

# Электронные архивы: разработка постановки для задачи обеспечения долговременной сохранности электронных деловых документов

А.В. СОЛОВЬЕВ

**Аннотация.** В статье сформулирована задача обеспечения долговременной сохранности электронных деловых документов как проблемы стабилизации объекта управления в условиях параметрических возмущений. Приведена математическая постановка задачи долговременной сохранности, основные ограничения и допущения. Статья служит для создания теоретической основы управления долговременным хранением электронных документов в динамически меняющейся программно-аппаратной среде.

**Ключевые слова:** электронный документооборот, электронный архив, система управления электронными документами, электронный документ, долговременное хранение, оптимальное управление.

## Введение

Современные реалии развития компьютерных технологий породили бум развития программно-аппаратных средств электронного документооборота (СЭД), регламентирующего жизнедеятельность организаций.

Практика показывает, что проблема долговременной сохранности электронных документов (ЭлД) является крайне актуальной. Поиск решения проблемы ведут все ведущие страны. Со временем количество электронных документов в мире будет стремительно возрастать, следовательно, порядок сложности решения задачи долговременной сохранности будет стремительно увеличиваться. Предположительно в ближайшие 5-10 лет вытеснение бумажных документов станет массовым явлением, и подходы к их хранению должны быть выработаны уже сейчас [1-5]. Фактически уже сейчас в Пенсионном Фонде РФ, Федеральной налоговой службе РФ, страховых компаниях во многих государственных и коммерческих организациях ЭлД начинают активно замещать документы бумажные.

Юридически ЭлД, согласно законодательным актам РФ (например, см. №379-ФЗ от 2014 года), становятся равнозначны документам бумажным. Однако только защитой ЭлД с помощью электронной подписи (ЭП) достичь автоматически долговременную сохранность ЭлД невозможно.

Задача обеспечения сохранности ЭлД видится крайне нетривиальной. Причина в том, что современная программно-техническая среда, т.е. набор

технических средств и программных средств – это крайне динамично меняющаяся во времени структура, работая с которой важно учитывать такие факторы как технологическое старение, износ, обновление программной среды, изменение форматов ЭлД, средств отображения ЭлД и т.д.

Сложившаяся противоречивая ситуация определяет необходимость решения важной научно-технической задачи обеспечения сохранности ЭлД, включая доступность, аутентичность (неизменность), интерпретируемость (читаемость) ЭлД в динамически меняющейся программно-аппаратной среде в течение всего длительного (годы, десятилетия) срока хранения.

Однако для того, чтобы задача была решена, она должна быть сформулирована. Иначе неправильная или неполная формулировка могут привести к необоснованным затратам и, в конечном итоге, к неразрешению проблемы сохранности.

## 1. Определения и основные понятия

*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).

*Долговременное 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) взято максимальное время хранения документов в оперативных архивах СЭД. Электронные документы могут храниться в течение десятилетий или даже столетий или «бессрочно» в зависимости от их важности.

*Сохранность* – свойство электронного документа существовать в качестве доступного и аутентичного свидетельства (доказательства) в произвольный момент времени.

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

*Доступность информации* – «возможность реализации беспрепятственного доступа к информации субъектов, имеющих на это надлежащие полномочия» [7] и, одновременное, «избежание 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].

*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. Разработка постановки для задачи обеспечения долговременной сохранности

Задача долговременной сохранности может быть сформулирована в общем виде следующим образом: необходимо создать условия для долговременного хранения электронных документов, при этом обеспечить аутентичность, интерпретируемость (читаемость), подтверждение авторства документа, а также надежность, устойчивость среды хранения документа к внешним воздействиям в течение всего срока хранения.

Для того, чтобы не учитывать историю состояний Элд до начала хранения, и в то же время искусственно не суживать возможные формы представления Элд, должны быть приняты следующие допущения:

- аутентичность документа на момент передачи его в архив 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;
- при долговременном хранении истекают сроки действия сертификатов ЭП;
- в течение срока хранения могут быть сняты с технической поддержки операционные системы (ОС) или их версии, например Windows, и прикладного программного обеспечения, обеспечивающего интерпретацию Элд.

Тем самым, 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 со средним сроком технологического хранения оборудования (обычно 5-7 лет), можно легко убедиться, что в течение такого срока хранения несколько раз будет обновлено как системное программное обеспечение, так и носители информации.

**Утверждение 1.** 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  $f$  для организации обратной связи для стабилизации объекта управления [13] (см. рис. 1).

