

Решение проблемы интерпретации цифровых данных долговременного хранения*

А.В. СОЛОВЬЕВ¹

¹ Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

Аннотация. В условиях стремительной цифровизации экономики возрастает как объем цифровых данных, так и их ценность. Искажение этих данных, безусловно, недопустимо и приведет к большим материальным и нематериальным потерям. Одной из проблем при работе с цифровыми данными, особенно при их долговременном хранении, является задача интерпретации данных. Предложено решение проблемы интерпретации цифровых распределенных данных долговременного хранения с помощью разработки формата долговременного хранения, который наряду с математической моделью цифровых данных долговременного хранения позволяет решить проблему интерпретируемости. Проведен обзор существующих форматов цифровых данных. Определены их достоинства и недостатки. Доказана необходимость создания собственного формата, который бы учитывал, с одной стороны, разнородность цифровых данных, с другой – позволял организовать доступ к отдельным частям распределенных данных, с третьей – позволял легко извлекать данные в случае выхода из строя или отсутствия программных средств интерпретации. В качестве практического результата в рамках проведенного исследования создан формат долговременного хранения на основе подмножества языка XML. Созданный формат прошел практическую апробацию. В дальнейшем планируется его более широкое внедрение и модернизация.

Ключевые слова: цифровые данные, долговременное хранение, файловый формат, электронный документ, цифровизация, цифровая экономика.

DOI: 10.14357/20790279210206

Введение

В целом ряде статей автора, посвященных проблеме долговременного хранения (см., например, [1–3]), было показано, что проблема долговременной сохранности цифровых данных является крайне актуальной.

В частности, в работе [1] описана проблема интерпретируемости цифровых данных в новых информационных условиях. Решение проблемы интерпретируемости в общем виде подразумевает наличие возможности раскодировать формат цифровых данных через десятилетия и представить их в понятном для человека виде (отобразить на экране, распечатать).

Если цифровые данные интерпретируются на небольшом временном интервале, например, пять-десять лет в локальной цифровой среде (например, масштаба предприятия), то в этом случае проблема интерпретации не представляется достаточно острой. Смена форматов и исчезновение

программно-технических средств интерпретации здесь маловероятна.

Совсем другое дело, когда данные в рамках цифровизации становятся распределенными и должны быть интерпретированы не только в рамках локальной цифровой среды, а в экономике в целом и не на коротком временном отрезке, а в течение десятилетий. Тогда проблема интерпретации становится крайне актуальной.

Отсутствие стратегии в вопросе обеспечения интерпретируемости данных при долговременном хранении и информационном обмене в рамках цифровой экономики может привести к тому, что хранилища цифровых данных превратятся в склады разноформатных бинарных последовательностей байтов. Это может привести к тому, что часть информации невозможно будет раскодировать из-за отсутствия (устаревания) средств интерпретации хранимых форматов данных, а также из-за утери описания хранимых форматов, в случае использования закрытых проприетарных форматов.

* Работа выполнена при частичной финансовой поддержке РФФИ в рамках научных проектов № 18-29-03070 и № 18-29-03085.

Часто приходится сталкиваться с мнением, что все современные форматы уже стандартизировались (см., например, [4–6]). Однако практически ни один из разработчиков не дает гарантии того, что данный формат будет поддерживаться в течение нескольких последующих десятилетий.

1. Краткий обзор цифровых форматов

Итак, для решения задачи интерпретации при долговременном хранении цифровых распределенных данных необходимо использовать форматы, гарантия поддержки которых со стороны производителя существует на протяжении хотя бы нескольких десятилетий.

Важным требованием к цифровому формату является также то, что он должен быть хорошо документированным, желателен открытый, для него должны существовать стандартные программные средства интерпретации.

К таким форматам, которые в настоящее время могут предположительно использоваться для долговременного хранения, можно отнести, в первую очередь, следующие:

