребуются стандарты для процесса сертификации, — стандарты, которые обеспечат гарантию качества (возможно, подобная или даже специальная версия МОС 9001). Пользующийся популярностью появившийся стандарт OAIS для открытых архивов перечисляет 12 областей, где стандарты необходимы [15].

ТЕХНИЧЕСКАЯ ИНФРАСТРУКТУРА

Большая часть того, что я разместил в блоке, названном “техническая инфраструктура”, на рис. 1, уже была объяснена в предыдущих разделах. Здесь я бы хотел просто прокомментировать пункт “И т. д., и т. д.”, который является держателем места не только для вещей, о которых я не имею представления, но также, как я надеюсь, для вещей, еще не изобретенных. Поскольку требования цифрового хранения будут определяться в дальнейшем, я полагаю, что эти требования будут включать средства технического и программного обеспечения, методы или компоненты, которые сделают цифровое хранение более легким и дешевым. И если мы добьемся успеха в изменении культуры, поставщики будут рады внедрить их.

ИДТИ ВПЕРЕД: ДВОЙНОЙ ПОДХОД

Ясно, что глубокая инфраструктура, за которую ратует данная статья, потребует для своего создания много лет. По той причине для учреждений разумно принять двойной подход: прагматический подход и стратегический подход. Прагматический подход должен состоять из хранения цифровых объектов *в настоящее время* в самом лучшем виде, в каком только можно. Уже есть несколько руководств по доступной наилучшей практике, охватывающей многие аспекты цифрового хранения. Более категорично я утверждаю, что учреждения должны также принять *стратегический* подход. Этот подход должен иметь дело со средне- и долгосрочными проблемами и включать следующее:

- Создание совместных структур
- Разработку стандартов
- Создание надежных архивов
- Поддержку научных и исследовательских центров централизованного цифрового хранения.

В ранее описанных в данной статье условиях эти виды деятельности должны помочь изменить культуру на такую, где наше цифровое культурное наследие будет защищено.

ПРИМЕЧАНИЯ И ЛИТЕРАТУРА

1. Preservation digital information: Report of the Task Force on Archiving of Digital Information, Commission on Preservation and Access and The Research Libraries Group, Inc., May 1, 1996, <<http://www.rlg.org/ArchTF>>.

2. За точки зрения, выраженные в этой статье, полностью отвечает автор, и они необязательно отражают точки зрения моих коллег.

3. Признаки надежного цифрового хранилища приводятся на сайте <<http://www.rlg.org/longterm/attributes01.pdf>>.

4. Другой аспект был предположением отчета относительно того, что перемещение является единственной жизненно важной стратегией цифрового хранения. О причинах сомнения в этом см. Stewart Granger. Emulation as a digital preservation strategy // D-Lib Magazine. <<http://www.dlib.org/dlib/october00/granger/10granger.html>>, October 2000, а также работы Digital preservation and emulation: From theory to practice // ICHIM 2001, Proceedings, Sept 3–7, Milan, ISBN 1-885626-24-X, доступную на: <<http://dSPACE.dial.pipex.com/stewartg>>.

5. Russel Kelly. Preserving digital scholarly resources: Progress through collaboration // Assignment.— Vol. 18, No. 2.— ISSN 0265-2587.

6. Информация о Коалиции имеется на <<http://www.jisc.ac.uk/dner/preservation/prescoalition.html>>.

7. Lynch Clifford. The battle to define the future of the book in the digital world // First Monday, 2001 at: <http://firstmonday.org/issues/issue6_6/lynch/index.html>.

8. Для упрощения двух сложных проблем: справедливое использование есть право использовать информацию об авторском праве без разрешения для определенных ограниченных целей, например для просмотра работы. Первая продажа применяется к физическому объекту, такому как книга, — владелец авторского права может контролировать продажу книги и устанавливать цену, но как только кто-то покупает книгу, он имеет полное право копировать ее и поступать с ней по своему усмотрению.

9. The digital dilemma. — Computer Science and Telecommunications Board, National Research Council, National Academy Press.— 2000.— ISSN 0-309-06499-6.

10. Понятия “справедливое использование” и “первая продажа” используются здесь просто для обеспечения разумного баланса между интересами, но я не берусь предсказывать, каким будет оптимальное решение.

11. Rothenberg Jeff. Avoiding technological quicksand: Finding a viable technical foundation for digital preservation.— 1999. <<http://www.clir.org/pubs/reports/reports.html>>.

12. Lynch Clifford A. The case for new economic models to support standardization efforts. <<http://www.niso.org>>.

13. Feenstra Bendert, Haag Den. Standards for a DESP: Standards for the implementation of a deposit system for electronic publications // NEDLIB Report 4. <<http://www.kb.nl/nedlib/results/dsepsstandards.rtf>>.

14. Не все они являются полными международными стандартами, некоторые представляют собой справочники.

15. Sawyer Don, Reich Lou. Reference model for an Open Archival Information System (OAIS)/NSA and CSC report: Red Book, Issue, May, URL: <http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html>.