



ПЕРЕДМОВА

Випуск дайджесту присвячено досвіду установ світу щодо зберігання і використання інформації в сучасному інформаційному суспільстві.

У публікації «Национальные Архивы Великобритании: Использование искусственного интеллекта для отбора на архивное хранение электронных государственных документов» розповідається про дослідження потенціалу інструментів на основі штучного інтелекту (ШІ) з відбору, із напівструктурованих та неструктурованих масивів, документів для архівного зберігання.

Сьогодні штучний інтелект не може замінити досвід фахівців з управління документами, проте ці продукти продовжують еволюціонувати. Зараз слушний час для того, щоб налагодити взаємодію з постачальниками послуг на основі ШІ та вплинути на напрямки розвитку цих продуктів.

У публікації «Искусственный интеллект: Кому можно доверять?» розповідається, що на думку людей, штучний інтелект і роботи заслуговують на більшу довіру.

Але залежність від алгоритму в багатьох випадках призводить до дискримінаційної практики та непоправних наслідків через невраховані соціальні фактори.

У публікації «Управляемое машинное обучение: Искусственный интеллект без головной боли» розповідається, що у штучного інтелекту (ШІ) та машинного навчання є величезний потенціал. Однак відсутність персоналу, що володіє необхідними знаннями і навичками, створює бар'єр, що не дозволяє багатьом компаніям успішно його використовувати.

У публікації «Искусственный интеллект: При использовании предварительно обученных ИИ-алгоритмов в других организациях результаты могут отличаться» розповідається, що спроби застосовувати навчений алгоритм до даних, створених чи зібраних інакше, ніж вихідні навчальні дані, який завжди можуть давати хороші результати.

У публікації «Как добиться поддержки генеральным и финансовым директорами Вашей программы управления документами?» розповідається, що документи, інформація та дані порівнюються з «кров'ю» та «новою нафтою», без яких організації не можуть прожити. Але єдиного способу оцінки інформації, який влаштовував би всіх і кожного, не існує.

Тому без узгодженої з генеральним та фінансовим директорами моделі визначення цінності інформації, яку Ви зберігаєте, вона завжди буде витратною та фінансуватиметься за залишковим принципом.

У публікації «Норвегия: Коварная проблема управления документами» розповідається про фундаментальну проблему відсутності достатньої документації щодо рішень та діяльності державного сектора Норвегії.

Через появу більшої кількості цифрових каналів виріс обсяг і різноманітність інформації – від Facebook до електронної пошти – що

призвело до значного збільшення виробництва документів. Оскільки ця інформація не реєструється у відповідних системах, слід замислитись про її збереження. Мета публікації представити методологічну концепцію вирішення проблем управління документами, поділитися досвідом.

У публікації «Мари-Анн Шабен: Вытесняются и затмеваются ли документы данными?» розповідається про використання слова «дані», його значення та взаємозв'язок між «документами» та «даними». На сьогодні немає єдиного загальноприйнятого поняття «дані».

У статті порівнюються «Загальні правила захисту персональних даних» (General Data Protection Regulation, GDPR, що діють з травня 2018 року) та стандарт управління документами ISO 15489:2016 «Інформація та документація — Управління документами».

Зроблено висновок, що домінування «даних» на сьогоднішній день швидше маргіналізувало, ніж скинуло поняття «документ», змушуючи той адаптуватися в технічному, культурному та лінгвістичному плані.

У публікації «США: «Аккредитованный комитет по стандартизации X9» опубликовал новый стандарт ANSI X9.141-2021 защиты персональных данных и уведомления об утечках» розповідається, що стандарт X9.141 складається з двох частин у яких розглядається, що відбувається «до» захисту персональних даних та «після» повідомлення про витік. Істотно те, що стандарт доповнює та розширює важливі керівництва Національного інституту стандартів та технологій (NIST) та охоплює всі види чутливих даних, як фінансових, так і нефінансових. Вказана адреса за якою його можна придбати.

У публікації «ИСО и МЭК: Готовится новая редакция стандарта ISO/IEC 27005 «Руководство по менеджменту рисков и возможностей в области информационной безопасности»» наведено рекомендації з наступних питань:

- реалізація вимог щодо ризиків інформаційної безпеки;
- посилання на ключові документи у сімействі стандартів ISO/IEC 27000, що підтримують діяльність з управління ризиками інформаційної безпеки;
- дії, спрямовані на обробку ризиків, пов'язаних із інформаційною безпекою;
- впровадження рекомендацій щодо менеджменту ризику із стандарту ISO 31000 у контексті інформаційної безпеки.

У публікації «Система запису файлів на мікрофільм Zeutschel Archive Writer OP800/ OP 800 HR» наведено переваги зберігання інформації на мікрофільмах, призначення та технічні характеристики КОМ – систем Archive Writer OP800 та Archive Writer OP 800 HR.

НАЦИОНАЛЬНЫЕ АРХИВЫ ВЕЛИКОБРИТАНИИ: ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОТБОРА НА АРХИВНОЕ ХРАНЕНИЕ ЭЛЕКТРОННЫХ ГОСУДАРСТВЕННЫХ ДОКУМЕНТОВ

Источник: сайт Национальных Архивов Великобритании
<https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/research-collaboration/using-ai-for-digital-selection-in-government/>

Цифровая трансформация в органах правительства Великобритании привела к увеличению объёмов и разнообразия государственных документов, – одновременно с уменьшением внимания к организации и структурированию данных. Традиционные процессы, разработанные для бумажных документов, не в состоянии справиться с объемами, разнообразием, сложностью и распределенной природой электронных документов министерств и ведомств.

В рамках выполненного Национальными Архивами Великобритании проекта был исследован потенциал инструментов на основе искусственного интеллекта (ИИ) для решения этой проблемы. Пять поставщиков ИИ-решений применили свои инструменты для классификации предоставленного Национальными Архивами набора данных. Оценивались следующие инструменты и платформы: Adlib Elevate, Amazon Web Services, Microsoft Azure, InSight от компании Iron Mountain и Records365 от компании RecordPoint. В целом, были получены обнадеживающие результаты, при этом ни один инструмент или подход не смог систематически превзойти все другие при выполнении всех задач.

В рамках проекта было обнаружено, что, хотя искусственный интеллект и не может заменить опыт специалистов по управлению документами, однако доступные на рынке ИИ-инструменты и конвейеры могут успешно помочь в решении задачи по отбору документов из полуструктурированных и неструктурированных массивов документов. Данные продукты продолжают эволюционировать, и сейчас подходящее время для того, чтобы наладить взаимодействие с поставщиками и повлиять на направления развития этих продуктов.

В документе под названием «Использование искусственного интеллекта для отбора электронных документов в государственных органах» (Using AI for Digital Selection in Government, <https://cdn.nationalarchives.gov.uk/documents/using-ai-digital-selection-in-government.pdf>) описаны основные результаты данного проекта, а также даны рекомендации для государственных органов, желающих внедрить собственное решение на основе искусственного интеллекта. Руководство

адресовано должностным лицам государственных органов, которые знакомы с процессами экспертизы ценности, отбора, анализа и передачи документов на постоянное хранение в Национальные Архивы.

В рамках проекта для проведения анализа имеющихся на рынке инструментов искусственного интеллекта был привлечён опытный маркетолог. Поставщики подготовили отчёты, объясняющие причины выбора определённых продуктов. Каждый поставщик предоставил отчёт в конце проекта (ссылки на них приведены ниже). В рамках проекта Национальные Архивы также разработал «инструмент для сравнительного анализа» (benchmarking tool); в упомянутом выше документе представлена подробная информация об этом инструменте.



Рекомендации и основные результаты проекта

- «Использование искусственного интеллекта для отбора электронных документов в государственных органах» (Using AI for Digital Selection in Government), <https://cdn.nationalarchives.gov.uk/documents/using-ai-digital-selection-in-government.pdf>

Дополнительные материалы проекта

- Отчет о первом этапе исследования рынка (Report of First Phase Market Research), <https://cdn.nationalarchives.gov.uk/documents/phase-1-market-research-ai.pdf>
- Шаблон оценки ИИ-инструментов для отбора документов на архивное хранение (AI for Selection Evaluation Template), <https://cdn.nationalarchives.gov.uk/documents/ai-for-selection-evaluation-template.xlsx>
- Инструмент Национальных Архивов для сравнительного анализа (The National Archives Benchmarking Tool),

<https://cdn.nationalarchives.gov.uk/documents/ai-for-selection-national-archives.pdf>

Отчёты поставщиков

- «Использование ИИ Национальными Архивами для отбора документов на архивное хранение» (The National Archives - AI for Digital Selection), компании Deloitte и Adlib, <https://cdn.nationalarchives.gov.uk/documents/deloitte-adlib-national-archives.pdf> (обзор решения Adlib Elevate компании Adlib);
- «Всесторонняя оценка веб-сервисов Amazon Web Services» (AWS Comprehensive Evaluation), компания Kainos, <https://cdn.nationalarchives.gov.uk/documents/kainos-aws-national-archives.pdf>
- «Использование ИИ Национальными Архивами для отбора документов на архивное хранение» (The National Archives - AI for Digital Selection), презентация компании Iron Mountain, <https://cdn.nationalarchives.gov.uk/documents/iron-mountain-insight-national-archives-ai-for-selection-report.pptx>
- «Оценка решения Records365 - Использование ИИ Национальными Архивами для отбора документов на архивное хранение» (Records365 Evaluation Report - The National Archives - AI for Digital Selection), компания RecordPoint, <https://cdn.nationalarchives.gov.uk/documents/recordpoint-records365.pdf>
- «Использование машинного обучения для классификации документов» (Machine Learning for Document Classification), <https://cdn.nationalarchives.gov.uk/documents/adatis-azure-national-archives.pdf> (подход компании Adatis на основе сервисов Microsoft Azure).



ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: КОМУ МОЖНО ДОВЕРЯТЬ?

Источник: блог компании Formtek <https://formtek.com/blog/artificial-intelligence-who-can-you-trust/> Автор: Дик Вейсингер

Хорошо, Вы прочли *это* в Интернете, значит, *это* должно быть правдой ... А если это не так, то кому мы можем доверять?

А как насчет искусственного интеллекта - Вы бы ему доверяли? Готовы ли вы довериться беспилотному транспортному средству (<https://towardsdatascience.com/why-deep-learning-is-not-a-silver-bullet-for-autonomous-vehicles-371c74c1a02b>), действующему на основе алгоритма, те которые сами его изобретатели не вполне понимают, хотя при этом он, похоже, хорошо работает?

Результаты опросов показывают, что, по мнению людей, искусственный интеллект (ИИ) и роботы заслуживают большего доверия.

Например, опрос фирмы Oracle (<https://www.oracle.com/news/announcement/money-and-machines-021021.html>) показал, что две трети опрошенных предпочли бы, чтобы их финансы управлялись алгоритмом, а не человеком, - и более половины заявили, что алгоритм, вероятно, сможет выполнить работу лучше, чем они.

Аспирант университета Джорджии Эрик Богерт (Eric Bogert, <https://www.terry.uga.edu/directory/mis/eric-bogert.html>) полагает, что «алгоритмы способны выполнять огромное количество задач, и список таких задач расширяется практически с каждым днем. Похоже, что есть предрасположенность в большей степени полагаться на алгоритмы по мере усложнения задач, и этот эффект оказывается сильнее, чем склонность полагаться на советы других людей» (<https://www.sciencedaily.com/releases/2021/04/210413114040.htm>).

Но, возможно, мы слишком доверчивы.