- ODF (Open Document Format) – открытый формат файлов документов для хранения и обмена редактируемыми офисными документами. Основан на XML (eXtensible Markup Language – расширяемый язык разметки, рекомендованный Консорциумом Всемирной паутины (W3C)). В 2006 году формат принят в качестве международного стандарта ISO/IEC 26300:2006. В РФ принят в качестве государственного стандарта ГОСТ Р ИСО/МЭК 26300-2010;
- PDF/A (Portable Document Format / A (Archive)) – открытый документированный формат, который позиционируется как формат для долговременного хранения (гарантия фирмы Adobe – 50 лет поддержки формата), является подмножеством формата PDF. Принят в качестве стандарта ISO 19005-1:2005. Существует более 15 лет;
- TIFF (*Tagged Image File Format*) – формат хранения растровых изображений, поддерживает сжатие без потерь, существует уже около 35 лет. Формат хорошо документирован, имеет достаточно простую структуру. Хорошо сжимает черно-белые изображения (TIFF Group 4) без потерь качества;
- JPEG (*Joint Photographic Experts Group*) – формат хранения изображений. Алгоритм сжатия с потерями и без потерь. Наилучшим образом подходит для сжатия цветных изображений с плавными переходами яркости и цвета. Существует более 25 лет. Принят в качестве стан-

дарта ISO/IEC 10918-1. Описана спецификация формата;

- PNG (*Portable Network Graphics*) – формат хранения изображения со сжатием без потерь. Свободный, открытый формат изображений. Принят как стандарт ISO/IEC 15948:2004. Существует свободно распространяемая библиотека для работы с форматом libpng.

В данный список не включен популярный формат RTF (*Rich Text Format*), существующий уже более 30 лет. При создании формат позиционировался как стандартный универсальный, простой для интерпретации для текстовых офисных документов. Однако формат является проприетарным. К тому же в процессе развития формата было создано большое количество версий, которые могут корректно интерпретироваться только в узком наборе программного обеспечения, как, например, версия для MS Word. Несмотря на хорошую документированность, RTF формально не является открытым форматом.

Для долговременного хранения, в принципе, можно использовать и популярный офисный формат DOC, тем более что фирма Microsoft открыла спецификацию формата [7]. Однако формат не является свободным, кроме того, он достаточно сложен для интерпретации.

Интересными с точки зрения долговременного хранения является формат WebP. Он предназначен для хранения графических изображений, по оценке фирмы производителя, на 25-34% компактнее изображений формата JPEG [8]. WebP является открытым, свободно распространяемым форматом, однако существуют сравнительно недолгое время (десять лет). Имеется свободно распространяемая библиотека для работы с форматом libvpx.

Отдельной проблемой является задача долговременного хранения видео- и аудио-документов, презентаций, анимационных файлов, программного кода, исполняемых файлов и их компонентов. Хотя рассмотрение этих форматов пока выходит за рамки данного исследования, можно привести общую рекомендацию. Для долговременного хранения видео- и аудио-документов также необходим выбор простых, открытых форматов. Возможно, это открытый формат контейнера видео WebM или существующие достаточно продолжительное время форматы AVI и MPEG, рекомендованные крупнейшим электронным архивом NARA (National Archives and Records Administration, USA) [6].

Проведем краткое сравнение форматов текстовых и графических документов с точки зрения долговременного хранения. Для этого введем следующие показатели:

- продолжительность использования (лет существования формата) – 1. ПИсп;
- заявленная поддержка фирмы-изготовителя (лет) – 2. ЗаявПод;
- наличие «текстовой» структуры, позволяющей извлечь текст в условиях отсутствия средств интерпретации – 3. ТкстСтр;
- наличие документации на спецификацию – 4. НСпДок;
- возможность свободного использования – 5. СвобИсп;
- наличие развитых средств интерпретации, позволяющих получить визуальное отображение документа – 6. НИнт;
- возможность воспроизвести оригинал документа один в один (показатель означает, что внешний вид документа будет гарантированно совпадать с оригиналом после конвертации (не сканировании) в данный формат) – 7. 1:1;
- независимость от используемых шрифтов – 8. ШрНз;
- кроссплатформенность – 9. КрПлат;
- наличие национальных или международных стандартов – 10. СТД;
- возможность хранения метаданных – 11. МетаДн.

Приведем сравнение по данным показателям в виде табл. 1.

Табл. 1

Сравнение форматов с точки зрения долговременного хранения

Показатель	ODF	PDF/A	TIFF	JPEG	PNG	RTF	DOC	WebP
1. ПИсп	15	25 ¹	33	28	25	37	22	10
2. ЗаявПод	-	50	-	-	-	-	-	-
3. ТкстСтр	+	+	- ²	-	-	+	-	-
4. НСпДок	+	+	+	+	+	+	+	+
5. СвобИсп	+	+	-	-	+	-	-	+
6. НИнт	+	+	+	+	+	+	+	-
7. 1:1 ³	+/-	+/-	+ ⁴	+ ⁵	+	+/-	+/-	+ ⁶

¹ С момента создания PDF (PDF/A выделен в отдельный формат в 2005, но, по существу, он является подмножеством PDF.

