

АРХИТЕКТУРА ПЛАТФОРМЫ ОТКРЫТЫХ ДАННЫХ РОССИЙСКОЙ ЭЛЕКТРОЭНЕРГЕТИКИ

Ф.С. Непша

ООО «РТСофт – Смарт Грид»
Россия, 105264, Москва, Верхняя Первомайская ул., 51
E-mail: nepsha_fs@rtsoft.ru

В.А. Воронин

Кузбасский государственный технический университет им. Т.Ф. Горбачева
Россия, 650000, Кемерово, Весенняя ул., 28
E-mail: voroninva@kuzstu.ru

С.П. Ковалёв

Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65
E-mail: kovalyov@sibnet.ru

Ключевые слова: энергетика, открытые данные, цифровая платформа, инженерия данных, экосистема.

Аннотация: В докладе рассматриваются основные проблемы, связанные с накоплением и применением открытых данных электроэнергетики, в том числе в условиях развития технологий искусственного интеллекта, и возможные пути их решения на основе платформенного подхода. Приведен краткий обзор ряда публикаций по вопросам организации открытых данных в энергетике. Перечислены источники открытых данных и проблемы, связанные с извлечением данных из разнородных источников. Рассмотрен экосистемный подход к организации платформы открытых данных. Представлены ключевые заинтересованные стороны платформы. Идентифицированы основные сценарии использования платформы. Предложена архитектура платформы, позволяющая удовлетворить потребности заинтересованных сторон и обеспечить ускоренное развитие инноваций в энергетике. Для эффективного хранения данных в платформе предложено использовать озеро-склад данных (data lakehouse).

1. Введение

Открытыми данными называются общественно значимые материалы, публикуемые в машиночитаемом формате и пригодные к использованию для создания полезных для граждан и бизнеса информационных сервисов, в том числе на базе искусственного интеллекта. Такие данные должны обладать следующими свойствами: первичность, полнота, актуальность, пригодность к машинной обработке, недискриминационный доступ, отсутствие частных закрытых форматов, лицензионная чистота [1].

В ряде публикаций отмечается высокая полезность открытых данных для раскрытия потенциала электроэнергетики в условиях энергетического перехода. В работе [2] предложена структура платформы открытых данных для реализации концепции электронного правительства. Авторы работы [3] отмечают, что открытые данные необходимы для повышения качества научных исследований в

электроэнергетике. В исследовании корейских ученых [4] предлагается структура платформы, позволяющей сформировать граф знаний для описания семантических отношений между собираемыми данными, что позволяет повысить доступность и эффективность использования данных. В статье [5] исследователи отмечают, что экосистема обмена открытыми данными в энергетике является ключом к развитию новых сервисов и технологий. В [6] представлены результаты анализа источников данных в европейской энергетике и условий предоставления данных. Китайские ученые [7] приводят анализ источников открытых данных по возобновляемым источникам энергии (ВИЭ) в Китае. Ученые отмечают, что имеющиеся данные имеют низкое качество и недостаточны для обеспечения устойчивого развития энергосистемы. В связи с этим актуально создание платформ открытых данных, предоставляющих удобный формат взаимодействия источников данных с их пользователями.

Существует ряд проектов, нацеленных на устранение барьеров к распространению и использованию открытых данных электроэнергетики. В США для этого создаётся Национальная инфраструктура искусственного интеллекта для электрических сетей (NI4AI). Цели национальной инфраструктуры охватывают три основных аспекта: развертывание платформы для работы с временными рядами, предоставление наборов данных разработчикам для тестирования новых алгоритмов и формирование экосистемы профессионалов. Подобные проекты представлены и в других государствах.

В Российской Федерации накопление и применение открытых данных приобрело организованную форму с принятием Постановления Правительства РФ от 10 июля 2013 г. № 583 «Об обеспечении доступа к общедоступной информации о деятельности государственных органов и органов местного самоуправления на их официальных сайтах в информационно-телекоммуникационной сети "Интернет" в форме открытых данных». Постановление определяет порядок отнесения информации к открытым данным, периодичность обновления открытых данных и правила их размещения на сайтах государственных органов. Каждый набор данных снабжается паспортом – минимальным набором идентифицирующих метаданных, в том числе наименование, обладатель, гиперссылка в сети Интернет и формат. Для консолидации ведомственных наборов открытых данных введен в эксплуатацию федеральный портал data.gov.ru и региональные порталы, такие как московский data.mos.ru.