Доцент университета Джорджии Аарон Шектер (Aaron Schecter, <https://www.terry.uga.edu/directory/mis/aaron-schecter.html>) отмечает, что «одна из распространенных проблем с ИИ - это когда он используется при принятии решения о предоставлении кредита или одобрении выдачи кому-либо ссуд. Хотя это субъективное решение, при его принятии принимается во внимание множество чисел, таких, как доход и кредитный рейтинг, поэтому люди считают, что это подходящая задача для алгоритма. Но мы знаем, что зависимость от алгоритма во многих случаях приводит к дискриминационной практике из-за неучтенных социальных факторов» (<https://www.sciencedaily.com/releases/2021/04/210413114040.htm>).

Педагог и исследователь Аянна Ховард (Ayanna Howard, <https://howard.ece.gatech.edu/>) говорит: «если мы чрезмерно доверяем этим системам, и эти системы принимают решения, приводящие к разным результатам для разных групп людей - скажем, если система медицинской диагностики работает по-разному с женщинами и с мужчинами - то это значит, что мы сейчас создаем системы, которые усиливают существующее сегодня неравенство. Это проблема. И когда Вы подключаете ИИ-системы к решению вопросов, связанных со здоровьем или транспортом - и в той, и в другой сфере могут возникать ситуации, в которых решается вопрос жизни или смерти, - плохое решение может привести к непоправимым последствиям. Так что нам действительно нужно исправить такое положение дел» (<https://www.technologyreview.com/2021/05/13/1024874/ai-ayanna-howard-trust-robots/>).

УПРАВЛЯЕМОЕ МАШИННОЕ ОБУЧЕНИЕ: ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ БЕЗ ГОЛОВНОЙ БОЛИ

Источник: блог компании Formtek <https://formtek.com/blog/managed-machine-learning-ai-without-the-headaches/> Автор: Дик Вейсингер

У искусственного интеллекта (ИИ) и машинного обучения имеется огромный потенциал, - однако в отсутствие обладающего значительными знаниями и навыками обученного персонала барьер для эффективного использования ИИ высок, что не позволяет многим компаниям успешно его использовать.

Директор по менеджменту продуктов принадлежащей Google Cloud платформы искусственного интеллекта (AI Platform) Крейг Уайли (Craig Wiley, <https://www.linkedin.com/in/craigwiley/>) говорит: «С моей точки зрения, машинное обучение в корпоративных условиях переживает кризис. Как человек, проработавший в этой сфере ряд лет, могу сказать, что если Вы посмотрите «Гарвардское деловое обозрение» (Harvard Business Review), или обзоры аналитиков, или что-то ещё, то буквально каждый из этих документов отмечает, что подавляющее большинство компаний либо инвестируют, либо заинтересовано в инвестировании в машинное обучение, но при этом они не получают от этого отдачу. Это нужно изменить, такое положение должно измениться» (<https://techcrunch.com/2021/05/18/google-cloud-launches-vertex-a-new-managed-machine-learning-platform/>).

Исследователь из компании Google Кэсси Козырьков (Cassie Kozyrkov, <https://www.linkedin.com/in/kozyrkov/>) сравнила преимущества использования платформы управляемого машинного обучения (Managed Machine Learning) с преимуществами швейцарского армейского ножа. Такая платформа способна обеспечить пользователям более быстрое конфигурирование, настройку и обучение, по сравнению с вариантами, когда приходится начинать с нуля или использовать существующие рамочные структуры. (<https://www.infoq.com/news/2021/06/google-vortex-ai/>).

Главный аналитик компании Omdia по ИИ-платформам, аналитике и менеджменту данных Брэдли Шиммин (Bradley Shimmin, <https://omdia.tech.informa.com/authors/bradley-shimmin>) полагает, что «корпоративные специалисты-практики в области информатики, надеющиеся внедрить ИИ в деловой деятельности организаций, не хотят разбираться с различными соответствующими инструментами. Скорее, они хотят получить инструменты, которые способны «укротить» жизненный цикл машинного обучения. К сожалению, это не простая задача. Для её решения требуется поддерживающая инфраструктура, способная объединить пользовательский опыт, используя сам ИИ в качестве помощника, и сделать данные самой сердцевинкой процесса - при одновременном поощрении гибкого внедрения

разнообразных технологий» (<https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-launches-vertex-ai-unified-platform-for-mlops>).

По словам вице-президента Google Cloud Эндрю Мура (Andrew Moore, <https://www.linkedin.com/in/andrew-moore-016b751/>), преимущество использования предварительно настроенной платформы, специально нацеленной на решение проблем ИИ, заключается в том, что она может помочь любой компании «серьезно взяться за перевод ИИ из чистилища пилотных проектов в полномасштабную промышленную эксплуатацию» (<https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-launches-vertex-ai-unified-platform-for-mlops>).

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ПРИ ИСПОЛЬЗОВАНИИ ПРЕДВАРИТЕЛЬНО ОБУЧЕННЫХ ИИ-АЛГОРИТМОВ В ДРУГИХ ОРГАНИЗАЦИЯХ РЕЗУЛЬТАТЫ МОГУТ ОТЛИЧАТЬСЯ

Источник: блог компании Formtek <https://formtek.com/blog/artificial-intelligence-your-miles-may-vary-when-ai-algorithms-go-into-production/> Автор: Дик Вейсингер

Алгоритмы искусственного интеллекта (ИИ) можно обучить (*не следует, однако, забывать о том, что искусственный интеллект не сводится к решениям на основе машинного обучения и нейронных сетей*). Но важно помнить, что обученные алгоритмы настроены с использованием специфических наборов обучающих данных так, чтобы хорошо работать в очень специфических условиях. Попытки применять обученный алгоритм к данным, которые были созданы или собраны иначе, чем исходные обучающие данные, не всегда могут давать хорошие результаты.

Профессор Стэнфордского университета и пионер ИИ Эндрю Нг (Andrew Ng, <https://www.andrewng.org/>) в интервью, данном Текле Пери (Tekla Perry, <https://www.linkedin.com/in/tekla-perry-33b4a211/>) для сайта IEEE Spectrum (см. <https://spectrum.ieee.org/view-from-the-valley/artificial-intelligence/machine-learning/andrew-ng-xrays-the-ai-hype>), говорил о проблемах, которые могут возникнуть, когда алгоритмы ИИ, настроенные для работы с тестовыми данными в лаборатории, переводятся в промышленную эксплуатацию и используются для анализа данных из приложений в реальной жизни.

Эндрю Нг (<https://www.linkedin.com/in/andrewyng/>) отметил, что «когда мы собираем данные Стэнфордского госпиталя, затем обучаем

алгоритм и тестируем его на данных из того же госпиталя, то, действительно, мы можем публиковать статьи, показывающие, что алгоритмы сопоставимы по эффективности с радиологами-людьми в плане выявления определенных симптомов. Оказывается, однако, что когда Вы берёте ту же самую модель и ту же самую ИИ-систему в более старую больницу вниз по улице, где стоит старый компьютер и техник использует несколько иной протокол захвата графических образов, возникающее смещение в данных может приводить к значительному ухудшению эффективности ИИ-системы. Напротив, любой рентгенолог-человек может прогуляться по улице до этой более старой больницы и отработать там как надо. Таким образом, даже если в определенный момент времени на конкретном наборе данных мы можем показать, что ИИ работает, - клиническая реальность такова, что все эти модели всё ещё нуждаются в основательной доработке, прежде чем они будут готовы к промышленной эксплуатации».

Нг сказал, что «у искусственного интеллекта повсеместно, а не только в сфере здравоохранения, есть разрыв между проверкой работоспособности концепции и промышленной эксплуатацией. Полный цикл проекта машинного обучения - это не одна только разработка модели. В него входит поиск правильных данных, развертывание и мониторинг ИИ-решения, обеспечение обратной связи к модели, демонстрация безопасности - делается всё, что необходимо сделать для подготовки к промышленной эксплуатации. Это куда большее, чем получение хороших результатов на тестовой выборке, в чём, к счастью или к сожалению, мы преуспеваем в сфере машинного обучения».



КАК ДОБИТЬСЯ ПОДДЕРЖКИ ГЕНЕРАЛЬНЫМ И ФИНАНСОВЫМ ДИРЕКТОРАМИ ВАШЕЙ ПРОГРАММЫ УПРАВЛЕНИЯ ДОКУМЕНТАМИ?

Источник: блог Meta-IRM <https://metairm.substack.com/p/how-to-get-your-cfo-and-ceo-to-support>. Автор: Карл Мелроуз австралийский специалист в области управления документами и информацией.

Как добиться поддержки генеральным и финансовым директорами Вашей программы управления документами? Сформулировать ответ на этот вопрос несложно, сложно добиться единодушного согласия с ним, - и этот ответ не найти ни в одной из программ, которые я когда-либо видел.

Чтобы заручиться поддержкой Вашего финансового директора (CFO) и генерального директора (CEO), необходимо иметь согласованную моделью определения ценности информации, которую Вы храните.

Единственными людьми, чьё согласие с моделью определения ценности способно повлиять на деятельность организации, вероятно, будут Ваш генеральный и финансовый директора (или лица, занимающие эквивалентные должности).

В нашей отрасли сейчас так много говорят о документах, информации и данных как о «крови» и «новой нефти», - и о том, как организации не могут прожить без нас.

- Стоимость крови - 210 долларов США за упаковку;
- Цена сырой нефти сегодня составляет 82,21 доллара за баррель;
- Цена негодной информации и плохой практики управления документами в Вашей организации составляет – ... (сколько?). Будет ли это сумма стоимостей приводимых аналогий (крови, нефти) т.е. 292,21 доллара?

Если бы ущерб от утраты одного документа составлял 85 долларов, мы бы относились к ней так, как если бы он стоил 85 долларов. Тогда мы бы знали, что короб, содержащий 100 документов, стоит 8500 долларов, - поэтому его хранение может на самом деле быть экономически оправдано.

Наши финансовые отделы отреагируют на это так же, как и на недостающие 9 долларов в выписке по вашей кредитной карте, которые, как Вы уверены, Вы не тратили.

Мы могли бы посылать руководителям счета за утрату документов или за их неправильное размещение в системе управления документами, и заставить их отнестись к этому серьезно, - потому что такая оценка ценности документов была согласована с генеральным директором и/или финансовым директорами.

Модели определения ценности прекрасны еще и тем, что не существует единого способа оценки информации, который бы устраивал всех и каждого.

Для одних учреждений будет иметь большое значение стоимость замены информации. Для других – какие-то виды метрик производительности или рыночная стоимость. Может учитываться в каком-то виде непреходящая историческая ценность или репутационная ценность для организации, и то, как, как это может повлиять на дееспособность организации. Мы также можем показать, что определённая информация является активом, а кое-какая – балластом / пассивом, - и поработать в проектах, исходя из их способности повысить ценность информационного актива.

Это всё потому, что контекст имеет значение.

Документ, захваченный в систему, которая позволяет нам автоматически доставить его нуждающемуся в этом документе представителю службы поддержки клиентов, имеет более высокую ценность, чем документ, хранимый так, что доступ к нему может быть получен лишь через два дня.

По своей сути, это вопрос стратегического управления информацией (information governance).

Стратегическое управление всегда связано с распределением в организации полномочий по принятию решений.

На данный момент никто не принимает решений о ценности информации в их организации с финансовой точки зрения.

Чему же тогда удивляться, если к нашей деятельности относятся как к затратной и не приносящей отдачи, а наши бюджеты постоянно сокращаются?



НОРВЕГИЯ: КОВАРНАЯ ПРОБЛЕМА УПРАВЛЕНИЯ ДОКУМЕНТАМИ

Источник: сайт LinkedIn <https://www.linkedin.com/pulse/wicked-problem-records-management-espen-sj%C3%B8voll/> Автор: директор Национальных Архивов Норвегии Эспен Шёволла

В 2019 году Национальные Архивы представили амбициозное видение: «К 2025 году ни один государственный служащий не должен управлять документами вручную». Вы вполне справедливо можете посчитать такие намерения весьма амбициозными – так что же нужно для достижения этой цели?