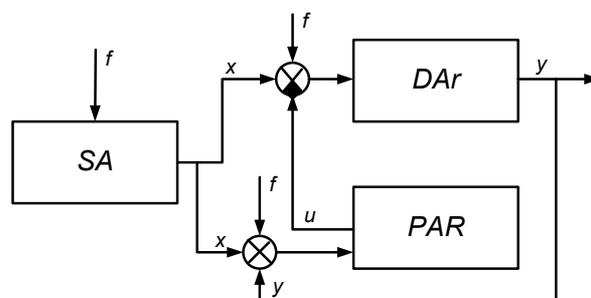


Рис. 1. Общая схема задачи обеспечения долговременной сохранности

### 3. О математической постановке для задачи обеспечения долговременной сохранности

Математическая постановка для задачи долговременной сохранности может быть сформулирована в следующем виде.

Дано:

- 1) *Объект управления*: множество ЭЛД  $D = \{DAr_i\}$ .
- 2) *Множество возмущений среды хранения  $D$*   $f = \{ \varepsilon(\alpha(t)), \varepsilon(\zeta(t)), \varepsilon(\varphi(t)), \varepsilon(\rho(t)), \varepsilon(v(t)), \varepsilon(\sigma(t)) \}$ , время  $t$ .

*Параметрические возмущения  $f$ :*

- $\varepsilon(\alpha(t))$  – нарушение аутентичности  $D$ ,
- $\varepsilon(\zeta(t))$  – нарушение интерпретируемости  $D$ ,
- $\varepsilon(\varphi(t))$  – изменения программно-аппаратной среды  $D$ ,
- $\varepsilon(\rho(t))$  – нарушение надежности хранения  $D$ ,
- $\varepsilon(v(t))$  – нарушение устойчивости хранения  $D$ ,
- $\varepsilon(\sigma(t))$  – нарушение информационной безопасности хранения  $D$ .

- 3) *Среда хранения SA (операционная система, техническое обеспечение, системы защиты информации) подверженная  $f$  и осуществляющая воздействие  $x$  на объект управления.*
- 4) *Множество требований к параметрам среды хранения  $T = \{ \alpha_T, \zeta_T, \varphi_T, \rho_T, v_T, \sigma_T \}$ .*

Найти:

- 1) *Множество математических моделей оценки параметров среды хранения, характеризующих сохранность  $D$ ,*  
 $\mu(t) = \{ \alpha(t), \zeta(t), \varphi(t), \rho(t), v(t), \sigma(t) \}$ ,  
 $y$  – сохранность  $D$ ,  
 $\mu(t)$  – функция оценки сохранности  $D$ .
- 2) *Множество алгоритмов  $A(t) = \{ A_i(t) \}$ , обеспечивающих стабилизацию объектов управления  $D$ , т.е. позволяющих при реализации обеспечивать такой сигнала управления  $u$ , при котором воздействия  $f$  и  $x$  компенсируются, и результирующая сохранность  $y$  находится в допустимых пределах.*
- 3) *Модель ЭЛД  $D = \{ DAr_i \}$ , описывающая ЭЛД с точки зрения состава информации, необходимой для долговременного хранения.*
- 4) *Программно-аппаратный регулятор PAR (ЭА), реализующий функции контроля параметрических возмущений  $f$ , действующих на объект управления через SA ( $x$ ), и стабилизации  $u$  объекта управления с помощью реализации алгоритмов  $A(t)$ .*
- 5) *Множество подсистем PAR (ЭА)  $\Pi = \{ \Pi_i \}$  и элементов  $\Theta = \{ \Theta_i \}$ .*
- 6) *Множество функций подсистем PAR (ЭА)  $\Phi = \{ \Phi_j \}$ .*

- 7) *Множество ограничений по критичности подсистем PAR (по времени выполнения  $T_d = \{ T_{dj} \}$  и по критичности выполнения функционала  $\Phi_d = \{ \Phi_{dj} \}$ ).*
- 8) *Методику оценки надежности функционирования среды хранения ЭЛД SA и PAR, модели и алгоритмы оценки устойчивости среды хранения  $D$ .*

### 4. Задача контроля сохранности объекта управления

**Утверждение 2.** Задача контроля сохранности электронного документа как объекта управления представляет собой проблему оптимального выбора по многим критериям.

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

Тогда можно утверждать, что проблема долговременной сохранности формулируется, как задача достижения максимума функции  $\mu(t)$  по  $t$  на произвольном временном интервале  $t$  (при  $t \rightarrow \infty$ ), т.е.

