

Система показателей для количественной оценки сохранности электронных документов

А.В. СОЛОВЬЕВ

Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

Аннотация. В статье предложена система показателей, с помощью которой можно количественно оценивать сохранность цифровых данных и электронных документов различных типов. Рассмотрен краткий обзор проблемы, основных рисков долговременной сохранности. Выполнена формальная постановка задачи определения системы показателей для оценки долговременной сохранности. Преимуществом предложенной системы показателей является их простота и унификация с существующими расчетами стандартизованных показателей надежности.

Ключевые слова: долговременная сохранность, цифровые данные, электронные документы, надежность, показатели сохранности.

DOI: 10.14357/20790279240105 **EDN:** RWDIZU

Введение

При организации долговременного хранения цифровых данных и электронных документов при цифровой трансформации экономики и общества, при которой данные в цифровой форме становятся ключевым фактором производства (см. [1]), крайне важно комплексно решить проблему сохранности данных и документов, обеспечив аутентичность, интерпретируемость, надежность и безопасность их хранения. Технология долговременного хранения описана автором в виде совокупности математических моделей и алгоритмов в работе [2].

Однако важно отметить, что нигде в современных исследованиях не поставлен вопрос, как оценить сохранность документов при организации долговременного хранения. Не предложена система показателей. Это, однако, становится крайне актуальным при возникновении необходимости хранить электронные документы и цифровые данные в течение десятилетий. Полагаться только на надежность технических средств вряд ли имеет смысл, т.к. даже самые надежные технические средства могут не обеспечить такие важные показатели сохранности электронных документов как аутентичность и интерпретируемость.

Разработке системы показателей, с помощью которой можно будет количественно оценивать сохранность цифровых данных и электронных документов, посвящена данная статья.

1. Краткий обзор источников и проблемы сохранности

Как было отмечено выше, автор не нашел в современных исследованиях, посвященных проблеме долговременной сохранности электронных документов системы количественных показателей, позволяющих количественно оценить сохранность. Весь упор делается на то, что техническое обеспечение обеспечит надежное хранение данных, организационные меры и средства защиты информации защитят их от злоумышленников. Однако, когда срок хранения электронных документов существенно превышает срок жизни технических и программных средств, возникает проблема миграции документов, а, следовательно, появляется дополнительная угроза их сохранности.

С другой стороны, в работах, посвященных проблеме надежности (см., например, [3-7]) не рассматривается проблема долговременной сохранности электронных документов и цифровых данных. Хотя и затрагиваются вопросы надежности программного обеспечения (см. [6, 7]).

Если обратиться к нормативным документам по надежности (см., например, [8]), то в это понятие включены долговечность и сохраняемость. Долговечность определяется как «свойство объекта, заключающееся в его способности выполнять требуемые функции в заданных режимах и условиях использования, технического обслуживания

и ремонта до достижения предельного состояния». Сохраняемость как «свойство объекта сохранять способность к выполнению требуемых функций после хранения и (или) транспортирования при заданных сроках и условиях хранения и (или) транспортирования».

Ни то, ни другое определение не подходит для электронных документов, для которых невозможно определить такие понятия как «предельное состояние», «транспортировка», «техническое обслуживание». Хотя, безусловно, можно попытаться перефразировать терминологию [8] для электронных документов. Есть ли в этом смысл – другой вопрос.

Так, в нормативных документах, посвященных долговременной сохранности электронных документов, подобные термины не вводятся (см., например, [9]). Но и показатели сохранности, в отличие от нормативных документов по надежности, в которых показатели надежности определены, не вводятся.

Тогда возникает противоречивая ситуация. С одной стороны проблема долговременной сохранности электронных документов важна и актуальна. С другой стороны нет разработанного математического аппарата для количественной оценки сохранности документов.

Данная работа посвящена разработке системы показателей сохранности электронных документов и цифровых данных, которые позволяли бы количественно оценить их сохранность.

2. Обзор основных рисков долговременного хранения

Необходимо отметить, что в первую очередь риск повреждения или утраты документов при долговременном хранении связан с несоблюдением условий хранения, которые стандартизованы и подробно описаны, например, в [10].

Кроме этого, есть проблема с долговечностью цифровых носителей информации (см., например, [11, 12]).