До 2025 года осталось ещё несколько лет. В рамках усилий по реализации этого видения, Национальные Архивы Норвегии провели в 2020 году проект Arkivflokken – или, в переводе, «Коварная проблема управления документами» (*по-норвежски это звучит даже сильнее – «проклятье документов»*). В проекте использовался сервисно-ориентированный подход, разработанный для решения сложных проблем в государственном секторе. В отчёте о проекте описан весь его процесс и извлечённые по его ходу уроки.

В первые десятилетия цифровой трансформации ведение документации стало «функцией самообслуживания сотрудников». Следствием этого является то, что каждый регистрирует свои собственные документы, что рассматривается как серьезная потеря рабочего времени и как проблема. Во всех административных структурах управление документами воспринимается как задача, которую нужно выполнять в дополнение к своей основной работе. Сотрудники воспринимают соответствующие системы как сложные для использования. Кроме того, служащие государственного сектора должны на свой страх и риск принимать множество решений о том, что следует сохранить. Результатом этого стали серьёзные пробелы в документировании и в документах государственного сектора.

В отчете «Коварная проблема управления документами» исследуется вопрос, который является фундаментальной проблемой для демократического общества: в государственном секторе Норвегии отсутствует достаточная документация по решениям и деятельности, поскольку эта информация не регистрируется в соответствующих системах.

Масштабы проблемы увеличились с ростом объёма и разнообразия информации ввиду появления большего количества цифровых каналов, посредством которых граждане и сотрудники государственных служб вступают в многоканальный – от Facebook до электронной почты и личных встреч - диалог.

Выросшие объём и разнообразия информации приводят к производству буквально гор документов, о сохранении которых следует задуматься.

Наше общество сталкивается с рядом сложных взаимосвязанных проблем, которые отдельные действующие лица не в состоянии решить в одиночку. Эти проблемы как раз можно назвать большими коварными социальными проблемами, решение которых требует нововведений, сотрудничества, инноваций и внедрения методов работы, отходящих от традиционной проектной методологии и знакомых структур административного управления. В данном отчёте выделены несколько проблем и предложена методологическая концепция того, как Национальные Архивы Норвегии могли бы решить коварные проблемы управления документами в сотрудничестве с другими государственными органами и учреждениями, а также с действующими на рынке поставщиками.



Мы считаем, что многие из извлеченных в ходе выполнения проекта уроков будут актуальны для архивных служб и ответственных за цифровую трансформацию государственных органов в правительствах по всему миру. Хотя некоторые вопросы актуальны только для Норвегии, мы полагаем, что данный отчёт может быть интересен тем сотрудникам государственных органов и учреждений, которые стремятся найти новые подходы к решению межотраслевых вопросов. Цель публикации «Коварной проблемы управления документами» заключается в том, чтобы поделиться нашим опытом поддерживаемых проектированием инноваций (design-driven innovation) в государственном секторе, и изучить, каким образом идеи могут

способствовать новым способам мышления и более ориентированным на пользователя государственными услугами. Мы надеемся, что отчёт окажется и полезным, и вдохновляющим.

Если Вы хотите узнать больше, то полный текст 35-отчета на английском языке доступен на сайте Национальных Архивов Норвегии по адресу: <https://www.arkivverket.no/...>

Точное название отчёта – «Какие уроки мы можем извлечь из коварной проблемы управления документами?» (What lessons can we learn from the wicked problem of records management?). Содержание этого документа следующее:

1. Введение и цель
2. История проекта Arkivflokken
3. Каким образом государственный орган может проявлять предприимчивость и решать проблему, у которой нет единого ответственного «владельца»?
4. Экосистема мер по решению основных проблем.
5. Как вызвать движение на вошедшем в ступор рынке?
6. Пицца для размышлений: Проект Arkivflokken как миссия

МАРИ-АНН ШАБЕН: ВЫТЕСНЯЮТСЯ И ЗАТМЕВАЮТСЯ ЛИ ДОКУМЕНТЫ ДАННЫМИ?

Источник: <https://www.marieannechabin.fr/2021/11/les-donnees-ont-elles-evince-ou-eclipse-les-documents-3-3/> Автор: Мари-Анн Шабен – известная французская специалистка в области управления документами и архивного дела



Часть 1: Господство данных

Данные (données). Наши данные. Данные компании. Открытые данные. Персональные данные. Сеть данных ... Данные повсюду. Это господство «данных».

Слово «данные» все чаще появляется в заголовках публикаций в СМИ, статей в журналах и постов на блогах.

Поэзия данных

Ниже приведена подборка заголовков статей последних лет, посвященных многообразным аспектам данных (составленный в хронологическом порядке список несколько длинноват, однако материал богат, и читать его увлекательно; также обратите внимание на то, что из выборки были исключены статьи, содержащие в заголовках англоязычный термин « data », поскольку его использование – дурной тон для публикаций на французском языке!):

[Список очень длинный, поэтому здесь он не приводится. С ним можно познакомиться по адресу <https://www.marieannechabin.fr/2021/11/les-donnees-ont-elles-evince-ou-eclipse-les-documents-1-3/>]

Но что такое «данные»?

Этот яркий выход - или вторжение - «данных» на авансцену (в зависимости от Вашей точки зрения) ставит ряд вопросов:

Что именно означает слово «данные» (во множественном или единственном числе)?

Что не является «данными»?

- Являются ли «данные» недавно появившейся оригинальной сущностью, или же это метаморфоза чего-то ранее существовавшего?

- Какова связь между данными и документами (documents), о которых мы обычно говорили: личными документами, документами компаний (а также связь с вопросами доступа и классификации документов и т.д.)?

- Вытесняются ли и будут ли навсегда вытеснены - и отправлены к «праотцам» (то есть в архивы) - документы данными с информационной сцены, или же мы являемся свидетелями всего лишь временного падения интереса к документам в том «платье», которое они носили в последние десятилетия, и ожидаем возрождения документов, теперь уже облаченных в новые цифровые одежды?

Изучение первого вопроса (что такое данные, что они собой представляют?) сразу же позволяет увидеть вопиющее несоответствие между существующими определениями (общими и техническими) и тем, что мы можем себе представить, просматривая приведенные выше заголовки публикаций, которые можно воспринимать как сводку характерных проблем, связанных с данными в 21-м веке - или как поэму нового времени.

Толковые словари, оставаясь ценными для истории использования термина «данные», явно устарели и не отражают текущее использование данного слова (словари все ещё отслеживают, как это слово используется, но в этом они, похоже, сильно отстают от жизни).

Словарь Французской Академии (Le dictionnaire de l'Académie française, <http://www.dictionnaire-academie.fr/article/A9D3040>) описывает «данные» как математический термин, появившийся в 18-м веке (данные задачи, которую нужно решить).

Наиболее распространённое сегодня определение ссылается на понятие «рассуждение»: «То, что известно и признано, и что служит основой для рассуждений, изучения или исследования» (« Ce qui est connu et admis, et qui sert de base, à un raisonnement, à un examen ou à une recherche », CNRTL, <https://www.cnrtl.fr/definition/donnée>), и ту же идею мы находим в словаре Larousse (<https://www.larousse.fr/dictionnaires/francais/donnée/26436>), в Википедии (<https://fr.wikipedia.org/wiki/Donnée>) или в онлайн-словаре Робер (Le Robert, <https://dictionnaire.lerobert.com/definition/donnee>).

Свои определения предлагаются в различных научно-технических публикациях. Например, на сайте www.techno-science.net читаем (см. <https://www.techno-science.net/definition/222.html>): «В сфере информационных технологий (ИТ) данные – это элементарные описания вещей, деловых операций, событий и т.д., часто представленные в виде кодов, Данные могут различным образом храниться и классифицироваться: бумажные, электронные, алфавитно-цифровые, графические изображения, аудиозаписи и т.д. ».

Мы перешли, как нетрудно заметить, от математики к информатике.

Существует немало глоссариев, предлагающих определения понятия «данные» (или копирующих их из словарей или других глоссариев), так что перечислять их все не имеет особого смысла. Со своей стороны, я придерживаюсь первоначального определения, данного в опубликованном в 2001 году стандарте открытой архивной информационной системы (Open Archival Information System, OAI, ныне это международный стандарт ISO 14721, о нём см. https://fr.wikipedia.org/wiki/Open_Archival_Information_System), посвящённом обеспечению долговременной сохранности научных данных.

В стандарте OAI сказано следующее: «Данные: повторно интерпретируемое представление информации в формализованном виде, пригодном для передачи, интерпретации или обработки. В качестве примеров данных можно назвать последовательность битов, таблицу чисел, символы на странице, запись издаваемых говорящим человеком звуков или образец лунного камня».

Это определение тем более интересно, что оно перекликается с определением понятия «информация», которое слишком часто ассоциируют с данными, не зная при этом, как объяснить разницу между этими двумя понятиями. В OAI «информация» определяется как «любые знания, которыми можно обмениваться. При обмене она представлена в виде данных. Примером может служить строка битов (данные), сопровождаемая описанием того, как интерпретировать строку битов в виде чисел, отражающих результаты измерений температуры, представленные в градусах Цельсия (информация о представлении)».

Это определение OAIS вдохновило меня на создание определения понятия «данные» для моего «Нового глоссария архивной терминологии» (Nouveau glossaire de l'archivage, 2010, https://www.arcateg.fr/wp-content/uploads/2017/03/Nouveau_glossaire_de_l_archivage.pdf - о нём также см. пост http://rusrim.blogspot.com/2010/03/blog-post_13.html): «Данные: Слово, число, сигнал, цепочка символов, последовательность битов, часть материала (*morceau de matière*) или иной первичный компонент (*élément brut*), записанный в информационной системе, где его можно соотнести с другими объектами и интерпретировать, превратив тем самым в информацию», сопровождающееся примечанием: «Данные - это только один компонент информации или документа. Следовательно, архивирование элементарных данных не имеет смысла, в отличие от операции резервного копирования, которая направлена на восстановление элементов системы в случае сбоя». Но к понятию «документ» (document) я вернусь позже.

Очевидно, что в начале 21 века мы перешли из сферы информационных технологий в области экономики, повседневной жизни и управления населением. Это огромное изменение, такого масштаба, что «компетентным органам» стоило бы взяться за данный вопрос и предложить определение, которое охватывало бы, скажем, как минимум 90% реальных значений терминов «данные» (*données* - во множественном числе) и «элемент данных» (*donnée* - в единственном числе), в которых те сегодня используются в публикациях. Это, возможно, помешало бы опубликовать «нелепые и неотшлифованные» определения, такие, как то, что можно прочесть в конце последней работы Орели Жан (Aurélie Jean) «Создают ли алгоритмы закон (Les algorithmes font-ils la loi ? , издательство l'Observatoire, <https://start.lesechos.fr/innovations-startups/tech-futur/aurelie-jean-les-algorithmes-ne-sont-pas-coupables-les-seuls-responsables-sont-ceux-qui-les-concoivent-1357204>), где в глоссарии предлагается следующее: «Данные (*Data*): это информация в форме данных (*données*), которая описывает физическое или юридическое лицо, страну, компанию, объект или даже сценарий» - заявление столь же шокирующее с лингвистической, как и со смысловой точки зрения (могу лишь предположить, что автор не перечитывала корректуру ...).