Тем не менее, в Российской Федерации до сих пор уделено недостаточное внимание вопросам сбора и применения открытых данных электроэнергетики, особенно в условиях быстрого развития технологий и средств искусственного интеллекта. Актуальна разработка цифровых платформ, направленных на максимальную универсализацию доступа к открытым данным и автоматизацию их жизненного цикла [8]. Предложения по созданию такой платформы представлены в настоящем докладе.

2. Открытые данные в российской электроэнергетике

Основными источниками данных в российской электроэнергетике являются системный оператор (СО ЕЭС), администратор торговой системы (АТС), Министерство энергетики, сетевые и сбытовые компании и др. Однако открытые данные об объектах электроэнергетики представлены ими не совсем адекватно, в том числе:

- сайты большинства субъектов энергетики не имеют удобного прикладного программного интерфейса для запроса данных;

- отсутствует ряд полезных данных в открытом доступе (данные о состоянии силового оборудования, аварийности, топологии электрической сети, генерации ВИЭ и др.);
- отсутствуют четкие регламенты использования данных с сайтов;
- отсутствуют единые требования к размещению данных по тарифам на сайтах энергосбытовых компаний;
- многие данные могут быть извлечены только путем синтаксического разбора наполнения сайтов, что значительно усложняет процесс доступа к данным и не гарантирует устойчивость их получения.

Вышеуказанные недостатки указывают на необходимость формирования единого подхода к распространению открытых данных, что возможно сделать путем создания соответствующей платформы и сопутствующей экосистемы. Это позволит системно организовать сбор и консолидацию данных с разных источников, их предобработку и предоставление в пользование заинтересованным сторонам.

3. Экосистема платформы открытых данных

Ключевым фактором обеспечения эффективности информационного пространства электроэнергетики является экосистема заинтересованных сторон создаваемой цифровой платформы. Их состав формируется из заинтересованных сторон энергетического перехода в целом [9] и включает следующие группы субъектов:

- поставщики данных (такие как энергетические компании, правительственные учреждения и производители), выступающие в качестве источников данных для платформы и постановщиков задач анализа данных;
- агрегаторы, собирающие данные в хорошо организованном и безопасном пространстве, обеспечивая качество данных и систематическую категоризацию;
- подготовители данных, выполняющие предварительный анализ данных и настройку конвейеров для очистки и нормализации данных по мере необходимости;
- разработчики программного обеспечения, заинтересованные в доступе к данным для тестирования своих решений и разработки алгоритмов, требующих больших наборов данных;
- университеты, проводящие исследования с использованием реальных данных от поставщиков;
- инжиниринговые компании, предоставляющие услуги по обработке данных для извлечения информации, позволяющей повысить эффективность бизнес-процессов в электроэнергетике;
- органы власти, ответственные за правовое регулирование открытых данных и реализацию мер поддержки компаний, создающих инновационные алгоритмы.

Среди возможных сценариев использования платформы открытых данных участниками экосистемы необходимо выделить следующие [10]:

- подключение источника открытых данных (в форме либо массива данных, либо внешней системы-поставщика данных);
- подключение приложения-потребителя открытых данных (в том числе на основе искусственного интеллекта);
- формирование задания на исследование данных;
- формирование запроса на получение необходимых данных для исследования;
- организация курсов переподготовки и повышения квалификации специалистов по сбору и обработке открытых данных;
- обработка данных о функционировании энергетического оборудования.

4. Архитектура платформы открытых данных

Архитектура предлагаемой платформы открытых данных представлена на рис. 1.