² Возможно размещение в пользовательском теге, но стандартные средства интерпретации формата данный тег проигнорируют.

³ «+/-» в колонке означает, что при идеальных условиях (наличие всех шрифтов, совпадение установок принтера, одинаковые операционные системы, совпадение алгоритмов переноса слов и т. д.) искажений не возникнет, в противном случае вероятность возникновения искажений (размеры страниц, переносы слов, шрифты и т. д.) достаточно высокая.

⁴ Гарантированно для «черно-белых» документов.

⁵ Гарантированно для цветных изображений с плавными переходами яркости и цвета при отсутствии потерь при сжатии.

⁶ Гарантированно для цветных изображений с плавными переходами яркости и цвета при отсутствии потерь при сжатии.

Показатель	ODF	PDF/A	TIFF	JPEG	PNG	RTF	DOC	WebP
8. ШрНз	-	+ ⁷	+	+	+	-	-	+
9. КрПлат	+	+	+	+	+	+	-	+
10. СТД	+ ⁸	+	-	+	+	-	-	-
11. МетаДн	-	+ ⁹	-	-	-	-	-	-

Из приведенной таблицы видно, что в наибольшей степени для долговременной сохранности подходит формат PDF/A. Его достоинства: наличие поддержки производителя, независимость от набора шрифтов, наличие текстового слоя внутри файла, свободное использование и др. позволяют использовать его для долговременного хранения.

К недостаткам можно отнести: вероятность возникновения искажений при проведении конвертации электронных документов в этот формат, связанных с отсутствием необходимых шрифтов, различием в алгоритме переноса слов и др., что, в свою очередь, не гарантирует точного совпадения внешнего вида с документом в оригинальном (первоначальном) формате.

Для точного совпадения оригиналу существует формат PDF/X, но это все же отдельный формат, не связанный с PDF/A, к тому же также не лишенный проблем с совпадением шрифтов и цветовой схемы.

Исследования, проведенные в работе [9] показывают, что в PDF/A удобно хранить точный внешний вид электронного документа в форматах TIFF (для черно-белых документов) и JPEG (для цветных документов), плюс отдельно в текстовом слое полный текст документа и, возможно, его метаданные.

Сходные рекомендации приведены в исследовании [10], где также рекомендуется использовать TIFF для сохранения в документе изображений хорошего качества и черно-белых документов, PNG для цветных изображений, в которых можно четко выделить границы объектов, и JPEG для остальных цветных изображений.

Тем самым можно утверждать, что точный внешний вид документа удобнее сохранять в графическом формате, текст хранить в том же файле PDF/A. При этом необходимо помнить, что такой документ со всеми его метаданными является неделимой единицей использования.

В документ формата TIFF в отдельный тег, определенный пользователем, можно поместить текст документа. Однако отсутствие стандартных

⁷ Шрифты помещаются внутрь PDF/A, однако конвертация в данный формат на машине, на которой отсутствуют необходимые шрифты, приведет к невозможности их помещения в файл.

⁸ Формат принят в качестве стандарта в РФ.

⁹ Возможно хранить метаданные в невидимом текстовом слое.

средств интерпретации для пользовательских тегов делает практически невозможным использование такого решения.

Общими недостатками рассмотренных форматов является то, что они разработаны не в РФ, следовательно, при некоторых неблагоприятных условиях воспользоваться технической поддержкой данных форматов может быть затруднительно через десятилетия.

К достоинствам форматов можно отнести тот факт, что спецификации форматов известны, значит теоретически возможно создание собственных средств интерпретации или конвертации хранящихся цифровых данных в эти или собственные форматы. Но это, безусловно, потребует дополнительных затрат, и при наличии огромных объемов данных может быть неприемлемым по времени.

2. Разработка и область применения формата долговременного хранения

Все вышеперечисленные форматы годятся для хранения тех или иных данных, которые представляют собой единую логическую единицу долговременного хранения, доступ к которой может быть только целиком, т. к. единица хранения будет представлять собой один файл.

Если рассмотреть более сложную модель цифровых данных, в которой они не только распределены между информационными системами участников цифрового обмена в рамках цифровой экономики, но и доступ к разным частям общего массива распределенных цифровых данных может быть различным (см. [11, 12]), то решением такой противоречивой ситуации может быть только разработка собственного формата.