К сожалению, в большинстве определений не делается различия между единственным (*donnée*) и множественным числом (*données*), а англоязычный термин *data* используется как имя существительное женского рода в единственном числе, в то время как оно и в латинском языке является существительным среднего рода во множественном числе (чтобы не пришлось использовать ужасно варварское «*datas*» - за что (вместе с «*esse*») живодёров от правописания могли бы и повесить?). Тем не менее, помимо элемента данных (в английском языке, *data element* - это форма единственного числа от *data*, которое, повторяюсь, является существительным во множественном числе), мы должны признать растущее употребление «данных» (*donnée*) в единственном числе как коллективного единственного

числа (singulier collectif) – подобно тому, как мы говорим «электромобиль» (la voiture électrique), « вакцина » (le vaccin) или « комикс » (la bande dessinée).



Часть 2: Данные затмевают документы

Честно говоря, мысль написать этот пост пришла мне в голову какое-то время тому назад, когда я читала книгу Жерома Дени (Jérôme Denis) «Невидимая работа данных. Элементы для социологии инфраструктур письменных материалов» (Le travail invisible des données. Éléments pour une sociologie des infrastructures scripturales, Париж, изд-во Mines, социологическая коллекция, 2018, <https://www.pressedesmines.com/produit/le-travail-invisible-des-donnees/>).

Эта книга побудила меня изучить вопрос о взаимосвязи между документами и данными. Она подсказала мне, где можно познакомиться с другими точками зрения, и заставила заново задуматься о других интерпретациях. Именно эти точки зрения я обсуждаю во второй части своего поста, после краткого обзора посвященных данным публикаций (см. первую часть поста «Le règne de la donnée», <https://www.marieannechabin.fr/2021/11/les-donnees-ont-elles-evince-ou-eclipse-les-documents-1-3/>) и перед тем, как дать анализ данного вопроса с точки зрения дипломатики (это будет часть 3).

«Данные» доминируют в сфере научных исследований

Книга «Невидимая работа данных» кажется мне репрезентативным представителем растущего числа научно-исследовательских публикаций (например, в области социологии, информатики), в которых термин «документ», когда-то очень распространенный, вытесняется термином «данные». Это субъективное впечатление, которое хорошо было бы подтвердить или опровергнуть путем надлежащего текстометрического анализа.

Жером Дени - социолог с большим опытом управления данными в инженерной среде. В своей книге о «невидимой работе данных» он

предпочитает оставить в стороне «социологию *с использованием данных*» с тем, чтобы сосредоточить внимание на «социологии *данных*». Отмечая, что пока еще не существует удовлетворительного определения понятия «данные», он выходит за рамки вопроса об определении понятия («Что такое данные?»), чтобы задать вопрос «**Когда это является данными?**» (Quand est-ce une donnée?). Несколько отстраняясь от общепринятых представлений о том, что а) мы являемся свидетелями «революции данных», и что б) все данные будут доступны в качестве больших данных, он стремится **наблюдать за «операциями записи и чтения», которые используются при создании данных и поддерживают их распространение.**

Жером Дени рассказывает о своем видении на примере личного опыта. Когда его отец умер, он и его семья столкнулись (как и многие другие семьи) с запутанным клубком процедур и процессов обмена информацией (административной, банковской, служебной, из учётной записи Facebook и т.д.) - необходимым шагом для того, чтобы эта реальность смерти любимого человека, в некотором смысле незафиксированные в письменной форме данные, стала достоверными данными в рамках общества.

Что меня сразу поразило при чтении книги - безусловно, с учётом моего собственного профессионального опыта и культуры - так это то, в какой степени понятие «документ», по-прежнему остающееся глубоко укоренившимся в административной деятельности и договорных отношениях с поставщиками всех типов, исчезло, затенённое понятием «данные». Конечно, мы говорим о документах, и особенно о документах, которые должны быть подписаны в такой ситуации, и Жером Дени признаёт, что документы, возможно, не являются столь уж второстепенными для понимания социальных фактов, как думают другие социологи. «Документы, сопровождающие смерть, - пишет Жером Дени, - позволяют нам получить довольно чёткое представление о роли, выполняемой определенной информацией. Они ясно показывают, что создание данных и связанные с ними в дальнейшем операции по чтению и записи, в полной мере участвуют в создании той реальности, которую мы разделяем». Хочу добавить, что это наблюдение не ограничивается контекстом траурных событий. Итак, документ здесь присутствует (это может быть сертификат, удостоверение, письмо, контракт, форма, выписка со счета ...), однако автор выдвигает на первый план понятие «данных», собранных в базах данных и включённых в различные информационные потоки, которые затем представляются в виде более формальных документов.

В свете этого личного, но в то же время и универсального опыта, Жером Дени проводит почти что этнографические наблюдения за данными, за тем, как они «работают» и за соответствующими «работниками» в различных условиях: в лаборатории, администрации, банке. В частности, он связывает эти манипулирование данными с понятием записи (inscription) или письменного документа (écrit) – к последнему термину я вернусь в части 3 моего поста.

Эта исследовательская работа по изучению данных, в среде которых мы находимся - когда они не окружают нас... - возможно, должна быть связана с обсуждениями, инициированными пятнадцать лет назад группой Роже Педок (Roger T. Pédaucque – *коллективный псевдоним группы авторов*) на тему «Документ в свете цифровых технологий» (Le document à la lumière du numérique, <https://bbf.enssib.fr/consulter/bbf-2007-04-0122-012>); данная публикация, в частности, настаивала на повторном документальном оформлении (redocumentarisation, <https://fr.wikipedia.org/wiki/Redocumentarisation> - *это идея повторной документальной обработки при переводе бумажных документов в электронную форму, включая разрыв связи между информацией и её носителем, обогащение метаданных и т.д. – Н.Х.*), поддерживаемой технологиями декомпозиции-перекомпоновки документов, - то было время, когда «данные» и «большие данные» еще не получили широкого распространения.

«Данные» преобладают в нормативном регулировании

Это постепенный уход в тень «документа» ввиду его вытеснения «данными» можно наблюдать как в речи, так и в реальности. По сути, термин «документ» постепенно приобретает оттенок архаичности; мы настолько привыкли к бумажным документам формата А4 и PDF-файлам, что сочетание термина «документ» с цифровыми объектами в иной форме (базы данных, видео, онлайн-формы, посты в социальных сетях и т.д.) становится всё менее и менее понимаемым.

Но что такое документ? Наиболее распространенное в информатике определение (см. <https://fr.wikipedia.org/wiki/Document>) - это сочетание информации и носителя, на котором она хранится. Такое определение весьма широкое и очень удобное для цифровых данных. Если даже мы вспомним, что этимология слова «документ» связана с тем, что он передает информацию или даже учит (на латыни *doceo* – «учить»), мы всё ещё придерживаемся представления, далеко выходящего за рамки «формата А4».

Для тех, кто никогда не читал брошюру Сюанны Брие (Suzanne Briet) «Что такое документация?» (Qu'est-ce que la documentation? , 1951, <http://martinetl.free.fr/suzannebriet/questcequeladocumentation/>), еще не всё потеряно, и я рекомендую им воспользоваться приведенной гиперссылкой. Среди прочего, они узнают, как пойманную исследователем антилопу можно рассматривать как первичный документ, с которым взаимосвязаны различные описательные научные документы.

Давайте дискутировать, - это всегда полезно, а нередко и приятно. Но давайте не будем забывать о влиянии нормативного регулирования на административную практику и на связанную с ней терминологию. Нормативная база также формирует реальность.

С этой точки зрения «Общие правила защиты персональных данных» (General Data Protection Regulation, GDPR, <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1532348683434&uri=CELEX:02016R0679-20160504> – закон Евросоюза о

защите персональных данных), действующие с мая 2018 года, являются **важным терминологическим событием.**

Действительно, этот ключевой для защиты персональных данных закон, который в значительной степени определяет стратегическое управление данными в организациях на протяжении более трёх лет, уже привёл к найму десятков тысяч «уполномоченных по защите персональных данных» (*délégués à la protection des données, DPO*); который, безусловно, нацелен на борьбу с произволом GAFAM (*большая пятёрка американских транснациональных технологических компаний, включающая Google, Amazon, Facebook, Apple и Microsoft*), но также определяет поведение десятков тысяч организаций, - **этот закон говорит исключительно о «данных», и никогда - о «документах».**

Официально в мире защиты персональных данных у «документа» больше нет прав гражданства. И это не фейковые новости!

Что еще более поразительно, из англоязычной версии закона GDPR полностью исключён термин *record* (документ) в пользу *data* (данные). Однако англоязычный термин «*record*» куда более точен, чем термин «*document*» во французском языке (в англоязычной архивной терминологии понятия *record* и *document* - две очень разные вещи); он также намного шире в плане формы и формата, потому что в английском языке «*record*» относится не только к листам формата A4, но и к любому типу записи данных.

Закон GDPR, безусловно, является ударом по дисциплине управления документами. Как только «документы» - по крайней мере, с терминологической точки зрения, - «уходят» из области применения GDPR, становится ещё труднее, чем раньше, предложить организации развернуть проект «управления документами» (во французской терминологии английскому термину «*records management*» соответствуют термины «*archivage managérial*», «*gouvernance des documents engageants*»); и подобная эволюция международного нормативно-правового регулирования является потенциально смертельным ударом по стандарту управления документами ISO 15489 (его первая редакция вышла в 2001 году, вторая – в 2016-м), который когда-то так превозносился (и вполне справедливо), а сегодня, хотя и находится всё ещё в молодых годах, уже несколько подзабыт. Мой пост о лексике закона GDPR см. здесь: <https://www.arcateg.fr/2018/01/29/rgpd-gdpr-choix-mots/>.

Каждый может проверить, в 99 статьях закона GDPR слово «документ» практически не используется для обозначения документального объекта, частично или полностью состоящего из персональных данных. Единственным исключением, лишь подтверждающим закономерность, является статья 86 о персональных данных, содержащихся в официальных документах. Это, конечно, не означает, что документы – раз уж их определенное количество имеется в организациях - не содержат персональных данных или что персональные данные не могут быть «обработаны», чтобы сформировать документальный объект, именуемый «документ». Однако терминологическая революция здесь налицо.

Определенным утешением для нас может служить то, что GDPR, с другой стороны, в нескольких местах настаивает на необходимости «задокументировать» определённые ситуации, такие, как, например, утечки данных ...

Законодатель вполне сознательно при установлении порядка защиты персональных данных не делает различия между персональными данными в документах и в не-документах – и правильно, поскольку в противном случае закон обходился бы и превращался бы в посмешище путём заявлений о том, что «документов у нас нет, а есть только информация». А вот там, где оператор должен доказывать законность обработки персональных данных, адекватность принятых мер или информировать об утечках – в законе везде говорится именно о «документах» и «документировании»!

В статье 4 закона GDPR «Определения», персональные данные определяются как «любая информация, относящаяся к идентифицированному или идентифицируемому физическому лицу ...». **«Обработка персональных данных»**, новый и важный термин закона, определяется как «любая операция или набор операций, выполняемых с персональными данными или с наборами персональных данных, вне зависимости от того, используются или нет автоматизированные средства, - таких как сбор, запись, организация, структурирование, сохранение, адаптация или модификация, извлечение, просмотр, использование, раскрытие посредством передачи, распространения или любой другой формы предоставления; согласование или комбинирование, ограничение, стирание или уничтожение». **Было бы интересно сопоставить это перечисление с формулировкой Жерома Дени: «операции записи и чтению, поддерживающие распространение данных».**

В статье 13 закона GDPR говорится, что оператор (представитель юридического лица, собирающего персональные данные) должен сообщить субъекту персональных данных, помимо прочего, срок хранения персональных данных или, если это невозможно, критерии определения этого срока, - иногда довольно деликатная операция, к которой я вернусь в третьей части своего поста.

