

# Электронные архивы: о постановке задачи долговременного хранения электронных документов

А.В. Соловьев

**Аннотация.** В статье рассматривается концептуальная постановка задачи управления долговременным хранением электронных документов в программно-аппаратной среде.

**Ключевые слова:** электронный документ, электронный документооборот, электронный архив, система управления электронными документами.

## Введение

В работах автора [1-5] было показано, что проблема долговременной сохранности электронных документов является крайне актуальной. Поиском решения проблемы занимаются все ведущие страны, т.к. со временем количество электронных документов будет стремительно возрастать, следовательно, порядок сложности решения задачи будет стремительно увеличиваться.

Обозначены проблемы, которые являются ключевыми при решении задачи долговременной сохранности: сохранность аутентичности документа в течение всего срока хранения; «старение» носителей информации; неизбежное обновление программно-аппаратной среды хранения; проблема интерпретируемости и отображения электронных документов при длительном хранении.

В то же время пока не видно универсального решения данной проблемы, несмотря на активные поиски такого решения в последние десять лет.

Поиск решения должен быть однозначно связан с выработкой комплексной технологии долговременной сохранности. А для этого задача должна быть четко сформулирована. Постановке задачи долговременной сохранности посвящена данная статья.

## 1. Основные понятия и определения

Электронный архив (ЭА) – структурированное хранилище неизменяемых электронных оригиналов документов (электронных изображений бумажных документов), созданное на основе законов и правил ведения архивов на конкретной территории (в конкретной стране).

Длительное хранение – хранение электронных документов не менее 5 лет.

Определение не претендует на «абсолютность», т.к. в конкретных архивах эти сроки могут меняться. За основу срока (5 лет) взят максимальный срок хранения документов в оперативных архивах СЭД. Электронные документы могут храниться в течение десятилетий или даже столетий или «бессрочно» в зависимости от их важности.

Долговременная сохранность – «период времени, в течение которого электронные документы поддерживаются в качестве доступного и аутентичного свидетельства (доказательства)» [6].

Доступность информации – «возможность реализации беспрепятственного доступа к информации субъектов, имеющих на это надлежащие полномочия» [7] и, одновременно, «из-

бежание временного или постоянного сокрытия этой информации от пользователей, получивших права доступа» [8].

Доступность документа – свойство документа, состоящее в том, что форма представления документа обеспечивает физическую возможность измерения заданных параметров этого представления документа (содержания, атрибутов, технологии) заданными средствами в заданных точках за конечное время [9].

Аутентичный электронный документ – «электронный документ, точность, надежность и целостность которого сохраняются с течением времени» [6].

Деловые (управленческие) документы – официальные документы, регламентирующие деятельность организаций, учреждений, предприятий, фирм, а также документы, связанные с деятельностью официальных лиц.

Электронный деловой (управленческий) документ – деловой документ в электронной форме, может быть заверен электронной подписью (ЭП). Ниже, употребляя термин электронный документ (ЭлД), мы будем иметь в виду электронный деловой документ.

Электронная подпись (ЭП) – «информация в электронной форме, которая присоединена к другой информации в электронной форме (подписываемой информации) или иным образом связана с такой информацией и которая используется для определения лица, подписывающего информацию» [10].

## 2. Общая постановка задачи обеспечения долговременной сохранности, основные ограничения и допущения

Если исходить из приведенных выше определений, то постановка задачи долговременной сохранности ЭлД формулируется следующим образом: необходимо обеспечить долговременное хранение электронных деловых документов, при этом обеспечить аутентичность, интерпретируемость (читаемость), подтверждение авторства для документа, а также надежность, катастрофоустойчивость среды хранения документа в течение всего срока хранения.

При этом предполагается, что:

- аутентичность документа на момент передачи его в ЭА подтверждена;
- документы не искажены;
- сохранность документов полная;
- нет ограничений на форматы данных передаваемых в ЭА документов;
- аппаратно-программная среда, в которой функционирует ЭА, подвержена постоянному изменению (обновлению);
- ЭА требует периодической модернизации вслед за изменением аппаратно-программной среды;
- ЭА сертифицирован для работы со средствами ЭП.

Предполагаем, что при долговременном хранении истекают сроки действия сертификатов ЭП, заканчивается оперативное хранение документов в архивах подразделений, завершается поддержка версий некоторых операционных систем (ОС), например Windows, и прикладного программного обеспечения.