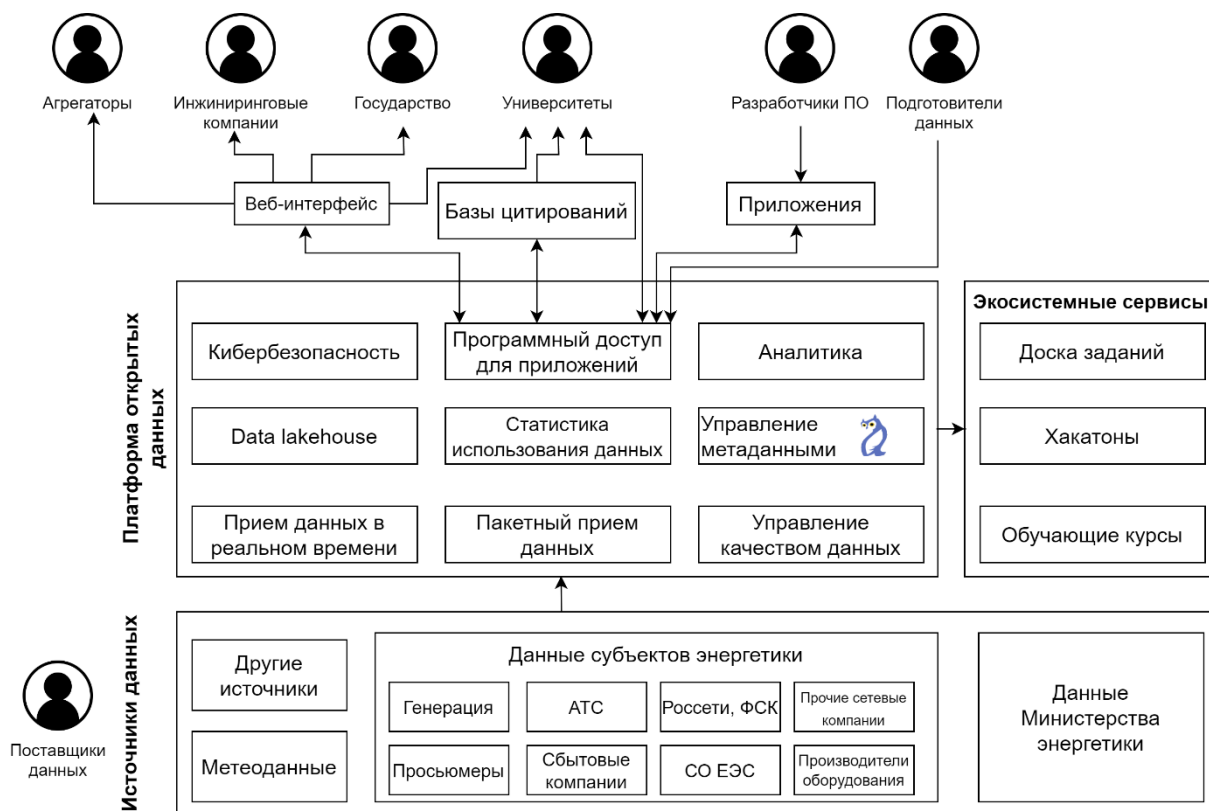


Рис. 1. Архитектура платформы открытых данных электроэнергетики.

Предлагаемая архитектура имеет следующие отличительные особенности [10]:

- 1) Для эффективного хранения данных в платформе используется озеро-склад данных (data lakehouse) – инновационное архитектурное решение, сочетающее преимущества озера данных (data lake), в том числе низкая стоимость хранения и многообразие способов доступа, и склада данных (data warehouse), такие как глубокое структурирование данных и широкий спектр функций управления [11].
- 2) Платформа включает систему управления метаданными с использованием онтологического подхода [12]. Использование онтологии позволяет не только формировать информационные модели данных для приложений, но и структурировать теоретические данные, организовывать курсы подготовки специалистов по обработке открытых данных.
- 3) Платформа позволяет работать как с историческими данными, так и с данными, получаемыми в реальном времени. Это позволяет оперативно формировать тестовые полигоны для проверки алгоритмов, требующих данных в реальном времени. Например, это может потребоваться для отработки интеллектуальных алгоритмов прогнозирования солнечной генерации.
- 4) Платформа имеет возможность размещения заданий по обработке данных, что позволит субъектам энергетики быстрее внедрять инновационные решения на основе искусственного интеллекта.