Один вопрос, который, с моей точки зрения, недостаточно изучен - это **степень пересечения области в пространстве информации, охватываемой законом GDPR, и области, охватываемой управлением документами.** Ведь очевидно, что как специалисты, отвечающие за защиту персональных данных в организации в соответствии с требованиями GDPR, так и те, кто отвечает за архивацию (то есть за обеспечение их защиты во времени), на практике заинтересованы в одних и тех же «вещах», какое бы имя мы им ни дали. Этот вопрос, заданный в 2017 году во время круглого стола, совместно организованного «Клубом ответственных за политику и проекты архивации» (Club des Responsables de Politiques et Projets d'archivage, CR2PA, см. <http://blog.cr2pa.fr/2017/10/cr2pa-rgpd-et-archivage-managerial-les-points-forts-de-la-table-ronde-du-11-octobre/>) и группой L'Oréal, позволил увидеть различные интерпретации такого пересечения. На продолжающуюся

дискуссию по данному вопросу, несомненно, негативно влияет нестыковка академических дисциплин и направлений деятельности в организации, хотя в значительном числе случаев обязанности уполномоченного по защите персональных данных и отвечающего за архивацию выполняются одними и теми же людьми.

«Документ», застрявший между «архивными документами» и «данными»

В 2015 году историк Бертран Мюллер (Bertrand Müller) написал следующее (в неопубликованном тексте, который автор передал мне): «Судьбы этих двух понятий [элемент данных (donnée) и архивный документ (archive)] теперь, кажется, взаимосвязаны, хотя эта связь является одновременно парадоксальной и недавней. Она парадоксальна, потому что эти два понятия вторглись в семантическое пространство интернета; они оба распространены повсеместно. В то же время эти два понятия также обозначают различные вещи или явления, которые были полностью переопределенные с появлением цифровых практик».

«Документ» (document) уже не обсуждается ...

Но если «документ» исчезнет, то что будет архивироваться завтра? Данные?

Традиционно архивные документы определяются как «документы», созданные при осуществлении деятельности, сохранение сведений о которой представляет интерес (эта формулировка является кратким изложением статьи 1 французского Закона об архивах 1979 года). Во время пересмотра Закона об архивах в 2008 году (этот закон интегрирован в Кодекс культурно-исторического наследия» (Code du patrimoine), созданный в 2004 году), была сохранена эквивалентность «archives = documents», с учётом широкой трактовки понятия document (охватывающего носители всех видов, все формы...), при этом сфера охвата сужается за счёт упоминания качества (qualité) и законченности (finalité) документов, которые и делают их «архивными документами» (documents d'archives).

«Данные» появляются в юридическом определении понятия «архивы» лишь в 2016 году (пять лет назад, за два года до GDPR), вместе со статьей 59 закона № 2016-925 от 7 июля 2016 года, которая заменяет формулировку «Архивы - это совокупность документов, вне зависимости от даты их создания. ...» на «Архивы - это совокупность документов, **включая данные**, вне зависимости от даты их создания... ». Мы видим, что в определение была сделана короткая вставка, вызывающая недоумение. Следует ли понимать это таким образом, что данные - это подмножество документов? Или что в документах обязательно есть данные, и данные вскользь упомянуты лишь для тех, кто в этом сомневается? Или что одни архивные документы состоят из данных, а другие нет? Означает ли это, что не включенные в документ данные не будут архивными документами и не будут архивироваться? В этом хотелось бы поподробнее разобраться, и данный вопрос заслуживает проработки для условий, в которых подавляющее большинство потенциально

подлежащей архивации информации создается с помощью цифровых технологий.

Давайте подождём.

Архивисты, с другой стороны, заинтересованы в данных, как показало проведённое с участием исследователей обсуждение вопроса об использовании архивных документов с помощью технологий. Как историки, так и архивисты долгое время работали с базами данных, опираясь на анализ и изучение архивных документов. Этот вопрос стал темой учебного семинара, организованного в июне 2019 года региональной группой Аквитании (Aquitaine) Ассоциации французских архивистов (l'Association des archivistes français, AAF) в партнерстве с рабочей группой «Цифровые исследования: от данных к системе / структуре» (Études digitales: des données aux dispositifs – E3D) исследовательской лаборатории «Посредничество, информация, коммуникация, искусство» (Médiation, information, communication, arts - MICA) университета Бордо-Монтень (Université Bordeaux-Montaigne). Название семинара **«Документ, бедный родственник данных. В поисках утраченных архивных фондов»** (Le document, parent pauvre de la donnée. À la recherche du fonds d'archives perdu, https://mica.u-bordeaux-montaigne.fr/wp-content/uploads/2019/06/Document_parent_pauvre_donnee_programme.pdf) особенно примечательно своим «триптихом» «архивы – документы – данные».

Этот подход новый и экспериментальный, и понятно, что он начинается с изучения уже собранных, существующих архивных фондов, созданных на аналоговых носителях, прежде чем погрузиться в полностью цифровой мир. Имеющийся опыт, что совершенно не случайно, больше относится к оцифровке бумажных архивных документов (процесс, следующий за созданием архивных документов), чем к производству изначально-электронных «документов». Интересно, однако, увидеть взаимосвязь между архивными документами и данными, которые могут быть извлечены из них, с возможной проекцией процесса в противоположном направлении, от данных к документам.

Аналитический отчет, написанный в 2017 году Жильбером Кутазом (Gilbert Coutaz) в сотрудничестве с Жилем Жанмоно (Gilles Jeanmonod) и озаглавленный **«Место персональных данных в исторических архивах: попытка интерпретации в контексте медицинских архивных документов в архивах кантона Во»** (La place de la donnée personnelle dans les archives historiques, essai d'interprétation à travers les archives de santé aux Archives cantonales vaudoises, https://www.vd.ch/fileadmin/user_upload/organisation/chancellerie/ACV/fichiers_pdf/dossier-thematique/Dossier-thematique-2017.pdf, 40 страниц), содержит размышления на стыке архивной науки и права, точнее, на перекрестке архивной науки и законодательства о защите персональных данных. В тот момент, когда писалась эта работа, европейские архивисты высказали свое мнение по поводу проекта европейского закона (будущего GDPR) ввиду определенных радикальных толкований «права быть забытым»,

ведущих к безвозвратному уничтожению архивных документов или к их несохранению, которые, однако, будучи следами жизни отдельных людей, являются незаменимыми источниками для коллективной истории.

В отчёте рассматривается «место персональных данных в процессах архивирования, и весь жизненный цикл документов от их разработки до окончательного решения их судьбы - либо уничтожения, либо передачи на хранение в архив для целей сохранения культурно-исторического наследия и проведения исторических исследований». Цитируя опубликованный в 2002 году отчет группы историков, касающийся истории и медицинских архивов, авторы напоминают, что проведение оценки содержащих персональные данные документов предполагает, что эти документы предлагаются для передачи на архивное хранение до того, как можно будет оценить необходимость их сохранения. Еще предстоит увидеть, как эволюционировало за последние два десятилетия производство медицинской информации, от цифровых файлов, являющихся наследниками бумажных документов прошлого, до баз данных информации о здоровье, подпитываемых постоянными, общими, связанными потоками; коллективно используемых, соединённых друг с другом и «обрабатываемых»... Между тем, подсчёт в этом тексте количества трёх рассматриваемых нами терминов дает следующие результаты: «данные» - 185; «архивы, архивные документы» - 251; «документ» - 50.



Часть 3: Документы не сдаются (пост 1)

Используя политическую метафору «господства» данных, можно сказать, что документ доминировал в «дореволюционную» эпоху. Однако смена режима, даже быстрая и жестокая (к счастью, в нашем случае мы не имеем дела с кровопролитием), не проходит без сопротивления среди населения (пользователей), без вмешательства, без возрождения в долгосрочной перспективе определенных идей.

Документ не сдаётся и не признаёт поражения (что было бы позором) концепции, носителем которых он был так долго, в частности, концепции

конечного, ограниченного, датированного информационного объекта, с ясным происхождением, на который можно ссылаться, или даже представляющего документ «вектора», идущего от автора, у которого есть намерение, к заинтересованному получателю – эти концепции не исчезают из-за преобладания данных. Я бы даже сказала, что, наоборот, документам предлагается защищать свою нишу и оттачивать свои аргументы, даже если это означает изменение их наименования. В политике так делают ...

«Документ» сопротивляется ...

Столкнувшись с «данными», которые навязываются в мире информации, в сфере деловой деятельности, в средствах массовой информации, термин «документ» сопротивляется. В противном случае был бы повод для беспокойства.

Можно наблюдать несколько форм такого сопротивления, рассматривая документ как объект, как правовое понятие и даже как привычное словоупотребление.

У многих людей, особенно в офисной среде, слово «документ» ассоциируется с образом прямоугольного объекта, чаще всего белого цвета, на который нанесены небольшого размера черные символы. Как в бумажной, так и в электронной форме, документ представляет собой объект, видимый на столе или на экране. При оценке силы слова не следует упускать из виду визуальный аспект.

Бумажный документ на самом деле не конфликтует с «данными». Он живет своей жизнью, постепенно приходя в упадок (бумажная почта была заменена электронной, административные документы превратились в онлайн-формы и т.д.). Бумажный документ в некотором роде пришёл на перекрёсток: слева архивы, справа данные. Либо понятие «документ» привязывается к практике прошлого с её старомодным шармом, либо документы начинают создаваться в изначально-электронном виде, что сейчас является нормой при повседневно обмене информацией. Документ должен сделать свой выбор.

Ситуация с электронным документом несколько иная. Действительно, электронный документ - это в большей степени представление в электронной форме документа, «продуманного на бумаге», чем продукт цифрового мира. Электронный документ родился в 1980-х годах благодаря инструментам оцифровки (сканирования) и системам электронного документооборота (*gestion électronique de documents*, GED, https://fr.wikipedia.org/wiki/Gestion_%C3%A9lectronique_des_documents#Histoire). Фактически, именно с этого, не столь отдалённого времени термин «документ» практически повсеместно ассоциируется с образом документа формата A4 (хотя это совсем иное). Формат PDF (см. https://fr.wikipedia.org/wiki/Portable_Document_Format) был изобретен десятью годами позже и стандартизирован Международной организацией по стандартизации (ИСО) в 2008 году (что стало большой победой компании Adobe). Сегодня «PDF» практически стало именем нарицательным (и используется даже как кошачье имя Pédée, см.

<https://www.marieannechabin.fr/wp-content/uploads/2016/10/Une-journee-dans-la-vie-de-lili-eting-versionpdf.pdf>).

Но эпоха, когда жили по принципу «я произвожу бумагу и сканирую её; я продолжаю производить бумагу, чтобы продолжить её сканирование», прошла. Представление об электронном документе как об отсканированном бумажном документе формата А4 уже еле держится ввиду распространения технологий производства изначально-электронных «освобождённых» материалов, то есть разработанных с учётом многих возможностей цифровой среды, сочетающих функциональные возможности программного обеспечения. и сетей. Сторонники электронного документооборота (GED), который когда-то был передовым, стали теперь отсталыми перед лицом «менеджеров данных». GED достиг почтенного возраста (сорок лет, см. <https://www.cnrtl.fr/definition/canonique>) и уже не сможет долго продержаться.

Это не только моё мнение. «Электронному документообороту» пришёл конец!» (La GED, c'est fini ! , <https://www.linkedin.com/pulse/la-ged-cest-fini-philippe-goupil>) - заявил Филипп Гупиль (Philippe Goupil) в 2017 году: «Сейчас уже слишком поздно рассматривать GED как решение на будущее!». Электронный документ освободился от оков, которыми его опутал его отец, GED, и бросился в объятия «данных». Игра ещё не окончена; я ясно вижу это в вопросах моих студентов, которые спрашивают меня о системах GED (потому что в компаниях, в которых они проходят стажировку, такие системы всё ещё порой существуют), - но, к сожалению для этого понятия, конец GED уже предрешён.

