

# Обзор: Цифровые тёмные века уже наступили? Утрата электронной информации в трёх шведских архивах

 [rusrim.blogspot.com/2016/10/blog-post\\_13.html](http://rusrim.blogspot.com/2016/10/blog-post_13.html)

В данной заметке Крис Эрикссон даёт обзор магистерской диссертации Анны-Марии и Эрика Андерхилл (Anna-Maria Underhill, Arrick Underhill) «Цифровые тёмные века уже наступили? Утрата электронной информации в трёх шведских архивах» (A Digital Dark Now? : Digital Information Loss at Three Archives in Sweden), защищённой в 2016 году в университете шведского города Лунд (Lund University, <http://www.lunduniversity.lu.se/>), по направлению архивного, музейного, библиотечного дела и информатики. Научный руководитель работы – Ларс Бьёрк (Lars Björk). Полный текст работы объёмом 92 страницы на английском языке доступен по адресу <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=8876749&fileId=8876760>



Целью настоящего исследования является изучение случаев утраты электронной информации в трёх шведских архивах. Обеспечение долговременной сохранности электронных материалов (электронной сохранности - digital preservation) является сложным вопросом, который большинство архивных учреждений пытается решить. Если всё внимание сосредоточить лишь на успехах, не обращая внимания на неудачи, то возникает риск создать «белое пятно», скрывающее существующие проблемы.

Под электронной информацией в данном исследовании понимаются электронные объекты и их метаданные. Исследование также охватывает внутренние электронные рабочие документы, которые используются для контекстуальной поддержки коллекций архива. Переход архивов на опережающие действия по обеспечению электронной сохранности анализируется с помощью «модели континуума» (Continuum Model). Далее используются предложенные Тибодё (Thibodeau) онтологические представления об электронной информации для выявления тех аспектов, которые могут быть утрачены. Затем анализируются технические аспекты утраты информации на основе модели SPOT для оценки риска, а также рассматриваются вопросы преимущественно организационного характера с использованием «Эталонной модели открытой архивной информационной системы» (OAI Reference Model).

Некоторые из полученных результатов вновь подтвердили выводы предыдущих исследований, например, о необходимости приоритетного внимания к организационным вопросам. Другие результаты отражают текущее состояние электронной сохранности в этих трёх архивах, которое включает в себя тонкое балансирование «между созданием систем для успешного решения задачи электронной сохранности в будущем, и одновременным управлением существующими электронными коллекциями, сохранность которых, возможно, не обеспечивалась надлежащим образом».

Некоторые учреждения не в состоянии перейти на использование предусматривающих опережающие действия форм обеспечения электронной сохранности ввиду характера хранимых ими материалов. В исследовании отмечается, что «при обсуждении вопроса электронной сохранности, сохраняется тенденция сначала думать об оцифрованных, а не об изначально-электронных электронных материалах». Утрата дел и

**A DIGITAL DARK NOW?**  
Digital information loss at three archives in Sweden

Anna-Maria Underhill & Arrick Underhill

Master's thesis (30 credits) in Archival Science within the Master's Program of Archival Science, Library- and Information Science and Museum Studies at Lund University  
Supervisor: Lars Björk  
Year: 2016

© Anna-Maria Underhill and Arrick Underhill

документов – не единственный вид потерь; возможна также потеря метаданных и связей между информацией, которая может быть более распространена, чем утрата целых электронных объектов. «И, наконец, один вопрос сопровождал это исследование с самого начала и до конца: Как можно узнать о том, что Вы потеряли нечто, о чьем существовании Вы никогда не подозревали?».

Обращается внимание на следующие моменты:

- При обсуждении вопроса об обеспечении электронной сохранности, важно разъяснить, что хранение и электронная сохранность – не одно и то же (*поскольку электронная сохранность предусматривает целый комплекс работ по выявлению объектов, подверженных риску технологического устаревания, по планированию и осуществлению их конверсии/миграции – т.е. это активная деятельность в отличие от пассивного «складского» хранения – Н.Х.*);
- Выживание информации зависит от поддержания соответствующей инфраструктуры и от миграции её в современные форматы;
- Сохранение аутентичности может быть серьезной проблемой для электронных документов, и аутентичность имеет важное значение для ценности документов в качестве свидетельств / доказательств;
- Эмуляция является ещё одним вариантом обеспечения электронной сохранности, при котором основное внимание уделяется среде оперативного использования информации, а не файлам;
- При использовании эмуляции в конечном итоге все равно потребуется проводить миграцию. Эмуляция может оказаться слишком сложной для того, чтобы быть жизнеспособной в долгосрочной перспективе;
- Иногда усилия в плане электронной сохранности не обеспечивают сохранения того, что она предполагалось сохранить, и это можно считать утратой информации;
- Моральное устаревание в настоящее время является одной из главных угроз для успешной электронной сохранности. Если файл не может быть прочитан, то это практически равносильно уничтожению документа;
- «В отсутствии знаний о происхождении документов и контекстных взаимосвязей между ними, невозможно доказать аутентичность и надежность документов, теряется их ценность как свидетельств / доказательств, и становится сложно их использовать в качестве источников сведений о происшедшем и для понимания того, что произошло».