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

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

- $\alpha(t)$  – функция оценки аутентичности, значение которой определяет вероятность сохранения аутентичности документа в момент времени  $T$ ;
- $\zeta(t)$  – функция оценки интерпретируемости, значение которой определяет вероятность интерпретации и визуализации документа в произвольный момент времени  $T$  (в данном случае под вероятностью мы понимаем оценку риска неинтерпретируемости документа). Под риском здесь и далее будем понимать количественную оценку в диапазоне  $[0, 1]$ , которая определяет степень приближения значения функции к максимуму, т.е. к 1;
- $\varphi(t)$  – функция оценки отчуждаемости, значение которой определяет 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 не извлечения документа).

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.

Оценка по данной шкале выполняется экспертом (группой экспертов), т.е. лицом (лицами), обслуживающим ЭА и знакомым с программно-аппаратной средой функционирования ЭА (табл. 1).

**Табл. 1**

Пример вербальной шкалы для приблизительной оценки вероятности извлечения или интерпретируемости ЭЛД

№	Наименование значений	Значение
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

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 персонал, обслуживающий ЭА, 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;

$\rho(t)$  – функция оценки надежности, значение которой определяет вероятность работоспособности программно-аппаратной среды хранения электронного документа в произвольный момент времени  $T$ ;

$\psi(t)$  – функция оценки устойчивости, значение которой определяет степень устойчивости программно-аппаратной среды хранения к внешним воздействиям в момент времени  $T$  (например, вероятность восстановления аппаратно-программной среды в момент времени  $T$ );

$\sigma(t)$  – функция оценки защищенности, значение

которой определяет вероятность устойчивости системы к угрозам нарушения информационной безопасности в момент времени  $T$  (в данном случае под вероятностью мы понимаем оценку риска нарушения защиты документа);  
 $t_0$  – момент времени создания ЭЛД.

Так как частные функции представляют собой вероятности, и общая функция сохранности также является вероятностной функцией, то необходимо представить функцию  $\mu(t)$  как взвешенную аддитивную свертку частных функций [14].

Тогда функция контроля сохранности примет вид:

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

Пусть  $\omega_i$  – весовые коэффициенты значимости соответствующей 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 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 функции, значение которой не может быть оценено статистическими методами.

Тем самым задача контроля сохранности декомпозируется на частные задачи достижения максимума для каждой из функций в (2).

Автором разработаны отдельные математические модели для каждой функции оценки параметров среды хранения, которые проверены экспериментально. В частности математическая модель для функции надежности  $\rho(t)$  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

Приведенная разработка постановки для задачи долговременной сохранности крайне необходима для создания комплексной технологии долговременного хранения электронных документов, которая, в свою очередь, позволит решить все проблемы долговременной сохранности на основе

единого комплексного подхода.

После разработки общей постановки задачи необходимо далее разрабатывать математические модели и методы решения частных проблем долговременной сохранности, о которых, в частности, подробно написано в работах автора [2-4].

Разработка единого комплексного подхода к обеспечению сохранности существенно упростит техническую разработку электронных архивов долговременной сохранности и позволит сэкономить массу ресурсов при реализации конкретной аппаратно-программной платформы архива долговременного хранения.

### Литература

1. Соловьев А.В. Электронные архивы: возможные решения проблем долгосрочного хранения данных / Акимов Г.П., Пашкин М.А., Пашкина Е.В., Соловьев А.В. // Труды ИСА РАН. Т. 63. Вып. 4. М. 2013. С. 39-49.
2. Соловьев А.В. Решение проблем оценки и сохранения аутентичности электронных документов при долговременном хранении // Журнал «Системы высокой доступности». №4. Т. 10. М.: Радиотехника. 2014. С. 99-106.
3. Соловьев А.В. Методология организации долговременного хранения электронных деловых документов // Труды XXI Международной научно-практической конференции «Документация в информационном обществе: нормативно-методическое обеспечение управления документами» (М. РГАСПИ. 18-19 ноября 2014 г.). М.: ВНИИДАД. 2015. С. 320-324.
4. Соловьев А.В. Электронные архивы: методологический подход к решению проблемы катастрофоустойчивости при долговременном хранении / Акимов Г.П., Пашкин М.А., Пашкина Е.В., Соловьев А.В., Соловьев Д.В. // Труды ИСА РАН. Т. 64. Вып. 3. М. 2014. С. 91-98.
5. Соловьев А.В. Электронные архивы: проблема определения понятия и характеристик электронного документа, как объекта долговременного хранения / А.В. Соловьев, А.С. Богданов // Информационные технологии и вычислительные системы. №4. 2016. С. 24-32.
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» / [Электронный ресурс] – 2008 – Режим доступа: [http://official.academic.ru/6177/%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/6177/%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» [Электронный ресурс] – 2015 – Режим доступа: [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. № 63-ФЗ «Об электронной подписи».
11. Емельянов С.В. Системы автоматического управления с переменной структурой. М.: Наука. 1967. – 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.** Заместитель директора ИСА ФИЦ ИУ РАН. Окончил в 1994 г. МГТУ им. Н.Э. Баумана. Доктор технических наук. Количество печатных работ: 55. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, математическое моделирование, электронный документооборот, электронный архив, долговременное хранение электронных документов. E-mail: [soloviev@isa.ru](mailto:soloviev@isa.ru)