Мы видим, что жизненный цикл хранения электронного документа в ЭА превосходит сроки жизни оборудования и программного обеспечения. Так, например, документы по личному составу должны храниться 75 лет. Сравнив этот срок с данными, приведенными в Табл. 1 (составлена по обзору [1]), можно легко убедиться, что в течение такого срока хранения несколько раз будет обновлено как системное программное обеспечение, так и носители данных.

Тем самым можно утверждать, что ЭлД становится объектом управления, а задача долговременной сохранности – это задача оптимального управления в условиях параметрических возмущений [11, 12]. А для оптимального управления необходимо научиться контролировать (измерять) параметры среды хранения и разработать алгоритмы компенсации возмущений для организации обратной связи с целью стабилизации объекта управления [13].

На Рис.1. приведена структурная схема задачи долговременной сохранности ЭлД.

Параметрические возмущения:

$\varepsilon(\alpha(t))$  – нарушение аутентичности ЭлД,

$\varepsilon(\zeta(t))$  – нарушение интерпретируемости ЭлД,

$\varepsilon(\varphi(t))$  – изменения программно-аппаратной среды,

$\varepsilon(\rho(t))$  – нарушение надежности хранения ЭлД,  
 $\varepsilon(v(t))$  – нарушение устойчивости хранения.  
 Сохранность объекта управления (ЭлД):  
 $\mu(t)$  – функция оценки сохранности ЭлД,  
 $x(\mu(t))$  – исходная сохранность ЭлД,  
 $y(\mu(t))$  – результирующая сохранность ЭлД.

### 3. Математическая постановка задачи обеспечения долговременной сохранности

Если говорить о математической постановке задачи долговременной сохранности, то она представляет собой задачу стабилизации объекта управления по многим параметрам, которая может быть сведена к задаче оптимального выбора по многим критериям.

Действительно, пусть сохранность электронного документа характеризуется целевой функцией  $\mu(t)$ , которая представляет собой вероятность сохранности электронного документа в произвольный момент времени.

Тогда можно утверждать, что задача долговременной сохранности формулируется, как задача достижения максимума функции  $\mu(t)$  на произвольном временном интервале  $t$  (при  $t \rightarrow \infty$ ), то есть

$$M(t) = \max_{t \in [0, \infty]} \mu(t). \quad (1)$$

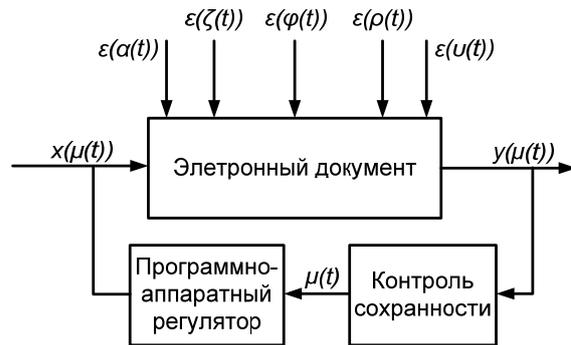


Рис. 1. Структурная схема задачи долговременной сохранности ЭлД

Функция  $\mu(t)$  определяется через частные целевые функции:

$\alpha(t)$  – функция аутентификации, определяющая вероятность сохранения аутентичности документа в момент времени  $t$ ;

$\zeta(t)$  – функция интерпретируемости, определяющая вероятность интерпретации и визуализации документа в произвольный момент времени (в данном случае под вероятностью мы понимаем оценку риска не интерпретируемости документа);

Под риском здесь и далее будем понимать количественную оценку в диапазоне  $[0, 1]$ , которая определяет степень приближения значения целевой функции к максимуму, т.е. значению 1.

Табл. 1. Общие проблемы обновления аппаратно-программной среды ЭА при длительном сроке хранения

Срок	Проблемы программной среды	Проблемы аппаратной среды
3 года	1) истечение сроков сертификатов ЭП	
5 лет	1) предельный срок сертификатов ЭП по 63-ФЗ 2) необходимы существенные доработки ПО ЭА	1) выход из строя носителей информации (CD, DVD, HDD, флэш-накопителей и др.)
10 лет	1) снимается с поддержки операционная система 2) устаревают программы просмотра 3) прекращается поддержка закрытых форматов данных	1) предельный срок службы SSD и FRAM (Ferroelectric Random Access Memory – сегнетоэлектрическая память с произвольным доступом) 2) проблемы с модернизацией техники
20 лет	1) значительные ограничения при использовании устаревшей ОС и устаревших программ просмотра 2) ОС можно использовать только в виртуальной среде 3) возможны проблемы даже с открытыми форматами данных 4) возможны изменения стандартов криптозащиты, как следствие - невозможность подтвердить ЭП	1) предельный срок службы аппаратных средств хранения