- 5) Платформа способна служить основой для организации площадки проведения хакатонов и создает среду для подбора необходимых специалистов.
- 6) Платформа позволяет привязывать данные к публикациям, что повышает воспроизводимость исследований и обеспечивает контроль качества публикаций.
- 7) Облегченные копии платформы могут стать полигонами для подготовки ИТ-разработчиков для электроэнергетики.
- 8) Предприятия энергетического сектора могут использовать платформу в качестве хаба для безопасного размещения информации для исследований.

5. Заключение

В данной работе рассмотрены проблемы создания платформы открытых данных в электроэнергетике. Приведены ссылки на примеры подобных платформ и проанализированы возможности доступа к открытым данным российской электроэнергетики. С учетом опыта создания платформ, описаны заинтересованные стороны, их потребности и интересы в накоплении и использовании открытых данных.

В качестве решения рассмотренных проблем предлагается разработать платформу открытых данных, которая призвана обеспечить удобный защищенный доступ к данным и поддерживать решение научных и практических задач с их использованием, в том числе на базе искусственного интеллекта. В докладе представлена архитектура платформы, которая способна лечь в основу технического задания на создание платформы.

Список литературы

1. Открытые данные. http://government.ru/open_data/ (дата обращения 10.12.2023).
2. Alhawawsha M., Panchenko T. Open Data Platform Architecture and Its Advantages for an Open E-Government // *Advances in Computer Science for Engineering and Education III. ICCSEEA 2020. Advances in Intelligent Systems and Computing*. Vol. 1247. Cham: Springer, 2021.
3. Pfenninger S., DeCarolis J., Hirth L., Quoilin S., Staffell I. The Importance of Open Data and Software: Is Energy Research Lagging Behind? // *Energy Policy*. 2017. Vol. 101. P. 211-215.
4. Lee J., Park J. An Approach to Constructing a Knowledge Graph Based on Korean Open-Government Data // *Applied Sciences*. 2019. Vol. 9, No. 19. P. 4095.
5. Manfred M., Nastasi B., Groppi D., Garcia D.A. Open Data and Energy Analytics - An Analysis of Essential Information for Energy System Planning, Design and Operation // *Energy*. 2020. Vol. 213. 118803.
6. Hirth L. Open Data for Electricity Modeling: Legal Aspects // *Energy Strategy Reviews*. 2020. Vol. 27. 100433.
7. Zhang X., Patino-Echeverri D., Li M., Wu L. A Review of Publicly Available Data Sources for Models to Study Renewables Integration in China's Power System // *Renewable and Sustainable Energy Reviews*. 2022. Vol. 159. 112215.
8. Староверов В.В. Открытые данные // *International Journal of Open Information Technologies*. 2016. Т. 4, № 10. С. 62-68.
9. Ковалёв С.П. Заинтересованные стороны цифровой децентрализованной малоуглеродной энергетики // *Энергобезопасность и энергосбережение*. 2019. № 2. С. 49-55.
10. Nepsha F.S., Voronin V.A., Kovalyov S.P. Open Data Platform as a Source of ML-driven Innovations in the Energy Sector // *Proceedings of the 2023 Belarusian-Ural-Siberian Smart Energy Conference (BUSSEC)*. Ekaterinburg, Russian Federation: IEEE, 2023. P. 1-6.
11. Oreščanin D., Hlupić T. Data Lakehouse - a Novel Step in Analytics Architecture // *Proceedings of the 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*. Opatija, Croatia: IEEE, 2021. P. 1242-1246.
12. Nepsha F.S., Nebera A.A., Andrievsky A.A., Krasilnikov M.I. Development of an Ontology for Smart Distributed Energy Systems // *IFAC-PapersOnLine*. 2022. Vol. 55, No. 9. P. 454-459.