## Electronic archives: the development of a formulation for the problem of ensuring the long-term preservation of electronic business documents

A.V. Solovyev

**Abstract.** The article discusses the development of a formulation for the problem of ensuring the long-term preservation of electronic business documents how the problem of stabilization of control object under parametric perturbations. This article is a continuation of creating a theoretical framework for the management of long-term storage of electronic documents in a dynamically changing software and hardware environment.

**Keywords:** *electronic document management, electronic archives, electronic document management system, electronic document, long-term storage, optimal control.*

### References

1. Solovyev A.V. Electronic archives: possible solutions to the problems of long-term data storage / G.P. Akimova, E.V. Pashkina, M.A. Pashkin, A.V. Solovyev // Proceedings of Institute of system analysis RAS (ISA RAS) – 2013 – T.63, Part.4 – P.39-49.
2. Solovyev A.V. The problems of assessment and conservation of authenticity of electronic documents for long term storage / A.V. Solovyev // High availability systems. – 2014 – №4, Part.10 – P.99-106.
3. Solovyev A.V. The methodology of the organization for long-term storage of electronic business documents / A.V. Solovyev // Proceedings of the XXI International scientific and practical conference “Documentation in information society: regulatory and methodological support document management” (Moscow, 18-19 November 2014) – 2015 – P.320-324.
4. Solovyev A.V. Electronic archives: methodological approach to the problem of fault tolerance for long term storage / G.P. Akimova, E.V. Pashkina, M.A. Pashkin, A.V. Solovyev, D.V. Solovyev // Proceedings of Institute of system analysis RAS (ISA RAS) – 2014 – T.64, Part.3 – P.91-98.
5. Solovyev A.V. Electronic archives: the problem of definition and characteristics of the electronic document as an object of long-term storage / A.V. Solovyev, A.S. Bogdanov // Information technology and computer systems. – 2016 – №4 – P.24-32.
6. GOST R 54989-2012 /ISO TR 18492:2005 Ensuring long-term preservation of electronic documents (entered into force with 01.05.2013).
7. The decision of the Council of CIS heads of state: “On the concept of cooperation of States-participants of the Commonwealth of Independent States in the sphere of information security and Complex plan of measures on realization of the concept of cooperation of States-participants of the Commonwealth of Independent States in the sphere of ensuring information security for the period from 2008 to 2010” [Electronic resource] – 2008 – Access mode: [http://official.academic.ru/6177/%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/6177/%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. *Financial dictionary “Finam”* [Electronic resource] – 2008 – Access mode: [http://dic.academic.ru/dic.nsf/fin\\_enc/22465](http://dic.academic.ru/dic.nsf/fin_enc/22465)
9. GOST R 52292-2004 Information technology. The electronic exchange of information. Terms and definitions. Approved and put into effect by the Federal Agency for technical regulation and Metrology 29 december 2004. № 135-st.
10. *Federal law of the Russian Federation 6 april 2011. №63-FZ «On the electronic signature».*
11. *Emelyanov S.V.* Automatic control systems with variable structure. – M.: Science, 1967 – 336p.
12. *Emelyanov S.V., Kostileva N.E., Matich B.L., Milovidov N.N.* System design automation. – M.: Mechanical engineering, 1978.
13. *Emelyanov S.V.* New types of feedback. – M.: Science, Fizmatlit, 1997 – 352p.
14. *Petrovski A.B.* The theory of decision-making. – M.: Publishing center “Academy”, 2009 – 400p.

**Solovyev Alexandr Vladimirovich** Deputy Director ISA FRC CSC RAS. BMSTU 1994. Number of publications: 55. Area of scientific interests: system analysis, database management system, reliability theory, mathematical modeling, electronic document management, electronic archive, long-term storage of electronic documents. E-mail: [soloviev@isa.ru](mailto:soloviev@isa.ru)