

Для цитирования:

Бурый А.С., Лопатин И.Н., Погодин И.М. Стандартизация данных и процессов в открытой информационной среде. Часть 1. Открытые данные / III Научно-практическая конференция «Стандартизация: траектория науки», приуроченная ко Всемирному дню стандартов, Москва, 15 октября 2025 г. // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 6(87). С. 469–474.

УДК 004.91+004.031.42

СТАНДАРТИЗАЦИЯ ДАННЫХ И ПРОЦЕССОВ В ОТКРЫТОЙ ИНФОРМАЦИОННОЙ СРЕДЕ. ЧАСТЬ 1. ОТКРЫТЫЕ ДАННЫЕ

*Бурый А.С., д-р техн. наук, ФГБУ «Институт стандартизации», Россия, г. Москва,
Лопатин И.Н., аспирант ФГБУ «Институт стандартизации», Россия, г. Москва,
Погодин И.М., аспирант ФГБУ «Институт стандартизации», Россия, г. Москва*

Целью исследования является концептуализация открытой информационной среды на основе интеграции функциональной и технологической стандартизации в контексте инновационного развития общества в парадигме цифровой трансформации. Разнотипность, масштабность и необходимость использования цифровых данных является причиной разработки требований к представлению данных, к их доступности, полноте, достоверности, открытости и другим факторам, отражающим как предметную область исследования, так и источники получения данных в информационной среде.

Ключевые слова: информационная система, метатехнология, искусственный интеллект, открытые системы, открытые данные, интероперабельность, стандартизация.

DATA AND PROCESSES STANDARDIZATION IN AN OPEN INFORMATION ENVIRONMENT. PART 1. OPEN DATA

*Buryi A.S., Doctor of Science in Technology, Russian Standardization Institute, Moscow
Lopatin I.N., graduate student, Russian Standardization Institute, Moscow
Pogodin I.M., graduate student, Russian Standardization Institute, Moscow*

The aim of the research is to conceptualize an open information environment based on the integration of functional and technological standardization in the context of the innovative development of society in the paradigm of digital transformation. The diversity, scale, and necessity of using digital data is the reason for the development of requirements for data presentation, accessibility, completeness, reliability, openness, and other factors reflecting both the subject area of research and the sources of data in the information environment.

Keywords: information system, meta-technology, artificial intelligence, open system, open data, interoperability, standardization.

Введение

Сегодня технологии больше не являются просто платформами, они трансформируются в базис, на который опирается вся современная индустрия в широком смысле: вычислительная, коммуникационная, научная, образовательная и ряд других сфер деятельности (предметных областей, далее – ПрО), которые являются источниками разрабатываемых прорывных технологий¹, обеспечивающих технологический суверенитет государства.

Именно искусственный интеллект (ИИ) представляет собой метатехнологию, которая играет роль «универсального движка», проникая в различные отрасли производства. Однако при этом некоторым «побочным продуктом» выступает сложность интерпретации результатов, построения прогнозов, которые порой лежат за гранью сегодняшнего понимания и результатов, и возможностей интерпретации. Это требует определенной осторожности построения прогнозов, не соблюдение которой, делает их фактически недостоверными.

¹ Топ-10 цифровых технологий в России и мире. Сайт: <https://issek.hse.ru/news/1000447295.html> (дата обращения: 01.10.2025).

Технологии искусственного интеллекта – это метатехнологии NBICS-комплекса². Именно в конвергенции указанных научных направлений при формировании междисциплинарных исследований открывается новое перспективное исследовательское поле [1]. Все НБИКС-технологии являются сверхнаукоемкими электронными науками, потому что в них интенсивно используются компьютерные методы обработки большого множества данных, требуется распределённое сетевое окружение, применяются методы совместного поиска решений [2].

Активное продвижение открытых систем в настоящее время представляется как основная тенденция в области информационных технологий. Результативность развития этого направления обеспечивается как научно-технической политикой, так и реализацией стандартов открытых систем (ОС) [3, 4]. Понятие *открытая система* тесно связано с понятием *открытые данные*. Оба понятия, как и многие другие концепции «открытости» (открытый код, открытый доступ, свободные лицензии), имеют множество практических областей применения как в общественной жизни, в научном обществе – область открытой науки [4], так и в экономической сфере [5].