На другом уровне «документ» также сопротивляется, потому что данное понятие было включено во французские правовые кодексы со времен Кодекса Наполеона. Даже если бы «данные» были вполне признаны законодательство, от двухвекового правового наследия за один день не избавишься.

Поиск, проведенный несколько дней назад на сайте Legifrance (*французский правовой портал*), дал следующие результаты: в статьях 67 кодексов слово «документ» встретилось 8631 раз. Много это или мало, сказать сложно - на чём будет основано сравнение? Но можно сказать, что аналогичный запрос выдает 5750 хитов для «данных» (données). Если учесть, что « donnée(s) » в более чем 50% случаев - это просто причастие прошедшего времени, образованное от глагола «donner» (что искажает сопоставление), то по большому счёту можно примерно прийти к выводу о том, что термин «документ» встречается в законах и нормативных актах в три-четыре раза чаще. Поскольку термин «данные» в основном используется в текстах последних лет, «баланс сил» терминов «документ» / «данные» вскоре должен измениться в обратную сторону. Анализ станет ещё более интересным, если мы включим в него 14061 упоминание термина «информация», - но на этом я закончу это отступление.

Несколько замечаний по поводу цитат из законодательства, которые Legifrance выдал в ответ на мой запрос:

- прилагательные или дополнения к именам существительным, связанные со словом «документ», обычно определяют содержание или процесс, в результате которого создаются указанные документы: бухгалтерские документы (documents comptables), документы планирования (documents d'aménagement), избирательные документы (documents électoraux), устанавливающие цели документы (documents d'objectifs), удостоверяющие личность документы (documents d'identité);
- другие квалификаторы подчеркивают статус документа: обязательный (obligatoire), подтверждающий (justificatif), необходимый (nécessaire), имущественный (patrimonial);
- «документ» ассоциирован с другими существительными в различных выражениях, которые отражают различные, взаимодополняющие роли между: «сведениями» (renseignements) и документами (Кодекс о защите прав потребителей), «книгами» (livres) и социальными документами, передаваемыми партнерам (Гражданский кодекс), документами и «отчетами» (rapports), передаваемыми аудитору (Коммерческий кодекс). «произведения» (œuvres) или документы порнографического характера (Кодекс кино), «правоустанавливающие документы» (titres) и документы (Кодекс о государственном имуществе), реестры (registres) и документы, ведение которых является обязанностью эксплуатирующих аэродромы организаций (Кодекс гражданской авиации).

Использование термина «данные» «вместо документа» началось сравнительно недавно и связано с защитой персональных данных, а точнее, со сроком хранения этих данных: «данные о гражданском состоянии и данные, относящиеся к социальной ситуации» (Кодекс въезда и пребывания иностранцев, и право на убежище, статья R142-56); «данные, касающиеся отслеживания обмена информацией, хранятся в течение 36 месяцев [...]» (Кодекс взаимоотношений между общественностью и администрацией, статья R114-9-7); «данные об уровне алкоголя в крови водителей не подлежат раскрытию, передаче или использованию. [...] » (Правила дорожного движения, статья L234-15).

Идёт процесс распространения термина «данные» («датыфикация») в нормативных документах. Однако мы можем задаться вопросом, смогут ли «данные» во всех случаях заменить «документы», даже если переосмыслить производство информации в цифровом контексте? Хм ...

Приведенные ниже два фрагмента из кодексов содержат оба слова, «документ» и «данные», в одном предложении.

Раздел 4 Кодекс общественного здравоохранения (Code de la santé publique) называется «Условия признания доказательной силы документов, содержащих персональные данные о здоровье, созданных или воспроизведенных в электронной форме, и уничтожения документов, хранящихся в неэлектронной форме», из чего мы можем сделать вывод о том, что электронные документы содержат данные (неважно, персональные или нет) и, возможно, что-то ещё (или нет), - но, исходя из этой формулировки,

мы не можем сказать, содержат ли данные также и неэлектронные документы.

Уголовный кодекс (code pénal) в статье 413-11 о военной тайне (secret de la défense nationale, см. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000020933031) перечисляет «вещи», несанкционированный доступ, уничтожение или разглашение которых карается пятью годами тюремного заключения и штрафом в размере 75 тысяч евро, и это: «процедура, объект, документ, информация, компьютерная сеть, элемент электронных данных (donnée informatisée) или файл». Потенциальные связи между этими различными сущностями не уточнены. Попутно отметим, что термин «элемент данных» в единственном числе (donnée) в нормативных текстах встречается очень редко.

Такого рода лексический анализ можно продолжать и продолжать, выявляя новые подробности, но общая тенденция уже ясно видна.

В заключение приведу два примера того, как «документ» оказывает сопротивление:

Единый документ по оценке рисков (document unique d'évaluation des risques, DUERP, <https://www.fonction-publique.gouv.fr/document-unique-devaluation-des-risques-professionnels-duerp>), обязательный для всех компаний при приёме на работу первого сотрудника, практически никогда не выглядит как традиционный документ формата А4; сегодня это файл формата Excel с вкладками или же полноценная база данных, однако название сохранилось.

Во время столь популярных в наше время утечек данных, «документ» занимает хорошее место: очень часто используется английское слово «papers» (бумаги, документы) (Панамские документы, <https://www.marieannechabin.fr/2016/04/panama-papers-4-remarques-sur-la-forme/>), документы Пандоры, <https://www.icij.org/investigations/pandora-papers/>); а средства массовой информации подчёркивают утечку миллионов «документов» (documents), наряду с данными, файлами и архивами (документами - records).

Часть 3: Документы не сдаются (пост 2)

Документ в виде совокупности данных

Давайте отложим в сторону термины и вернемся к реалиям, к которым они относятся.

Цифровые технологии разрушили ограничивающие рамки материального носителя и в каком-то смысле «освободили» контент.

Как следствие, «документ» имеет тенденцию уходить в тень «данных», которые он содержит (документ содержит данные) или которые его образуют (документ состоит из данных). В любом случае документ уходит на вторые роли, поскольку данные выходят на первый план и привлекают внимание.

Это особенно характерно для цифровой обработки книг в библиотеках или технической документации в инженерных службах. «Доцифровой» документ разбивается на индивидуализированные элементы данных, по

которым поисковая система может вести поиск, и которые можно выделить и извлечь. «Обеспечивая возможность компьютерной обработки самого контента, оцифровка шаг за шагом привела к изменению понимания вещей»; «документ представляет собой совокупность данных, которые можно использовать», - пишет Готье Пупо (Gautier Poupeau), автор блога Les petites cases («Маленькие ящички», <http://www.lespetitescases.net/>) в образовательной презентации под названием «Экскурсия по стране данных» (Visite guidée au pays de la donnée, <https://drive.google.com/drive/folders/1eAqAVKPEgVm4dgxiqIPw2NL5wyGtI7R?sort=13&direction=a>). Автор подчеркивает изменение степени детализации обрабатываемой информации благодаря современным технологиям.

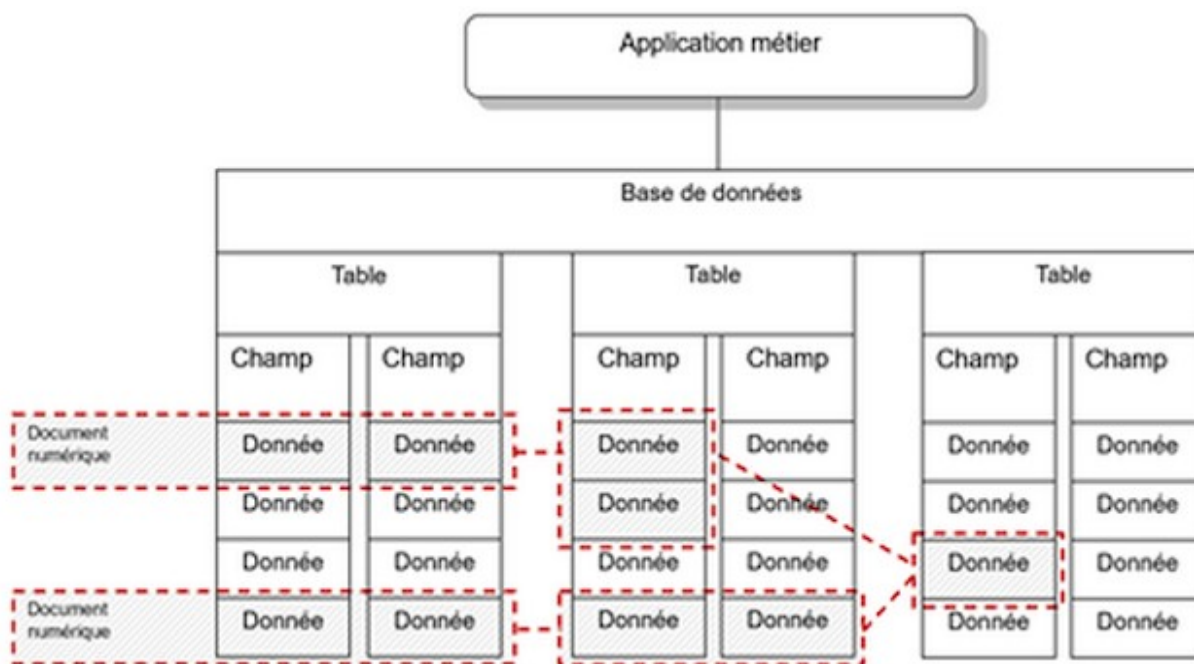
Это затенение «документа» «данными» или даже «элементами данных» также имеет место тогда, когда документ существует лишь виртуально, возникая после производства данных в результате агрегирования группы соответствующих данных.

В ряде международных стандартов по вопросам управления электронными документами (фр.: records management électronique, archivage des informations numériques engageantes) используется понятие агрегации (aggregation – *речь идёт об упорядоченных наборах документов различного уровня, от папки и дела и вплоть до фонда*). Это касается спецификаций MoReq2010 (<https://www.moreq.info/>) и стандарта ICA-Req (<https://www.ica.org/fr/node/14871>), опубликованного в 2008 году и впоследствии ставшего международным стандартом ISO 16175.

Термины «агрегация» (agrégation) или «агрегат, конгломерат» (agrégat) понимаются как совокупность документов, но также и как совокупность данных: это «любая совокупность документных объектов на уровне выше индивидуального документного объекта (документ, электронный объект), - например, электронное дело или серия». Хотя эти стандарты, написанные 10-15 лет назад, в первую очередь имеют в виду объединение (агрегацию) документов в дела, термин «агрегация» хорошо иллюстрирует логическую конструкцию, в рамках которой на основе существующих данных создаётся информационный объект, который в доцифровом мире назывался «документом», не задавая лишних вопросов.

Приведенный ниже рисунок, взятый из части 3 стандарта ICA-Req («Функциональные требования и рекомендации по управлению документами в деловых системах»), снабжён подзаголовком «Идентификация информационных компонентов / элементов данных, составляющих электронный документ в базе данных» (Identification des composants d'information (ou données) constituant un document électronique engageant dans une base de données).

Идея проста, когда мы имеем дело с так называемыми структурированными данными, образующими «документ» типа формы с хорошо индивидуализированными полями, такой, как платежная ведомость или счёт-фактура.



Для других случаев, то есть в случае документов, подпадающих под определение «неструктурированных данных» (<https://www.arcateg.fr/2018/09/06/donnees-structurees-et-donnees-non-structurees/>) агрегирование является более сложным вопросом.

Чтобы проиллюстрировать этот сценарий, я воспользуюсь материалами публикации, упомянутой в моём обзоре литературы в начале части 1 этого поста - «Является ли копия письменной экзаменационной работы набором персональных данных? (Une copie d'examen manuscrite est-elle un ensemble de données à caractère personnel?», <https://www.droit-technologie.org/actualites/copie-dexamen-manuscrite-ensemble-de-donnees-a-caractere-personnel/>), выложенной на отличном блоге «Право и технологии» (Droit et technologies). Речь шла о бумажном документе, однако приведенный анализ взаимосвязи между «документом» и «данными» весьма красноречив.