Если кратко резюмировать суть этих проблем, то можно утверждать, что усредненные данные по срокам хранения цифровых носителей информации при соблюдении условий хранения – 10–30 лет (CD/DVD-диски, SSD- и HDD-диски, флэш-накопители, магнитные ленты и др.). Не рассмотрены такие носители, как, например, считающиеся перспективными, кварцевые цифровые носители по причине их малой распространенности, дороговизны в настоящее время.

Еще одной проблемой долговременной сохранности является проблема долговечности ин-

формационного слоя, формирующего документ на соответствующем носителе. Имеется в виду проблема с долговечностью магнитных, оптических и иных слоев, собственно представляющими содержание информации на соответствующих носителях.

Рассмотрим типовые негативные влияния разрушающих факторов ($Des = \{dd_i\}$), которые воздействуют на электронные документы при организации долговременного хранения:

- Воздействие электромагнитных полей (dd_1).
- Потеря аутентичности (dd_2) из-за несанкционированного изменения документа.
- Неинтерпретируемость (невозможность интерпретировать последовательность бит, составляющих электронный документ) (dd_3), например, из-за отсутствия через продолжительное время программно-технических средств, позволяющих декодировать формат хранения документа.
- Воздействие изменений программно-аппаратной среды хранения (dd_4), приводящих к невозможности использовать документ.

Подробнее о проблемах долговременной сохранности цифровых данных и электронных документов см. работу автора [1].

Тем не менее, при организации долговременного хранения ценных и особо ценных документов нужно учитывать тот факт, что их потеря или повреждение могут обернуться колоссальными издержками: политическими, экономическими, имидживыми. Это, в свою очередь, может привести к негативным последствиям и затратам, которые могут превысить затраты на создание программно-технических решений обеспечения долговременной сохранности.

Сохранение ценных документов (электронных, бумажных и иных) – важная государственная задача, которая в условиях цифровизации должна решаться на качественно новом уровне.

Приведем постановку задачи определения системы показателей для оценки долговременной сохранности электронных документов и цифровых данных.

3. Формализация задачи определения системы показателей для оценки долговременной сохранности

Перед тем, как перейти к формальной постановке задачи обеспечения долговременной сохранности, введем ряд важных определений. Долговременное хранение – обеспечение сохранности документа в течение не менее 5 лет. Сохранность

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

При переходе к цифровой экономике сроки хранения документов могут намного превосходить сроки функционирования программно-аппаратной среды их хранения. Так, например, документы по личному составу должны храниться 50 лет (до 2003 года – 75 лет). Особо ценные документы могут храниться еще дольше.

Сложность решения задачи организации долговременного хранения документов состоит в том, что при их создании мало кто задумывается, что документ предстоит сохранять в течение длительного времени в неблагоприятной среде хранения. Электронная среда хранения подвержена частым изменениям (обновления программного обеспечения, смена аппаратной платформы, необходимость миграции данных с одного носителя на другой). Тем самым на электронный документ влияет целый ряд неблагоприятных факторов, способствующих его разрушению или порче.

В этой связи можно сформулировать задачу организации долговременного хранения как задачу стабилизации объекта хранения (электронного документа) в условиях среды хранения, подверженной параметрическим возмущениям, нарушающим стабильность объекта хранения.

Тогда математическая постановка задачи может быть сформулирована в следующем виде.

Дано:

- Множество электронных документов долговременного хранения $Doc = \{ d_i \}$.
- Сохранность электронных документов на момент передачи на долговременное хранение $SV_0 = \{ SV_{0i} \}$.
- Множество типов документов долговременного хранения $Typ = \{ dt_i \}$.
- Множество разрушающих факторов (воздействий) на документы $Des = \{ dd_i \}$.

Подробнее типы документов $Typ = \{ dt_i \}$ и множества требований к условиям хранения описаны в [10]. Формат данной статьи не позволяет привести их полностью.

Найти:

- Множество показателей оценки долговременной сохранности $MQual = \{ mq_i \}$.

Перейдем к решению поставленной задачи.

4. Разработка системы показателей для оценки долговременной сохранности

Специальных показателей, кроме выдерживания сроков хранения, для долговременного хранения не существует.