Табл. 2. Пример вербальной шкалы для приблизительной оценки вероятности извлечения или интерпретируемости Элд

Номер п.п.	Наименование значений	Числовой эквивалент
1	Не интерпретируется/не отчуждается	0
2	Скорее не интерпретируется/не отчуждается	0,25
3	Частично интерпретируется/отчуждается	0,5
4	Скорее интерпретируется/отчуждается	0,75
5	Полностью интерпретируется/отчуждается	1

Если испытания Бернулли в классическом понимании оценки вероятности невозможны или трудоемки, то предлагается следующая шкала значений оценки риска для приведенных показателей функции интерпретируемости и отчуждаемости (Табл. 2).

Оценка по данной шкале выполняется экспертом (группой экспертов), т.е. лицом (лицами), обслуживающим ЭА и знакомым с программно-аппаратной средой функционирования ЭА.

Для более точной оценки вероятности интерпретируемости и отчуждаемости Элд в программное обеспечение ЭА должна быть заложена функция периодической автоматической инвентаризации фонда Элд с целью автоматически определить возможность интерпретируемости и отчуждаемости каждого Элд с автоматическим подсчетом вероятности (испытания Бернулли).

Если интерпретируемость и отчуждаемость Элд (части Элд) автоматически выполнить невозможно, обслуживающий ЭА персонал должен быть уведомлен системой ЭА о нарушении принципа интерпретируемости и отчуждаемости по конкретному Элд. Также должна быть предоставлена возможность проверить проблемный Элд на возможность ручной интерпретируемости и отчуждаемости. Если и эта попытка не удалась, показатель вероятности извлечения данного Элд автоматически вычисляется системой ЭА как равный нулю.

$\varphi(t)$  – функция отчуждаемости, определяющая готовность документа к миграции (отчуждению документа от конкретной программно-аппаратной среды хранения), т.е. вероятность извлечения документа из конкретной аппаратно-программной среды в момент времени  $t$  (в данном случае под вероятностью мы понимаем оценку риска неизвлечения документа);

$\rho(t)$  – функция надежности, определяющая вероятность работоспособности программно-

аппаратной среды хранения электронного документа в произвольный момент времени;

$v(t)$  – функция устойчивости, определяющая степень катастрофоустойчивости программно-аппаратной среды хранения в момент времени  $t$  (например, вероятность восстановления аппаратно-программной среды в момент времени  $t$ );

$\sigma(t)$  – функция защищенности, определяющая вероятность устойчивости системы к угрозам нарушения информационной безопасности в момент времени  $t$  (в данном случае под вероятностью мы понимаем оценку риска нарушения защиты документа).

Так как частные целевые функции и общая функция сохранности представляют собой вероятности, то целевая функция  $\mu(t)$  есть взвешенная аддитивная свертка частных целевых функций для исключения несоответствия [14].

Тогда целевая функция сохранности примет вид:

$$\mu(t) = \omega_1 \alpha(t) + \omega_2 \zeta(t) + \omega_3 \varphi(t) + \omega_4 \rho(t) + \omega_5 v(t) + \omega_6 \sigma(t). \quad (2)$$

Пусть  $\omega_i$  – весовые коэффициенты значимости соответствующей целевой функции  $i=[1,6]$ , (коэффициенты выбираются путем экспертной оценки для каждой конкретной реализации программно-аппаратной среды хранения электронных документов).

Должно строго выполняться условие:  $\sum \omega_i = 1$ . Если нет предпочтений ни для одной из целевых функций, коэффициенты  $\omega_i$  можно принять равными  $1/6$  для  $i=[1,6]$ .

Должна быть определена шкала значений (минимум 3-5 значений, хотя бы на вербальной шкале), определяющая градации рисков для оценки каждой целевой функции, значение которой не может быть оценено статистическими методами. Выбор шкал значений для каждой целевой функции, которая оценивается как

риск, приведен ниже при создании модели расчетов целевых функций.

Тем самым задача разбивается на частные задачи достижения максимума для каждой из целевых функций, что и позволяет проводить оценку текущего состояния сохранности электронных документов, как конкретного, так и всей совокупности в ЭА.

Автором разработаны отдельные математические модели для каждой функции и проверены экспериментально. В частности, модель для функции  $\rho(t)$  надежности прошла практическую проверку в рамках нескольких крупных проектов. Исключением является функция  $\sigma(t)$ , для которой автор не делал разработку отдельной математической модели, предлагая выполнять приблизительную оценку степени соответствия ЭА выявленным моделям угроз информационной безопасности.