Проиллюстрируем подобную ситуацию следующим примером. Если описывать модель документа в системе межведомственного электронного документооборота, то документ будет состоять из текста документа, регистрационной карточки, контрольной карточки, которая, в свою очередь, состоит из резолюций, согласований, ознакомлений, связанных с различными должностными лицами разных организаций. Кроме того, документ принадлежит некоторому делу, связан с отделами организаций, с некоторыми ведомственными системами классификации и т. д.

Конечному пользователю может оказаться доступен только текст документа (точный внешний вид), тем самым хранить всю информацию документа, метаданные, резолюции и др. в одном файле становится невозможными с точки зрения модели безопасности.

Если же разбивать такой документ на несколько файлов, то как отследить целостность цифровых данных в целом?

Тем самым, с точки зрения разработанной автором модели документа при долговременном хранении [13], ни один из вышеперечисленных форматов до конца не пригоден. Т. к. кроме самого документа приходится хранить еще информацию о документе, метаданные, связанные файлы и классификаторы. Тем самым необходимость разработки формата долговременного хранения становится актуальной и непростой задачей.

Разработанная автором модель электронного документа при долговременном хранении [13] позволяет сформулировать ряд требований для определения области применения формата долговременного хранения распределенных цифровых данных:

1. Ограничением области применения является механизм автоматизированного информационного обмена цифровыми данными между субъектами цифровой экономики (см. [11, 12, 14]) с контролем доступа к различным данным единого цифрового массива. Тем самым в качестве единицы передачи (хранения) цифровых данных при информационном обмене можно считать в общем случае группу связанных файлов различных форматов (электронные документы, графические образы, мультимедиа-файлы и др.).
2. Созданные в рамках цифровой экономики данные содержат определенную семантику, следовательно, должны быть структурированы. Таким образом, в качестве формата долговременного хранения необходимо взять формат, позволяющий выполнять описания цифровых данных на основе типовых схем и структур.
3. Цифровые данные и (или) их части, как правило, связаны определенной иерархией, следовательно, формат долговременного хранения должен позволять описывать иерархические структуры данных произвольной глубины с отображением перекрестных связей между частями данных или отдельными электронными документами, образами и др.
4. Формат должен быть документированным и позволять интерпретировать цифровые данные в отсутствие средств интерпретации, т. е. цифровые данные, текст электронных документов, графические образы, в случае устаревания или исчезновения средств интерпретации, должны быть извлечены из общего массива цифровых данных и раскодированы.
5. Формат должен быть открытым для минимизации рисков, возникающих из-за отсутствия

поддержки формата или средств его интерпретации.

Всем перечисленным требованиям, в принципе, удовлетворяет формат eXtensible Markup Language – XML. Однако сам по себе стандарт XML является очень обобщенным форматом данных, вернее даже языком описания данных. Он создан консорциумом W3C (World Wide Web Consortium), состоящим из множества различных компаний. В результате в реализацию XML вошло очень много различных концепций и идей, в том числе и противоречивых. В частности, направленность формата одновременно и на размеченный текст, и на хранение структурированных данных. Как следствие, одну и ту же схему можно описать разными способами: например, тип элемента можно указывать через механизм типов или с помощью ссылки на другой элемент.

В связи с этим задача разработки формата долговременного хранения цифровых распределенных данных разбивается на несколько частных подзадач:

- создание подмножества XML-схем согласно модели документа при долговременном хранении (см. [13]) для описания хранимых и передаваемых цифровых данных;
- обеспечение возможности работы с цифровыми данными с помощью стандартного инструментария интерпретации XML-файлов;
- выбор форматов для долговременного хранения связанных с цифровыми данными электронных документов, графических образов и др. Таким образом при передаче цифровых данных можно легко включать или исключать те или иные части и (или) связанные электронные документы, образы и др.;
- разработка описаний формата цифровых данных и собственных средств интерпретации;
- создание механизма разделения доступа и деперсонализации данных для информационного обмена (см. [11]);
- создание нормализатора поступающих цифровых данных для включения их в связанный массив данных.

Заключение

Стремительная цифровизация экономик ведущих стран мира говорит о том, что в ближайшее время объем цифровых данных, в том числе предназначенных для долговременного хранения, станет стремительно возрастать.