Поэтому автор предлагает систему показателей ($MQual$), во многом аналогичную показателям надежности в технических системах согласно основным международным стандартам:

- IEC 60300-3-11. Dependability management. Part 3-11: Application guide: Reliability centred maintenance.
- IEC 60319. Presentation and specification of reliability data for electronic components.
- ISO 2394. General principles on reliability for structures.
- ГОСТ 27.102-2021. Надежность в технике.

Для сходства расчетов показателей надежности, таких как долговечность и сохраняемость, автор предлагает систему показателей, которая является новой и вводится впервые. Она разработана на основе опыта автора по созданию программных средств долговременного хранения. В первую очередь – электронного архива персонифицированного учета Пенсионного фонда РФ, в разработке которого автор принимал участие на протяжении 10 лет. В составе архива обеспечивается хранение электронных документов сроком до 75 лет. Приведем разработанные автором показатели сохранности и правила их расчетов ($MQual = \{ mq_i \}$):

- исходная сохранность документа (mq_1) – сохранность электронного документа на момент передачи на долговременное хранение (SV_{0i}). Диапазон значений $[0, 1]$. Характеризует исходное состояние документа d_i . Соответственно, для всех типов документов ($Typ = \{ dt_i \}$): $SV_0 = \{ SV_{0i} \}$;
- назначенный срок хранения документа (mq_2), определяемый как календарная продолжительность поддержания сохранности документа типа dt_i в заданных условиях хранения dc_i на уровне SV_{0i} : T_{dsi} . Соответственно, для всех типов документов ($Typ = \{ dt_i \}$): $T_{ds} = \{ T_{dsi} \}$;
- назначенный срок восстановления документа (mq_3), определяемый как календарная продолжительность обнаружения проблемы и восстановления документа типа dt_i в заданных условиях хранения dc_i до уровня сохранности SV_{0i} : T_{dri} . Соответственно, для всех типов документов $Typ = \{ dt_i \}$: $T_{dr} = \{ T_{dri} \}$;
- вероятность сохранности документа (mq_4), определяемая как статистическая оценка вероятности достижения документом типа dt_i в заданных условиях хранения dc_i назначенного срока хранения T_{dsi} : $P(T_{dsi}) = n_i/N_i$ (n_i – количество документов типа dt_i остающихся сохранными, т.е. в состоянии исходной сохранности SV_{0i} на периоде времени T_{dsi} , N_i – общее количество докумен-

- тов типа dt_i). Соответственно, для всех типов документов ($Typ = \{ dt_i \}$): $P(T_{ds}) = \{ P(T_{dsi}) \}$;
- вероятность восстановления документа (mq_5), определяемая как статистическая оценка вероятности обнаружения проблемы и восстановления (реставрации) документа за время не превышающее T_{dri} , относящегося к типу dt_i : $P(T_{dri}) = nv_i / Nv_i$ (nv_i – количество документов типа dt_i , которые удастся восстановить до состояния исходной сохранности SV_{0i} на периоде времени T_{dri} , Nv_i – общее количество документов типа dt_i). Соответственно, для всех типов документов ($Typ = \{ dt_i \}$): $P(T_{dr}) = \{ P(T_{dri}) \}$;
 - средний срок сохранности документа (mq_6), определяемый как математическое ожидание времени, в течении которого документ, относящийся к типу dt_i , остается полностью сохранен на уровне SV_{0i} : $T_{dsvi} = (\sum_{j=1, Ni} P_j(T_{dsi}) T_{dsij}) / N_i$ (где $P_j(T_{dsi})$ – вероятность сохранности j -го документа типа dt_i , T_{dsij} – время обнаружения проблемы и восстановления j -го документа типа dt_i). Соответственно, для всех типов документов ($Typ = \{ dt_i \}$): $T_{dsv} = \{ T_{dsvi} \}$;
 - среднее время восстановления документа (mq_7) определяемое как математическое ожидание времени обнаружения проблемы и восстановления (реставрации) документа, относящегося к типу dt_i : $T_{drvi} = (\sum_{j=1, Ni} P_j(T_{dri}) T_{drij}) / N_i$ (где $P_j(T_{dri})$ – вероятность восстановления j -го документа типа dt_i , T_{drij} – время обнаружения проблемы и восстановления j -го документа типа dt_i). Соответственно, для всех типов документов ($Typ = \{ dt_i \}$): $T_{drv} = \{ T_{drvi} \}$;
 - гамма-процентный срок сохранности документа (mq_8), определяемый как срок сохранности документа типа dt_i в заданных условиях хранения dc_i на уровне SV_{0i} , достигаемый с заданной вероятностью «гамма» (γ): $P_\gamma(t) = [t = [T_{dsi}, \infty)) P(T_{dsi}) dt = \gamma / 100$. Соответственно, для всех типов документов ($Typ = \{ dt_i \}$): $P_\gamma(t) = \{ P_{\gamma i}(t) \}$.
- В качестве основных показателей, опираясь на рекомендации стандартов [8, 9], можно определить следующие:
- средний срок сохранности документа (mq_6) $T_{dsv} = \{ T_{dsvi} \}$, если потеря сохранности документа не может привести к катастрофическим (материальным или нематериальным) последствиям;
 - гамма-процентный срок сохранности документа (mq_8) $P_\gamma(t) = \{ P_{\gamma i}(t) \}$, если потеря сохранности документа может привести к катастрофическим (материальным или нематериальным) последствиям;