За последние годы открытые данные превратились в одно из ключевых направлений обеспечения открытости государства, прозрачности корпораций и доступности результатов научных исследований для возможности их дальнейшего использования. Для реализации указанных направлений особенно актуальна подготовка нового поколения инженеров-системотехников, исследователей [6].

По прогнозам, к 2027 году до 60% мировых данных будут использоваться для управления предприятиями и почти 20% данных будут относиться к категории критически важных с точки зрения поддержки систем обеспечения жизнедеятельности человека. При этом более 25% всех создаваемых данных будет генерироваться в режиме реального времени.

Развитие мобильной связи, появление технологий 5G, развитие интернета вещей, внедрение систем машинного обучения – все это будет влиять на рост объема генерируемых данных, причем количество устройств, генерирующих потоки данных, растет с каждым годом, включая интернет-сервисы³.

Огромный потенциал в развитии методов, моделей и алгоритмов в области всего жизненного цикла (ЖЦ) получаемых в информационных системах данных открывается за счет стандартизации процессов переработки данных в ходе их извлечения, представления, хранения, интеграции и других этапов ЖЦ. Это обеспечивается за счет требуемого уровня качества данных [7, 8], а также качества самих документов по стандартизации в формате метрик (полнота, безызыбочность, непротиворечивость, востребованность) [9].

Объединяющую технологию указанных направлений развития открытой информационной среды можно представить в виде цифровых платформ нового поколения, целью которых является поддержка функционирования открытых цифровых экосистем, «умных» сервисов, построенных как автономные интеллектуальные системы, интегрирующие сетевые, вычислительные, управляющие функции, обеспечивая всестороннюю поддержку в процессе выработки и принятия решений [10].

Целью настоящего исследования является концептуализация информационной среды на основе интеграции функциональной и технологической стандартизации с учетом инновационного развития общества в парадигме цифровой трансформации.

Открытые данные. Состояние и направления развития

Под *открытыми данными* обычно понимают данные, опубликованные в публичном доступе в сети Интернет в машиночитаемом формате и под свободной (открытой) лицензией,

² NBICS – (акроним: N – нано-; B – био-; I – инфо-; C – когнитивные; S – социально-гуманитарные технологии) – это взаимная интеграция или конвергенция, пяти фундаментальных отраслей знаний.

³ Большие данные как драйвер развития бизнеса и систем безопасности. UML: https://www.vedomosti.ru/press_releases/2025/11/17/bolshie-dannie-kak-draiver-razvitiya-biznesa-i-sistem-bezopasnosti (дата обращения: 19.11.2025).

допускающей бесплатное повторное использование этих данных, их переработку и распространение [5].

Российской особенностью является то, что основные данные, как по объему, так и по значению, накапливаются в государственных информационных системах (ГИС). В докладе «Оценка открытости государственных информационных систем 2020 г.»⁴ приведены примеры таких систем и оценена их текущая открытость. Из доклада следует, что текущее государственное регулирование не обеспечивает открытости данных в ряде ГИС по умолчанию. Таких требований нет ни в конкурсной документации на разработку информационных систем (ИС), ни в методических рекомендациях, ни в типовых положениях для ИС. В мировой практике важным направлением развития нормативной базы является внедрение принципа «открытость по умолчанию» (open by default). Текущее отечественное регулирование пока не использует этот принцип и, как следствие, в большинстве примеров ГИС недоступными остаются большая часть хранимых в них данных.

Открытые данные необходимы для разработки систем искусственного интеллекта⁵, для которых они являются «бесплатным топливом», без которого невозможно обучение алгоритмов на, по-настоящему, больших объемах данных и реальных задачах [5].

Направлениями развития в сфере открытых государственных данных являются:

1. Взаимодействие с потребителями данных в выработке общих требований к раскрытию и обмену данными.

2. Максимальное проактивное раскрытие информации, которой располагает государство:

а) для граждан, как основных потребителей и бенефициаров этой информации, (здесь речь идет прежде всего о разговоре с обществом на понятном языке, предоставление информации в понятном для граждан виде);

б) для экспертов и активистов от лица общества как для помощников и союзников власти в борьбе с неэффективностью государства в целом – от неэффективного расходования бюджета до коррупции, (здесь речь идет, в том числе, о предоставлении данных в машиночитаемом виде, чтобы была возможность использовать новейшие технологии для автоматического поиска проблемных точек и их устранения).