С другой стороны, эти цифровые данные становятся ключевым фактором производства, и потеря или искажение их становятся крайне кри-

тичными. Этот факт повышает требования к защищенности данных.

С третьей стороны, риск неинтерпретируемости важных данных через десятилетия приводит к необходимости поиска путей решения этой задачи, которая становится также крайне актуальной.

В статье предложено решение задачи интерпретируемости цифровых распределенных данных долговременного хранения с помощью разработки формата, который наряду с математической моделью цифровых данных долговременного хранения [13] позволяет решить проблему интерпретируемости.

В статье проведен обзор существующих файловых форматов, которые могут быть использованы для долговременного хранения данных. Определены их достоинства и недостатки. Доказана необходимость создания собственного формата, который бы учитывал, с одной стороны, разнородность цифровых данных, с другой позволял организовать доступ не ко всем данным из-за проблемы конфиденциальности данных, с третьей позволял легко интерпретировать данные в случае выхода из строя или отсутствия программных средств интерпретации.

Необходимо также подчеркнуть, что изложенные в статье положения реализованы на практике. Создан формат долговременного хранения на основе подмножества языка XML. Созданный формат прошел практическую апробацию в рамках решения прототипа информационной системы. В дальнейших исследованиях планируется провести исследования данного формата при внедрении в различные информационные системы долговременного хранения, а также провести развитие данного формата.

Литература

1. *Solovyev, A.V.* Long-Term Digital Documents Storage Technology // Lecture Notes in Electrical Engineering. 2020. Vol. 641. P. 901-911.
2. *Solovyev, A.V.* Authentication control algorithm for long-term keeping of digital data // IOP Conference Series: Materials Science and Engineering (MSE), Vol. 862(5), 052080, 2020.
3. *Solovyev, A.V.* Digital media inventory algorithm for long-term digital keeping problem // IOP Conference Series: Materials Science and Engineering (MSE). Vol. 919(5), 052003. 2020.
4. *Open Government Partnership* UK National Action Plan 2013 to 2015. London. SW1A 2AS. 2013. 58 p.
5. *Pitman, N., Shipman, A.* A manager's guide to the long-term preservation of electronic documents. London. BIP 0089 BSI. 2008. 110 p.

6. *Universal Electronic Records Management (ERM) Requirements*. U.S. National Archives and Records Administration. 2017. URL: <https://www.archives.gov/records-mgmt/policy/universalemrequirements>
7. *Microsoft Office Word 97–2007 Binary File Format (.doc) Specification*. Microsoft Corporation. 2007. 210 p.
8. *WebP – a new image format for the Web*. [Электронный ресурс] – Google – 26/09/2013 – URL: <https://developers.google.com/speed/webp/>.
9. Николаев, Д.П., Постников, В.В., Усилин, С.А. Cognitive PDF/A – технология оцифровки текстовых документов для публикации в Интернет и долговременного архивного хранения // Труды ИСА РАН. 2009. Т. 45. С. 159–173.
10. Берестова, В.И. Средства и методы создания электронного документа, содержащего графические образы // Журнал «Делопроизводство». 2014. № 1. С.45–56.
11. Баканова, Н.Б., Соловьев, А.В. Алгоритм деперсонализации данных при информационном обмене в цифровой экономике // Труды ИСА РАН. 2020. 70 (2). С. 12–20.
12. Akimova, G.P., Pashkin, M.A., Soloviev, A.V., Tarkhanov, I.A. Modeling the methodology to assess the effectiveness of distributed information systems // *Advances in Science, Technology and Engineering Systems*. 2020. Vol. 5(1). P. 86–92.
13. Соловьев, А.В. Электронные архивы: разработка математической модели электронного документа при долговременном хранении // Информационные технологии и вычислительные системы. 2017. № 1. С. 46–61.
14. Devyatkin D.A., Grigoriev O.G., Sokolov I.A., Suvorov R.E., Tikhomirov I.A., Zhebel' V.V. On Creating a National System for Identifying Research and Development Priorities // *Scientific and Technical Information Processing*, Allerton Press, Inc., New York. 2019. Vol 46. Issue 1, P.14-19.