- среднее время восстановления документа (mq_7) $T_{drv} = \{ T_{drvi} \}$ – во всех случаях.

Предложенная система показателей еще требует уточнения и отладки, которая предполагается по мере создания проектов долговременного хранения электронных документов. Преимуществом предложенной системы показателей является их простота и унификация с существующими расчетами стандартизованных показателей надежности.

Заключение

В статье предложена система показателей, с помощью которой можно количественно оценивать сохранность цифровых данных и электронных документов различных типов.

Рассмотрен краткий обзор проблемы и источников литературы, проведен обзор основных рисков долговременного хранения, выполнена формальная постановка задачи определения системы показателей для оценки долговременной сохранности.

Приведен результат разработки системы показателей, определены основные и вспомогательные показатели. Преимуществом предложенной системы показателей является их простота и унификация с существующими расчетами стандартизованных показателей надежности.

В дальнейшем предложенная система показателей будет уточняться по мере создания и развития проектов долговременного хранения электронных документов.

Литература

1. Программа «Цифровая экономика Российской Федерации». Утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р. М.: 2017. 88 с.
2. Solovyev A.V. Long-Term Digital Documents Storage Technology // Lecture Notes in Electrical Engineering. 2020. Vol.641. Pp.901-911. ISSN 1876-1100. doi: 10.1007/978-3-030-39225-3_97.
3. Крылов Е.Г. Оценка показателей надежности автоматизированных систем. ВолгГТУ. Волгоград. 2018. 80 с.
4. Булатов В.В. Оценка надежности автоматизированных систем в процессе эксплуатации // Промышленные АСУ и контроллеры. 2021. № 6. С. 3-7.
5. Самохвалов А.А., Слабуха В.Н., Таранов А.И. Многоуровневая оценка надежности техники связи и автоматизированных систем управления по результатам испытаний на стадии эксплуата-

- тации // Проблемы технического обеспечения войск в современных условиях. Труды VII межвузовской научно-практической конференции. Санкт-Петербург. 2022. С.237-245.
6. Бутакова М.А., Гуда А.Н., Гнаденберг В.С. Методы оценки надежности и технологической безопасности управляющего программного обеспечения автоматизированных систем управления на железнодорожном транспорте // Вестник Ростовского государственного университета путей сообщения. 2011. №3(43). С. 22-32.
 7. Ретин С.И. Оценка качества программного обеспечения автоматизированных систем на основе одноступенчатого плана испытаний // I-methods. 2021. № 13 (1).
 8. ГОСТ 27.102-2021. Надежность в технике. Надежность объекта. Термины и определения. Введ. 2022–01–01. М.: ФГБУ «РСТ». 2021. 36 с.
 9. ГОСТ Р 54989-2012 (ISO/TR 18492:2005). Обеспечение долговременной сохранности электронных документов. Введ. 2013–05–01. М.: Стандартинформ. 2013. 23 с.
 10. Правила организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, научных организациях // Федеральное архивное агентство, 2022, URL: <https://archives.gov.ru/documents/rules/pravila-2020.shtml> (дата обращения 09.01.2024).
 11. Шавырова М.В., Грибков Д.Н. Современные технологии обеспечения сохранности документов // Электронное информационное пространство для науки, образования, культуры. Материалы VI всероссийской научно-практической конференции. Научный редактор и составитель Д.Н. Грибков. 2019. С. 218-224.
 12. Муродова С.Н. Продолжается погоня за носителями данных сверхвысокой емкости, которые переживут наше время // Современные технологии: актуальные вопросы теории и практики. Сборник статей II Международной научно-практической конференции. Пенза. 2022. С. 36-38.