Общие свойства ОС обычно формируются следующим образом [3]:

- расширяемость/масштабируемость (extensibility/scalability);
- мобильность (переносимость) (portability);
- интероперабельность (interoperability);
- дружелюбность к пользователю, в том числе – легкая управляемость.

Данные часто поступают из многих источников, особенно это касается сложных динамических объектов, в контуре управления которых циркулирует разнотипная информация (управляющая, технологическая, регистрируемая и др.). Поэтому необходимо связать данные с созданными «связанными данными». *Связанные данные* – это структурированные данные, которые можно преобразовать в машиночитаемый формат, к которому можно выполнить, например, семантический запрос. Это позволяет выполнять поиск данных, а также комбинировать различные наборы данных для создания на их основе ценности. Создание ценности на основе данных требует объединения больших данных (см. рисунок), полученных из различных и разнородных источников данных. Большие открытые и связанные данные на государственном уровне (на рисунке обозначены цифрами 1–3) становятся наиболее востребованными как обычными гражданами, так и организациям в современных условиях развития цифровых технологий, что указывает на меняющийся характер данных [11].

Цепочка создания ценности данных включает основные этапы их ЖЦ: сбор, организацию, анализ, хранение и использование. Сбор связан с извлечением, записью и представлением данных в определенном формате для передачи или для предварительной обработки. Организация

⁴ Оценка открытости государственных информационных систем 2020. UML: <https://ach.gov.ru/upload/pdf/>

⁵ Искусственный интеллект и открытые данные: важнейшее сочетание. <https://data.europa.eu/en/publications/datastories/ai-and-open-data-crucial-combination>

предполагает структурирование базы данных, включая обобщение определенных элементов данных и добавление заголовков и пояснительных примечаний. Это превращает данные в информацию. Анализ связан с интеграцией и обработкой различных типов данных, включая интеллектуальный анализ данных. Он преобразует информацию в знания. Хранение предполагает архивирование данных в доступных для извлечения формах. Использование предполагает применение знаний, основанных на данных, для прогнозирования и принятия решений в соответствующих организационно-технических системах.



Место открытых данных в информационной среде:

- 1 – открытые правительственные данные; 2 – связанные открытые правительственные данные;
 3 – большие открытые правительственные данные; 4 – большие открытые данные;
 5 – связанные открытые данные; 6 – связанные правительственные данные;
 7 – большие правительственные данные

Стандартизация данных и среды

Стандартизация открытой информационной среды – это процесс создания и применения общих правил, форматов и протоколов для обеспечения совместимости, интеграции и взаимодействия различных ИС, технологий и данных, что минимизирует затраты, повышает гибкость, прозрачность, долговечность и способствует инновациям в ИТ-инфраструктуре, используя принципы открытых систем. Это достигается за счет разработки национальных (ГОСТ Р) и международных (ISO) стандартов (например, серия ИСО 10303 для обмена данными, стандарты информационной безопасности), обеспечивая бесшовный информационный обмен и унификацию компонентов (форматы и структуры данных, программных модулей).

Ключевые аспекты стандартизации открытой информационной среды должны обеспечивать:

- в методическом аспекте – методологию открытых систем, за счет модульной архитектуры построения систем, позволяющей легко заменять их компоненты и интегрировать с другими системами [12];
- в технологическом аспекте – совместимость и интеграцию для поддержки взаимодействия разнородных систем и ресурсов через общие стандарты обмена данными (пример – CALS-технологии);
- в нормативном аспекте – разработку и повсеместное внедрение открытых стандартов для поддержки вышеперечисленных аспектов, отдавая на откуп производителям лишь качество предоставляемых информационных продуктов.

Перспективность методологии открытых ИС основывается, прежде всего, на сочетании решаемых прикладных задач и возможностей платформенных компонент [12], объединяющих операционный уровень (операционной системы), аппаратный, отвечающий за функциональное наполнение, и сервисный уровень ИС (системно-прикладной, баз данных и знаний и др.).

Отмеченное выше требование *интероперабельности*, предполагает не только взаимосвязь между различными системами, но и включает в себя разработку общих стандартов и процессов на различных уровнях для облегчения передачи информации, сетевые протоколы, контент данных, использующие стандарты на метаданные и перспективные Веб-сервисы, с целью автоматизированного распространения системных процессов и процедур. Для полностью открытых интерфейсов ИС интероперабельность – это способность продукта или системы, взаимодействовать и функционировать с другими продуктами или системами без каких-либо ограничений доступа.