Соловьев Александр Владимирович, Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Главный научный сотрудник, доктор технических наук. Количество печатных работ: 125. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, математическое моделирование, долговременное хранение электронных документов. E-mail: soloviev@isa.ru

Solving the problem of interpreting digital data for long-term keeping

A.V. Solovyev¹

¹Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia

Abstract. In the context of the rapid digitalization of the economy, both the volume of digital data and their value are increasing. Distortion of this data is certainly unacceptable and will lead to large material and non-material losses. One of the problems when working with digital data, especially with their long-term storage, is the problem of data interpretation. This article proposes a solution to the problem of interpreting digital distributed long-term keeping data by developing a long-term keeping format, which, along with a mathematical model of long-term keeping digital data, allows solving the problem of interpretability. The article provides an overview of the existing digital data formats. Their advantages and disadvantages are determined. The necessity of creating a custom format has been proved, which would take into account, on the one hand, the heterogeneity of digital data, on the other hand, it would allow organizing access to separate parts of distributed data, on the other hand, it would make it easy to interpret the data in the event of a failure or lack of interpretation software. As a practical result, within the framework of the study, a long-term keeping format based on a subset of the XML language was created. The created format has passed practical testing. In the future, its wider implementation and modernization is planned.

Keywords: *digital data, long-term keeping, file format, electronic document, digitalization, digital economy*

DOI: 10.14357/20790279210206

References

1. *Solovyev, A.V.* 2020. Long-Term Digital Documents Storage Technology. Lecture Notes in Electrical Engineering. 641: 901-911.
2. *Solovyev, A.V.* 2020. Authentication control algorithm for long-term keeping of digital data. IOP Conference Series: Materials Science and Engineering (MSE). 862(5): 052080.
3. *Solovyev, A.V.* 2020. Digital media inventory algorithm for long-term digital keeping problem. IOP Conference Series: Materials Science and Engineering (MSE). 919(5): 052003.
4. *Open Government Partnership* UK National Action Plan. 2013. London. SW1A 2AS. 58 p.
5. *Pitman, N., and Shipman, A.* 2008. A manager's guide to the long-term preservation of electronic documents. London. BIP 0089 BSI. 110 p.
6. *Universal Electronic Records Management (ERM) Requirements.* 2017. U. S. National Archives and Records Administration. 2017. Available at: <https://www.archives.gov/records-mgmt/policy/universalmrequirements> (accessed March 25, 2021).
7. *Microsoft Office Word 97–2007 Binary File Format (.doc) Specification.* 2017. Microsoft Corporation. 210 p.
8. *WebP – a new image format for the Web.* 2013. Google. Available at: <https://developers.google.com/speed/webp/> (accessed March 25, 2021).
9. *Nikolayev, D.P., Postnikov, V.V., and Usilin, S.A.* 2009. Cognitive PDF/A – tekhnologiya otsifrovki tekstovykh dokumentov dlya publikatsii v Internet i dolgovremennogo arkhivnogo khraneniya [Cognitive PDF / A – technology for digitizing text documents for publication on the Internet and long-term archival storage]. Trudy ISA RAN [Proceedings of the ISA RAS]. 45. P. 159–173.
10. *Berestova, V.I.* 2014. Sredstva i metody sozdaniya elektronnoho dokumenta, soderzhashchego graficheskkiye obrazy [Means and methods for creating an electronic document containing graphic images]. Deloproizvodstvo [J Office work]. 1. P. 45–56.
11. *Bakanova N.B., and Solovyev A.V.* 2020. Algoritm depersonalizatsii dannykh pri informatsionnom obmene v tsifrovoy ekonomike [Data Depersonalization Algorithm for Information Exchange in the Digital Economy]. Trudy ISA RAN [Proceedings of the ISA RAS]. 70(2). P. 12–20.
12. *Akimova G.P., Pashkin M.A., Soloviev A.V., and Tarkhanov I.A.* 2020. Modeling the methodology to assess the effectiveness of distributed information systems. Advances in Science, Technology and Engineering Systems. 5(1). P. 86–92.
13. *Solovyev A.V.* 2017. Elektronnyye arkhivy: razrabotka matematicheskoy modeli elektronnoho dokumenta pri dolgovremennom khraneni [Electronic archives: development of mathematical models of electronic documents for long-term storage]. Informatsionnyye tekhnologii i vychislitel'nyye sistemy [Information Technology and Computing Systems]. 1. P. 46–61.
14. *Devyatkin D.A., Grigoriev O.G., Sokolov I.A., Suvorov R.E., Tikhomirov I.A., Zhebel' V.V.* 2019. On Creating a National System for Identifying Research and Development Priorities. Scientific and Technical Information Processing, Allerton Press, Inc., New York. 46(1). P. 14-19.

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