Суд Европейского Союза (Cour de Justice de l'Union Européenne, CJUE) рассматривал иск провалившего экзамен ирландца, которому было отказано в возможности ознакомиться с его экзаменационной работой на том основании, что она «не представляет собой персональные данные в смысле законодательства о защите персональных данных». Оказалось, что да, экземпляр письменного теста содержит персональные данные: «в той мере, в какой в ней фиксируется информация, относящаяся к кандидату, она представляет собой набор персональных данных». Уточняется, что сам по себе рукописный текст может содержать информацию о человеке; а также что комментарии проверяющего также считаются персональными данными кандидата.

Из этого примера экзаменационной работы очевидно, что сегодня данные - это нечто намного большее, чем числа или значения, которые мы вводим в поля, хотя такого рода определение остаётся наиболее

распространенным в глоссариях и словарях, которые, как я отмечала в первой части моего поста, не являются актуальными.

Говоря иными словами, не следует ли нам считать, что экземпляр экзаменационной работы является неделимым, неразделимым (заполнение и отметки проверяющего) целым, и что если данные могут быть извлечены из неё для той или иной обработки, эта совокупность (файл, набор данных) тем не менее сохраняет свой статус «документа» (document) в смысле определений «зафиксированная информация либо объект, которые могут восприниматься и обрабатываться как единое целое» (ISO 15489-1:2001 – *это по сути дела определение понятия «информационный объект», который в быту также известен как «файл»*), «любой текст или запись, рассматриваемые как документальные единицы» (*это французское определение можно найти здесь: Glossaire PIAF, "Document", https://www.piaf-archives.org/sites/default/files/bulk_media/glossaire/glossaire_papier.pdf*)?

Именно поэтому тот факт, что слово «документ» удалено из французского и английского текстов закона GDPR удивил меня своим радикализмом. GDPR требует, чтобы устанавливались сроки хранения персональных данных, и чтобы о них был проинформирован субъект персональных данных. Но как применить сроки хранения в отношении элементарным данных? И как применить сроки хранения (и выполнить действия по их уничтожению) к группе данных, которая не рассматривается как отражение взаимодействия между двумя людьми (отправителем и получателем, источником и производителем) с конкретной целью, связанной с определённой датой?

Бертран Мюллер (Bertrand Müller), уже цитировавшийся во второй части этого длинного поста, далее говорит: «Изолированный фрагмент данных, то есть отделённый как от контекста его создания, так и от контекста его достоверности, является бессмысленными данными. Документирование данных означает фиксирование сведений о том, как они были спроектированы, что они означают, а также специфицирование их контента и структуры. Документация также имеет решающее значение для обеспечения долговременной сохранности данных».

«Данные и документы»

Моя цель - показать тем, кто не осознает, что доминирование «данных» возникло не спонтанно, и основано, посредством переименования понятий, как мы это делаем в политике, на существующем и часто довольно старом документальном субстрате. Например, на рисунке мы можем увидеть: структурированный документ (например, табель), поток поступающих данных и два тематических досье. Документ, как совокупность (агрегация) данных, является виртуальным - даже если он является предметом целевой записи информации - и его скрепляет воедино общая ценность собранных таким образом данных.

Часть 3: Документы не сдаются (пост 3)

Данные и «письменная форма»

Во второй части этой публикации я подчеркнул то значение, которое Жером Денни (Jérôme Denis) в его социологическом исследовании «данных» придает операциям записи и чтения, которые имеют место при создании и распространении данных.

Мы все больше и больше говорим о жизненном цикле данных так же, как мы говорили (и до сих пор говорим) о жизненном цикле документа, - с очень разных точек зрения: одни рассматривают статус данных от момента их создания и до повторного использования, в то время, как другие ограничивают вопрос стадиями управления и движения в информационной системе или в организации; и я вряд ли осмелюсь упомянуть парадоксально узкое логистическое видение Межведомственной службы Архивов Франции (Service Interministériel des Archives de France, SIAF) и Национальной комиссии по информатике и свободам граждан (Commission nationale de l'informatique et des libertés, CNIL – *регулятор в области защиты персональных данных*) - см. <https://www.cnil.fr/fr/les-durees-de-conservation-des-donnees> .

Мне кажется важным принять во внимание происхождение данных (будь то имена, значения, числа, знаки, мнения и т.д.), т.е. когда они впервые записываются, где-то регистрируются, формализуются, фиксируются перед передачей, сохраняются, «работают» (используя термин Жерома Денни), - чтобы, при необходимости, их можно было прочесть, интерпретировать, повторно использовать, перезаписать, перерегистрировать, переформализовать, зафиксировать в новом контексте с новыми действующими лицами, новыми намерениями и целями.

Со времён закона от 13 марта 2000 года (см. <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000399095/>) об адаптации порядка доказывания к информационным технологиям и касающийся электронной подписи (реализующего в национальном контексте Европейскую директиву от 13 декабря 1999 года по тому же вопросу), мы регулярно слышим от действующих лиц в сфере «цифровизации» и электронного документооборота (GED) о том, что **электронный документ имеет такую же юридическую силу, как и бумажный документ**, если его автор аутентифицирован и если целостность документа сохраняется с момента его создания. Однако **в законе не говорится в данном случае о «документе»; речь в нём идёт о «письменной форме» (écrit).**

В статье 1366 Гражданского кодекса конкретно сказано: «Электронный [документ] в письменной форме (écrit électronique) имеет такую же доказательную силу, что и бумажный [документ] в письменной форме, при условии, что лицо, от которого он исходит, может быть надлежащим образом идентифицировано, и что он создан и хранится в таких условиях, которые гарантируют его целостность».

Следует отметить, что до 2016 года это была статья 1316-1, формулировка которой была такой же, за исключением двух слов:

«письменный документ в электронном виде» (l'écrit sous forme électronique) впоследствии превратился в «электронный документ в письменной форме» (l'écrit électronique). Я рассматриваю это как обычную апроприацию. Возможно, в будущей версии мы увидим «цифровой документ в письменной форме» (l'écrit numérique)?

Определение письменной формы дано в статье 1365 того же Гражданского кодекса: «Письменная форма (écrit) состоит из ряда букв, знаков, цифр или любых других знаков или символов, наделенных понятным значением, независимо от их носителя». Эта формулировка, которая также датируется 2016 годом, основана на предыдущей статье 1316, определяющей понятие «письменного доказательства» (preuve par écrit): «Письменное доказательство (preuve littérale) или доказательство в письменной форме (preuve par écrit) состоит из ряда букв, знаков, цифр или любых других знаков или символов, наделенных понятным значением, независимо от их носителя и способов передачи». По этому поводу см. статью Беатрис Френкель (Béatrice Fraenkel) и Давида Понтия David Pontille «Юридический документ в письменной форме, удостоверенный электронной подписью - Прагматический подход» (L'écrit juridique à l'épreuve de la signature électronique, approche pragmatique, <https://www.cairn.info/revue-langage-et-societe-2003-2-page-83.htm>), опубликованную в журнале Langage et société, 2003/2 (выпуск 104), страницы 83–122.

Это определение «письменной формы», как мы видим, намного шире, чем определение «документа», которое одновременно означает гораздо больше (в других значениях) и вызывает гораздо меньшие ассоциации (ограничение, связанное с ассоциацией с прямоугольной формой документа в рамках электронного документооборота, см. выше).

«Письменная форма» вполне совместима с «данными».

Таким образом, термин «письменная форма» (écrit), несмотря на его несколько старомодный оттенок (можно подумать о рукописи и представить себе перо или авторучку, а не клавиатуру), вполне подходит для той вещи, которая возникает в результате записи информации физическим лицом, который выражает себя, и даже лицом, которое утверждает что-либо о других лицах.

Интерпретации понятия «документы в письменной форме» (documents écrits) можно преимущественно разделить на письменную форму, которая что-либо отслеживает (запись) и на документ, который наставляет и учит (docere, documentum). Письменная форма в смысле записи – это действие субъекта, автора; документ же информирует об объекте. Однако одна и та же сущность, скажем, группа данных – используя популярное слово, иногда может быть «письменной формой», а иногда – «документом», или, скорее, обоими сразу, но с разных точек зрения.

Например, упомянутая выше экзаменационная работа, рассматриваемая как набор персональных данных, в том числе за рамками слов (рукописный текст, написанный кандидатом - ввиду графического представления – является не одним словом, а всеми), - которую я бы

квалифицировала как неразделяемую агрегацию для целевого использования, - в первую очередь является продуктом совершенной кандидатом операции написания (*inscription*) на носителе, работа с которым не прекращается с закрытием экзаменационной комнаты, но ограничена экзаменационным процессом; и, следовательно, включает в себя внесение замечаний экзаменаторами. Этот экземпляр является и свидетельством юридически значимого действия, и основание для дальнейших усилий по обучению. Добавлю, что он имеет вдвойне обязывающий характер (и для кандидата, и для экзаменатора и учреждения, в котором тот работает).

По вопросу о «письменной форме» существует обширная литература. Документы в письменной форме уже произведены в больших масштабах и доступны (архивные документы). Это также происходит благодаря взрывному росту числа видов «письменной формы», допускаемых и предлагаемых цифровыми технологиями. **Это область применения дипломатики**, дисциплины с многовековой историей, традиционно занимавшейся исследованием подлинности древних актов с использованием ретроспективного подхода, но также занимающейся сейчас, в рамках более широкого и прогностического подхода, исследованием подлинности материалов в письменной форме, их полноты и воздействия. Какие формальные характеристики письменной формы позволяют надёжно использовать информацию, которую та передаёт? Какие её недостатки имеют негативные последствия, как сегодня, так и в долгосрочной перспективе, для автора материала в письменной форме, его носителя или заинтересованных сторон?

Мои размышления обо всём этом тесно взаимосвязаны с чтением главы, написанной Жоржем Тессье (Georges Tessier) в книге «История и её методы» (*L'Histoire et ses methods*, изд-во La Pléiade, 1961, <http://www.gallimard.fr/Catalogue/GALLIMARD/Encyclopedie-de-la-Pleiade/L-Histoire-et-ses-methodes#>), где «письменная форма» часто упоминается.

Прочитую отрывок со стр. 665: «Любой правовой акт может быть зафиксирован в письменной форме, предназначенном для использования в качестве доказательства. Затем мы имеем дело с письменным актом (*acte écrit*), инструментом, как когда-то было сказано, подражания римлянам, и который до сих пор упоминается в языке международного публичного права. Письменный акт – это инструмент, который «документирует» правовой акт. Он представляет собой свидетельство, и немецкое слово *Urkunde* имеет именно то достоинство, что тесно связывает идею свидетельства с понятием письменного акта. Однако письменный акт может свидетельствовать не о правовом акте, а о чём-то ином. **Письменный акт может свидетельствовать о «существенных фактах, которые не содержат волеизъявления, но, тем не менее, могут иметь правовые последствия»** [цитируется высказывание дипломата А. Дюма]».

Шестьдесят лет спустя это понимание легко переносится в цифровой мир соединённых сетями организаций, платформ и социальных сетей.

Заключение

Завершая этот тройной пост и отвечая на вопрос о том, не вытеснили ли «данные» и не затенили ли они «документы», я бы сказала, что **доминирование «данных» на сегодняшний день скорее маргинализировало, чем свергло понятие «документа», заставляя того адаптироваться** в техническом, культурном и лингвистическом плане, если он хочет защитить свои цвета (которые не сводятся к чёрному и белому...), и своё наследие.