Для упрощения предварительной обработки (на начальных этапах ЖЦ данных), разработки программного обеспечения и анализа данных крайне важно уменьшать требования к адаптациям, необходимым для обеспечения интероперабельности. Идеалом интероперабельности для исследователей является структура, которая позволяет легко использовать наборы данных различными программными продуктами независимо от их источника.

Именно стандарты определяют характеристики продукта, процесса или услуги. Используя существующие архитектуры аппаратно-программных и технологических структур, можно реализовать определенную стандартную структуру для создания полезного свойства или продукта в целом. Общая структура может помочь обеспечить совместимость между различными программными продуктами и сократить время, необходимое для повторения или реализации дополнительной предварительной обработки данных. Совместимость должна гарантировать такие показатели, как согласованность, когерентность, надежность и другие метрики понятия *качества данных* [7].

В настоящее время уже сложились основные принципы и стандарты в области публикации открытых данных. Они инвариантны для разных стран и могут успешно применяться, не требуя содержательной переработки, так как в целом отвечают и общечеловеческим ценностям, и современному состоянию технологий. К ним относятся следующие принципы и стандарты:

- принципы Хартии открытых данных (opendatacharter.net/principles);
- стандарт публикации открытых данных: Frictionless Data – это одновременно и стандарт публикации табличных открытых данных, и широкий набор открытых инструментов по подготовке и упаковке данных⁶;
- открытые стандарты публикации данных в Project Open Data – project-open-data.cio.gov/open-standards
- стандарты данных от The ODI (Open Data Institute) – это руководство и набор рекомендаций для создания, разработки и внедрения открытых стандартов данных.

Заключение

Одна из ключевых мотиваций для публикации открытых данных состоит в том, что предоставление данных общественности, в конечном счете, ведет к приумножению общественных благ. Иными словами, данными могут пользоваться граждане для создания новых бизнесов, развития систем общественного контроля, повышения удобства в городской среде, повышения оперативности и качества в предоставлении госуслуг, формирования фактически «умной» городской среды. Стандартизация – это важная часть открытости, позволяющая администраторам баз данных правильно раскрывать данные (так, чтобы их можно было использовать), а пользователям – легко в них разбираться и обрабатывать их средствами, ориентированными на соответствующий стандарт. Кроме того, это полезно, с точки зрения оценки качества данных.

На основе идеи междисциплинарной направленности развития современной науки необходимо строить и процесс обучения – подготовку соответствующих специалистов для

⁶ Open source for Opendata: Методические рекомендации по использованию решений с открытым кодом при работе с открытыми данными. UML: <https://www.infoculture.ru/wp-content/uploads/2021/06/OpenSourceForOpenData-1.pdf>

новых дисциплин и направлений исследований: астроинформатики, информатики поведения, биоинформатики, информатики здравоохранения, медицинской социальной информатики, основным материалом которых выступают данные, полученные в ходе нашей повседневной деятельности, а также в ходе проводимых экспериментов. Задачами такой подготовки выступает необходимость научить студентов «думать с помощью данных», «управлять данными», «проводить аналитику данных», «взаимодействовать с данными», «интерпретировать полученные результаты».

В дальнейшем исследовании предполагается рассмотреть роль стандартизации и открытых данных в развитии возможностей современных цифровых платформ для задач ситуативного управления и поддержки принятия решений.