## Заключение

Постановка задачи долговременной сохранности крайне необходима для создания комплексной технологии долговременного хранения электронных документов, которая позволит решить все проблемы долговременной сохранности на основе единого подхода.

После разработки общей постановки задачи необходимо далее разрабатывать математические модели и методы решения проблем долговременной сохранности, о которых подробно написано в работах [1, 3, 5].

Создание таких моделей и методов существенно упростит техническую разработку электронных архивов долговременной сохранности и позволит сэкономить массу ресурсов при реализации конкретной аппаратно-программной платформы архива долговременного хранения.

## Литература

1. Акимова Г.П., Пашкин М.А., Пашкина Е.В., Соловьев А.В. Электронные архивы: возможные решения проблем долговременного хранения данных // Труды Института системного анализа РАН (ИСА РАН), Том 63, выпуск 4, М.: 2013, С. 39-49.
2. Акимова Г.П., Пашкин М.А., Пашкина Е.В., Соловьев А.В. Проблемы доступности данных электронных архивов при долгосрочном хранении и возможные решения // Журнал «Системы высокой доступности» № 4 за 2013 г., М.: Радиотехника, 2013, С. 48-59.
3. Акимова Г.П., Пашкин М.А., Пашкина Е.В., Соловьев А.В. Проблемы долговременного хранения электронных документов // Труды XX Международной научно-практической конференции «Документация в информационном обществе: эффективное управление электронными документами» (Москва, РГАСПИ, 20-21 ноября 2013 г.).
4. Акимова Г.П., Пашкин М.А., Пашкина Е.В., Соловьев А.В., Соловьев Д.В. Электронные архивы: методологический подход к решению проблемы катастрофостойчивости при долговременном хранении // Труды Института системного анализа РАН (ИСА РАН), Том 64, выпуск 1, М.: 2014.
5. Акимова Г.П., Пашкин М.А., Пашкина Е.В., Соловьев А.В. Проблемы долговременного хранения электронных деловых документов // Журнал «Делопроизводство», №1, 2014.
6. ГОСТ Р 54989-2012 /ISO TR 18492:2005 Обеспечение долговременной сохранности электронных документов (вступил в силу с 01.05.2013).
7. Решение Совета глав государств СНГ: «О Концепции сотрудничества государств-участников Содружества Независимых Государств в сфере обеспечения информационной безопасности и о Комплексном плане мероприятий по реализации Концепции сотрудничества государств-участников Содружества Независимых Государств в сфере обеспечения информационной безопасности на период с 2008 по 2010 год» [Электронная публикация] ([http://official.academic.ru/61777/%D0%94%D0%BE%D1%81%D1%82%D1%83%D0%BF%D0%BD%D0%BE%D1%81%D1%82%D1%8C\\_%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%B8](http://official.academic.ru/61777/%D0%94%D0%BE%D1%81%D1%82%D1%83%D0%BF%D0%BD%D0%BE%D1%81%D1%82%D1%8C_%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%B8)).
8. Финансовый словарь «Финам» [Электронная публикация] ([http://dic.academic.ru/dic.nsf/fin\\_enc/22465](http://dic.academic.ru/dic.nsf/fin_enc/22465)).
9. ГОСТ Р 52292-2004. Информационная технология. Электронный обмен информацией. Термины и определения. Утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 29 декабря 2004 г. № 135-ст.
10. Федеральный закон Российской Федерации от 6 апреля 2011 г. N 63-ФЗ «Об электронной подписи».
11. Емельянов С.В. Системы автоматического управления с переменной структурой. – М.: Наука, 1967 г. – 336 с.
12. Емельянов С.В., Костылева Н. Е., Матич Б. Л., Миловидов Н. Н. Системное проектирование средств автоматизации. — М.: Машиностроение, 1978.
13. Емельянов С.В. Новые типы обратной связи. – М.: Наука, Физматлит, 1997. – 352 с.
14. Петровский А.Б. Теория принятия решений. – М.: Издательский центр «Академия», 2009. – 400 с.

**Соловьев Александр Владимирович.** Ведущий научный сотрудник ИСА РАН. Окончил МГТУ им. Н.Э. Баумана в 1994 году. Кандидат технических наук. Автор 41 печатной работы. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, влияние человеческого фактора, математическое моделирование, электронный документооборот, электронный архив. E-mail: alexsol@cs.isa.ru