Исходя из этого, мы можем представить себе два сценария: либо мир данных будет сидеть на своём технологическом «троне», игнорируя другие информационные культуры и обрекая их на уход на дрейфующий субконтинент; либо возникнет плодотворный диалог между инновациями и опытом с целью извлечения выгоды из знаний и из адаптации других культур.

Я готова признать, что мои рассуждения неполны. Эти несколько страниц больше похожи на блокнот, заполненный после прочтения материалов, отражающих различные точки зрения, своего рода «описание проблемы мелкими штрихами», которое является лишь началом иного исследования.

Как отметила в своём комментарии в LinkedIn одна из моих читательниц, безусловно, есть **повод для организации крупного коллоквиума по данным**, который бы собрал вместе специалистов различных дисциплин, связанных как с гуманитарными науками (социология, право, дипломатика), так и с точными науками – точнее, с так называемыми «цифровыми» науками (алгоритмы, глубокое обучение, визуализация и т.д.), а также привлёк руководителей государственных структур и политиков. Давайте немного помечтаем ...



США: «АККРЕДИТОВАННЫЙ КОМИТЕТ ПО СТАНДАРТИЗАЦИИ X9» ОПУБЛИКОВАЛ НОВЫЙ СТАНДАРТ ANSI X9.141-2021 ЗАЩИТЫ ПЕРСОНАЛЬНЫХ ДАННЫХ И УВЕДОМЛЕНИЯ ОБ УТЕЧКАХ

Источник: сайт комитета X9 <https://x9.org/asc-x9-issues-new-data-protection-and-breach-notification-standard/>

Новый стандарт - важная веха в усилиях по защите уязвимой информации.

«Аккредитованный комитет по стандартизации X9» (Accredited Standards Committee X9 Inc., X9, <https://x9.org/>) опубликовал новый национальный стандарт США ANSI X9.141-2021 **«Защита финансовых и персональных данных и извещения об утечках»** (Financial and Personal Data Protection and Breach Notification).

Стандарт X9.141 - это состоящий из двух частей документ, в котором рассматривается как то, что происходит «до» защиты персональных данных, так и «после» уведомления об утечке. Существенно то, что стандарт дополняет и расширяет важные руководства Национального института стандартов и технологий (National Institute of Standards and Technology, NIST) и охватывает все виды чувствительных данных, как финансовых, так и нефинансовых.

Обе части документа можно приобрести (как по отдельности, так и вместе) по адресу <https://webstore.ansi.org/Search/Find?in=1&st=x9.141> (*цена двух частей составляет 160 долларов*).

Нарушения безопасности данных продолжают подвергать риску как миллионы потребителей, так и коммерческие организации, их сотрудников и другие заинтересованные стороны. Защита внутренней деловой информации и информации о потребителях - это общая ответственность всех вовлечённых сторон, включая поставщиков унаследованных и облачных услуг; организации, которые хранят, передают или обрабатывают чувствительную информацию; финансовые учреждения; сотрудников коммерческих организаций и отдельных потребителей.

Новый стандарт четко определяет требования к защите персональных данных и к уведомлению об утечках для отрасли финансовых услуг. Он применим в любых организациях любой отрасли, которые обрабатывают чувствительную персональную информацию, - и его внедрение станет основательным шагом на пути к созданию надёжной защиты как от изощренных киберпреступников, так и от обычных похитителей данных.

«Ключевой особенностью стандарта X9.141 является то, что он согласуется с мерами и средствами обеспечения информационной безопасности в деловой деятельности, содержащимися в концепциях Национального института стандартов и технологий (NIST) обеспечения кибербезопасности и неприкосновенности частной жизни», - говорит исполнительный директор X9 Стив Стивенс (Steve Stevens). «Это значительно облегчает усилия организации по обеспечению исполнения законодательно-нормативных требований. Вместе с документами NIST, стандарт X9.141 указывает коммерческим организациям чёткий путь, по которому им следует идти при удовлетворении своих потребностей в защите персональных данных».

Первая часть стандарта X9.141-1:2021 «Часть 1: Защита данных» (Data Protection) содержит требования и рекомендации по защите персональных данных в сфере финансовых услуг. Она основана на специальной публикации NIST, выбранной из соображений обеспечения совместимости с преобладающими подходами к оценке безопасности и защиты

неприкосновенности частной жизни, используемыми в других отраслях. Документ также ссылается на некоторые другие стандарты NIST, Международной организации по стандартизации (ИСО) и Комитета X9.

Вторая часть стандарта X9.141-2:2021 «Часть 2: Уведомления об утечках» (Breach Notification) основывается на части 1 и учитывает федеральные законы США и законы штатов, касающиеся обеспечения неприкосновенности частной жизни. В этом документе предлагается последовательный процесс уведомления об утечках с целью защиты и информирования потребителей и других заинтересованных сторон по всей стране, включая требования и всесторонние рекомендации.

«Основные отраслевые группы уже в течении ряда лет признавали очевидную потребность в стандарте защиты персональных данных и уведомления об утечках. В 2017 году в совместном письме Конгрессу (см. <https://www.aba.com/-/media/documents/letters-to-congress-and-regulators/joint-financial-services-trades-supporting-strong-data-security-legislation-110117.pdf?rev=fc2ade466a8b44799525576cbdf3e66d>) семь организаций призвали к созданию единого строгого национального стандарта для всех персональных данных, как финансового, так и нефинансового характера, чтобы устранить текущее противоречивое лоскутное одеяло законов. Стандарт X9.141 - это реализация данной цели», - отмечает Алан Тиманн (Alan Thiemann), адвокат и партнер компании Han Santos, PLLC и главный юрисконсульт компании Conexxus.

Тиманн далее говорит: «Крайне важно, чтобы коммерческие организации, будь то финансовые учреждения или любые другие отраслевые субъекты, имели возможность сконцентрировать свои ресурсы по обеспечению соответствия законодательно-нормативным требованиям на едином унифицированном стандарте защиты всех персональных данных, от данных о потребителях / платежах и до персональных данных сотрудников, претендентов на рабочие места, продавцов или поставщиков».

ИСО И МЭК: ГОТОВИТСЯ НОВАЯ РЕДАКЦИЯ СТАНДАРТА ISO/IEC 27005 «РУКОВОДСТВО ПО МЕНЕДЖМЕНТУ РИСКОВ И ВОЗМОЖНОСТЕЙ В ОБЛАСТИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ»

Источник: сайт ИСО <https://www.iso.org/standard/80585.html>
<https://www.iso.org/obp/ui/#!iso:std:80585:en>



В ноябре 2021 года сайт Международной организации по стандартизации сообщил о том, что идёт голосование по проекту международного стандарта **ISO/IEC DIS 27005 «Информационная безопасность, кибербезопасность и защита неприкосновенности частной жизни – Руководство по менеджменту рисков и возможностей в области информационной безопасности»** (Information security, cybersecurity and privacy protection — Guidance on managing information security risks), см. <https://www.iso.org/standard/80585.html> и <https://www.iso.org/obp/ui#!/iso:std:80585:en>.

Документ, подготовленный техническим подкомитетом SC27 «Информационная безопасность, кибербезопасность и защита неприкосновенности частной жизни» (Information security, cybersecurity and privacy protection) Объединённого технического комитета Международной организации по стандартизации (ИСО) и Международной электротехнической комиссии (МЭК) JTC1 «Информационные технологии», заменил ныне действующую редакцию ISO/IEC 27005:2018.

Во вводной части документа отмечается:

В настоящем документе содержатся рекомендации по следующим вопросам:

- реализация требований в отношении рисков информационной безопасности, установленных в стандарте ISO/IEC 27001;
- ссылки на ключевые документы в семействе стандартов ISO/IEC 27000, поддерживающие деятельность по менеджменту рисков информационной безопасности;
- действия, направленные на обработку рисков, связанных с информационной безопасностью (см. ISO/IEC 27001:2013, разделы 6.1 и 8);
- внедрение рекомендаций по менеджменту риска из стандарта ISO 31000 в контексте информационной безопасности.

Данный документ содержит подробные рекомендации по менеджменту риска и дополняет рекомендации из стандарта ISO/IEC 27003.

Целевая аудитория данного документа включает:

- организации, намеренные создать и внедрить систему менеджмента информационной безопасности (СМИБ) в соответствии со стандартом ISO/IEC 27001;
- лица, осуществляющие менеджмент риска информационной безопасности или участвуют в нём (например, специалисты по СМИБ, владельцы рисков и другие заинтересованные стороны);

- організації, намереваючися удосконалити свої процеси менеджмента ризика інформаційної безпеки.

Зміст стандарту наступний:

Предисловие

Вступление

1. Область применения
2. Нормативные ссылки
3. Термины и определения
4. Структура настоящего документа
5. Менеджмент рисков информационной безопасности
6. Определение условий деловой деятельности (контекста)
7. Процесс оценки риска информационной безопасности.
8. Процесс обработки риска информационной безопасности.
9. Функционирование
10. Использование взаимосвязанных процессов системы менеджмента информационной безопасности (СМИБ)

Приложение А: Методы, поддерживающие процесс оценки риска –
Примеры

Библиография

СИСТЕМА ЗАПИСУ ФАЙЛІВ НА МІКРОФІЛЬМ ZEUTSCHEL ARCHIVE WRITER OP800/ OP 800 HR

Джерело інформації: <https://arinsy.com/product/systema-zapysu-fayliv-na-mikrofil-m-zeutschel-archive-writer-op800-op-800-hr/>

Висока продуктивність, автоматична система керування кольором ICC та великий вибір програмних засобів для пакетної корекції файлів дозволяють легко отримувати високоякісні мікрофільми.

Переваги зберігання інформації на мікрофільмах:

- довготривале зберігання;
- мікрофільм стійкий до електричних та магнітних полів;
- висока якість відтворення оригіналу;
- низька собівартість процесу зберігання.



Система запису файлів на мікрофільм (КОМ – система) Archive Writer OP800/ Archive Writer OP 800 HR створена для перетворення файлів на мікрофільм для системи використання або довготривалого зберігання.

Технічні характеристики

Виробник	Zeutschel GmbH
Роздільна здатність	132 млн пікселів – OP800, це 6815 dpi для 35 мм плівки 300 млн пікселів – OP800 HR це 10223 dpi для 35 мм плівки
Формати файлів	Всі стандартні формати в ч/б, напівтоновому чи кольоровому зображенні, наприклад TIFF, TIFF LZW, JPG, JP2, BMP, PDF, PDF A, JPM, GIF
Мікрофільм	рулонна ч/б, напівтонова чи кольорова мікроплівка 35 мм та 16 мм
Розмір кадру	стандартний 32 x 45 мм, максимальний 32,2 x 57 мм
Формат розташування кадрів	Повнокадровий, подвійний та багатокадровий

ЗМІСТ

Передмова	1
Национальные Архивы Великобритании: Использование искусственного интеллекта для отбора на архивное хранение электронных государственных документов	3
Искусственный интеллект: Кому можно доверять?	5
Управляемое машинное обучение: Искусственный интеллект без головной боли	7
Искусственный интеллект: При использовании предварительно обученных ИИ-алгоритмов в других организациях результаты могут отличаться	8
Как добиться поддержки генеральным и финансовым директорами Вашей программы управления документами?	9
Норвегия: Коварная проблема управления документами	11
Мари-Анн Шабен: Вытесняются и затмеваются ли документы данными?	13
США: «Аккредитованный комитет по стандартизации X9» опубликовал новый стандарт ANSI X9.141-2021 защиты персональных данных и уведомления об утечках	35
ИСО и МЭК: Готовится новая редакция стандарта ISO/IEC 27005 «Руководство по менеджменту рисков и возможностей в области информационной безопасности»	37
Система запису файлів на мікрофільм Zeutschel Archive Writer OP800/ OP 800 HR	39