Список литературы / References

1. Ковальчук М.В., Нарайкин О.С., Яцишина Е.Б. Конвергенция наук и технологий – новый этап научно-технического развития // Вопросы философии. 2013. № 3. С. 3–11. / Kovalchuk M.V., Narajkin O.S., Yatsishina E.B. Konvergenciya nauk i tekhnologij – novyj etap nauchno-tehnicheskogo razvitiya. Voprosy filosofii. 2013; 3: 3–11. (In Russ.).
2. Родзин С.И., Титаренко И.Н. NBIC-технологии, искусственный интеллект и электронная культура // Информатика, вычислительная техника и инженерное образование. 2013. № 2(13). С. 60–74. / Rodzin S.I., Titarenko I.N. NBIC-tehnologii, iskusstvennyj intellekt i elektronnaya kul'tura. Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie. 2013; 2(13):60–74. (In Russ.).
3. Кузнецов С. Открытые системы, процессы стандартизации и профили стандартов. https://citforum.ru/database/articles/art_19.shtml (дата обращения: 01.10.2025). / Kuznetsov S. Otkrytye sistemy, processy standartizacii i profili standartov. (In Russ.). Available at: https://citforum.ru/database/articles/art_19.shtml
4. Бурый А.С., Ловцов Д.А. Открытая наука как движущая сила информационного общества // Правовая информатика. 2024. № 3. С. 4–12. – DOI 10.24682/1994-1404-2024-3-4-12. / Buryi A.S., Lovtsov D.A. Otkrytaya nauka kak dvizhushchaya sila informacionnogo obshchestva. Legal informatics. 2024;3:4-12. (In Russ.).
5. Работа с открытыми данными: особенности публикации и использования в российском правовом поле. Аналитический доклад. URL: <https://fnkcrr-database.ru/OpenDataReview.pdf> (дата обращения: 01.10.2025). / Rabota s otkrytymi dannymi: osobennosti publikacii i ispol'zovaniya v rossijskom pravovom pole. Analiticheskij doklad. Available at: <https://fnkcrr-database.ru/OpenDataReview.pdf> (accessed 12.10.2025). (In Russ.).
6. Цао Л. Образ мышления в науке о данных: Наступающая научно-техническая и экономическая революция. – СПб.: Издательство Европейского университета в Санкт-Петербурге, 2022. – 552 с. / Cao L. Obraz myshleniya v nauke o dannyh: Nastupayushchaya nauchno-tehnicheskaya i ekonomicheskaya revolyuciya. Sankt-Peterburg: Izdatel'stvo Evropejskogo universiteta v Sankt-Peterburge, 2022, 552 p. (In Russ.).
7. Бурый А.С., Погодин И.М. Оценка качества больших данных. Часть 1. Основные понятия и метрики // Информационно-экономические аспекты стандартизации и технического регулирования. 2024. № 3 (78). С. 49–58. / Buryi A.S., Pogodin I.M. Ocenka kachestva bol'shih dannyh. Part 1. Osnovnye ponyatiya i metriki. Information and economic aspects of standardization and technical regulation. 2024;3(78):49–58. (In Russ.).
8. Лопатин И.Н. Многоуровневые системы качественных данных на основе моделей искусственного интеллекта: проблемы и решения // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 1(82). С. 70–75. / Lopatin I.N. Mnogourovnevye sistemy kachestvennyh dannyh na osnove modelej iskusstvennogo intellekta: problemy i resheniya. Information and economic aspects of standardization and technical regulation. 2025;1(82):70–75. (In Russ.).
9. Гарбук С.В. Метод оценки влияния параметров стандартизации на эффективность создания и применения систем искусственного интеллекта // Информационно-экономические аспекты стандартизации и технического регулирования. 2022. № 3(67). С. 4–14. / Garbuk S.V. Metod ocenki vliyaniya parametrov standartizacii na effektivnost' sozdaniya i primeneniya sistem iskusstvennogo intellekta. Information and economic aspects of standardization and technical regulation. 2024;3(67):4–14. (In Russ.).
10. Городецкий В.И., Ларюхин В.Б., Скобелев П.О. Концептуальная модель цифровой платформы для кибер-физического управления современным предприятием. Часть 1. Цифровая платформа и цифровая экосистема // Мехатроника, автоматизация, управление. 2019. Т. 20, № 6. С. 323–332. – DOI 10.17587/mau.20.323-332. / Gorodetsky V.I., Laryukhin V.B., Skobelev P.O. Conceptual Model of a Digital Platform for Cyber-Physical Management of a Modern Enterprises Part 1. Digital Platform and Digital Ecosystem. Mekhatronika, Avtomatizatsiya, Upravlenie. 2019;20(6):323-332. (In Russ.). <https://doi.org/10.17587/mau.20.323-332>
11. Charalabidis Y., et al. The world of open data // Public Administration and Information Technology. Cham: Springer International Publishing. 2018, vol. 28, 245 p. Available at: <https://doi.org/10.1007/978-3-319-90850-2> (Accessed 12.10.2025).
12. Лукинова О.В. Методологические аспекты управления жизненным циклом информационной системы на основе инструментов функциональной стандартизации // Программные продукты и системы. 2016. № 4. С. 27–35. / Lukinova O.V. Metodologicheskie aspekty upravleniya zhiznennym ciklom informacionnoj sistemy na osnove instrumentov funkcional'noj standartizacii. Programmnye produkty i sistemy. 2016;4:27-35. (In Russ.).