Соловьев Александр Владимирович. Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» г. Москва, Россия. Главный научный сотрудник. Доктор технических наук. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, математическое моделирование, долговременное хранение электронных документов. E-mail: soloviev@isa.ru

System of indicators for quantitative assessment of the safety of electronic documents

A.V. Solovyev

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Moscow, Russia

Abstract. The article proposes a system of indicators with which it will be possible to quantitatively assess the safety of digital data and electronic documents of various types. A brief overview of the problem, an overview of the main risks of long-term safety, and a formal formulation of the problem of determining a system of indicators for assessing long-term safety were carried out. The advantage of the proposed system of indicators is their simplicity and unification with existing calculations of standardized reliability indicators.

Keywords: *long-term preservation, digital data, electronic documents, reliability, safety indicators*

DOI: 10.14357/20790279240105 **EDN:** RWDIZU

References

1. The Government of the Russian Federation. 2017. PROGRAM “Digital Economy of the Russian Federation” approved by the order of at July 28, 2017. № 1632-r: 88.
2. Solovyev A.V. 2020. Long-Term Digital Documents Storage Technology. Lecture Notes in Electrical Engineering. 641: 901-911. doi: 10.1007/978-3-030-39225-3_97.
3. Krylov E.G. 2018. Assessment of indicators of reliability of automated systems. VolgGTU. Volgograd, 80 p.
4. Bulatov V.V. 2021 Assessment of the reliability of automated systems during operation. Industrial ACS and controllers, 6: 3-7.
5. Samokhvalov A.A., Slabukha V.N., Taranov A.I. 2022. Multilevel assessment of the reliability of

- communication technology and automated control systems based on the results of tests at the operational stage. Problems of technical support of troops in modern conditions. Proceedings of the VII interuniversity scientific-practical conference: 237-245.
6. Butakova M.A., Guda A.N., Gnadenberg V.S. 2011. Methods for assessing the reliability and technological safety of the control software for automated control systems in railway transport. Bulletin of the Rostov State University of Communications, 3(43): 22-32.
 7. Repin S.I. 2021. Assessing the quality of software for automated systems based on a single-stage test plan. I-methods, 13 (1).
 8. GOST 27.102-2021. 2021. Reliability in technology. object reliability. Terms and Definitions. FGBU "RST". 36 p.
 9. GOST R 54989-2012 (ISO/TR 18492:2005). 2013. Ensuring long-term safety of electronic documents. Standartinform. 23 p.
 10. Federal Archival Agency (ROSARCHIV). 2022. Rules for the Organization of Storage, Acquisition, Accounting and Use of Documents of the Archival Fund of the Russian Federation and Other Archival Documents in State and Municipal Archives, Museums and Libraries, Scientific Organizations. Available online: <https://archives.gov.ru/documents/rules/pravila-2020.shtml> (accessed 09.01.2024).
 11. Shavyrova M.V., Gribkov D.N. 2019. Modern technologies for ensuring the safety of documents. Electronic Information Space for Science, Education, Culture, Proceedings of the Materials of the VI All-Russian Scientific and Practical Conference: 218–224.
 12. Murodov S.N. 2022. The race continues for ultra-high-capacity storage media that will outlive our time. Modern Technologies: Topical Issues of Theory and Practice, Proceedings of the Collection of Articles of the II International Scientific and Practical Conference: 36–38.

Solovyev A.V. Chief Researcher, Doctor of Technical Sciences. Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, 44/2 Vavilova str., Moscow, 119333, Russia. E-mail: soloviev@isa.ru