В работе приведено определение краткосрочной, среднесрочной и долговременной сохранности:

- Краткосрочная сохранность - решения, которые используются на коротком интервале времени, в течение максимум 5 лет;

- Среднесрочная сохранность - решения, которые используются в течение срока службы системы, в течение максимум 10 лет;
- Долговременная сохранность - решения, которые используются после того, как первоначальная система выведена из эксплуатации, сроки варьируются от 10 до 50 лет.

**Мой комментарий:** Сразу возникает вопрос, а что с материалами, сроки хранения которых превышают 50 лет? Придумывать для них отдельную категорию «постоянной сохранности» в общем-то бессмысленно. Как отмечали в одной из своих концепций британские электронные архивисты, на практике задача сегодняшних архивистов состоит в том, чтобы передать электронные архивные документы через 30-50 лет в более-менее приличном состоянии следующему поколению архивистов, и далее нести эту эстафету.

С моей личной точки зрения, более практично привязывать понятия «краткосрочности» и т.д. не к конкретным календарным срокам (которые, кстати говоря, названы вполне разумные), а исходя из необходимости выполнения специальных работ. Если электронный документ можно хранить как есть, обеспечивая целостность и неизменность потока битов в рамках одной и той же информационной системы и не собирая дополнительных сведений – это кратковременная сохранность, даже если она затянется на столетие. Если у электронного документа, подписанного сегодня утром, к вечеру истек срок действия сертификата ключа проверки, это уже как минимум среднесрочная сохранность, потому что потребуются дополнительные усилия и сбор поддерживающей документации для подтверждения подлинности такого документа.

«Темные архивы» (т.е. архивы, не выдающие свои материалы исследователям – Н.Х.) часто используются для того, чтобы отделить оригинальные мастер-копии файлов от экземпляров / копий, к которым пользователи реально имеют доступ. Доступ к этим «темным архивам», как правило, осуществляется только тогда, когда в них помещаются новые материалы, а в остальном они защищаются в целях сохранения аутентичности оригиналов, путем помещения их в среду, которая в максимальной степени предотвращает возможность манипуляций и ошибок».

Подлежат сохранению шесть важнейших для целей электронной сохранности свойств:

- Доступность
- Идентичность (Identity)
- Неизменность
- Отображаемость (Renderability)
- Понятность
- Аутентичность

**Мой комментарий:** С этим списком я бы поспорила. Аутентичность охватывает идентичность; доступность, отображаемость и понятность вполне объединятся в одну категорию – пригодность к использованию; забыта полнота, которая вместе с неизменностью объединяется в категорию

*«целостность»;* да и не стоило бы забывать о конфиденциальности.

В ходе исследования выявлены типы фактической и потенциальной утраты информации:

- Утрата электронных объектов или их частей во время миграции;
- Потеря взаимосвязей между аналоговой и цифровой информацией, хранимой одним и тем же архивом;
- Потеря информации вследствие её сохранения в неподходящем формате;
- Потеря данных в связи с технологическими изменениями;
- Потеря электронной информации при её хранении вместе с аналоговыми материалами;
- Потеря информации из-за устаревания оборудования;
- Потеря метаданных из-за того, что базы данных не являются программным обеспечением с открытым исходным кодом.

Причинами такой реальной и потенциальной потери информации являются:

- Человеческие ошибки в процессе создания информации;
- «Неэлектронный» подход к пониманию и обработке электронной информации;
- Отсутствие у организации соответствующей структуры и стратегии электронной сохранности;
- Недостаток ресурсов;
- Технологические ограничения;
- Отсутствие необходимых компетенций у персонала, работающего с электронной информацией.

### **Крис Эриксон (Chris Erickson)**

**Мой комментарий:** Значительную часть диссертации составляет пересказ содержания бесед студентов с сотрудниками архива, и здесь можно найти немало любопытного. Например, в «Архиве А» были случаи утраты информации в результате неудачного преобразования документов в формат PDF/A (это Росархиву и Минкомсвязи на заметку).

Интересно, что опрошенные архивисты дружно подчеркнули, что одни технические решения проблему электронной сохранности не решат, и что очень большое значение имеют знания, навыки и добросовестность персонала архива. Они отметили, что эти специфические знания очень легко потерять.

Источник: блог Digital Preservation Matters

<http://preservationmatters.blogspot.ru/2016/06/a-digital-dark-now-digital-information.html>