

ОБ ЭКОНОМИКЕ ДАННЫХ

Введение. Экономика данных – это очередной (третий) этап процесса цифровой трансформации, который должен решить проблему управления данными государства и бизнеса как целостным объектом. Переход от того, что называется «цифровая экономика», к тому что называется «экономика данных» становится политикой многих государств [1]. В такой экономике жизненные ситуации становятся агрегированными потребностями граждан, клиентские пути представляют собой комплекс процессов для удовлетворения этих потребностей, а домены тесно связаны с клиентскими путями и охватывают области деятельности государственных органов. Данные становятся активами, которые требуют оценки их стоимости. В настоящее время стоимость таких активов не определяется, отсутствуют методики такой оценки. Особенностью данных, как финансового актива является то, что несмотря на его высокую стоимость, определяемую затратным методом, его ценность может быть сведена к нулю в силу низкого качества данных. В докладе рассматривается этот аспект экономики данных и предлагаются пути его совершенствования.

Основная часть. Прежде всего укажем факторы, определяющие стоимость данных при оценке их затратным методом. Такая стоимость оценивается совокупностью множества факторов. Среди них: цена аппаратного оборудования центра обработки данных (ЦОД); стоимость программного обеспечения ЦОД; операционные расходы по эксплуатации и обслуживанию ЦОД; расходы на сбор, аналитику, предоставление данных; расходы на масштабируемость, на обеспечение безопасности и т.п. В цифровой экономике такие затраты распределяются на множество информационных систем в различных доменах «Государственное управление», «Финансы», «Строительство и ЖКХ» и др. В результате стоимость данных, определяемая совокупностью указанных факторов по множеству информационных систем, приобретает существенное значение. Теперь обратим внимание на следующие три обстоятельства.

1. Затраты значительны, но каково качество конечного продукта, то есть данных? Если оно низкое, или недостаточно высокое, то их использование в бизнес-процессах или ограничено, или рискованно, или даже недопустимо. То есть, перечисленные выше затраты на продукт бессмысленны. Необходимо иметь метрику качества, системы мониторинга и аналитики качества по такой метрике, декларировать

собственниками информационных ресурсов уровень их качества для целей использования данных, нести ответственность за предоставление данных по качеству ниже декларированных.

2. Для текущего состояния цифровизации характерно развитие технологий искусственного интеллекта (ИИ). В ИИ-системах в форме ИИ-агентов, ИИ-ассистентов, имеет место этап машинного обучения (ML-MachineLearning) для которого используются данные, в том числе, государственных информационных ресурсов. Обучение по данным низкого качества ведет к низкому качеству ИИ-систем. То есть, затраты на ML могут быть бессмысленны, если данные рожают риски использования ИИ.

3. Экономика данных строится в значительной степени путем использования технологий больших данных BigData с созданием корпоративных хранилищ данных КХД [2]. Создание КХЛ осуществляется по множеству источников данных. Проблема в том, что эти источники содержат данные разных форматов, разной семантики, несовместимые между собой. То есть, на создание КХДи аналитику данных необходимы новые затраты, поднимающие стоимость данных.

Указанные обстоятельства обозначили три проблемы: 1) качество данных, в том числе, государственных информационных ресурсов не известно, не документируется и не нормируется; 2) отсутствуют единые правила управления качеством данных для аналитики и обучения ИИ, аналитики больших данных (BigDataAnalytics, BDA) в организациях; 3) сложность, а соответственно большая затратность формирования КХД из-за отсутствия в стране интероперабельности множества информационных систем, количество которых исчисляется сотнями.

Решение первой из указанных проблем автор предлагает осуществить путем принятия на национальном уровне единой модели качества данных. Модель должна содержать характеристики, которые подлежат использованию при оценке качества данных – продукта информационных систем. Качество такого продукта – это степень, с которой данные удовлетворяют требованиям, определенные собственником/поставщиком данных. Эти требования отражаются в модели качества данных через ее характеристики. Примеры таких характеристик:

– характеристика полноты, которая может касаться как атрибутов баз данных, так и отсутствия части данных (например, отсутствие данных об объектах недвижимости реального мира в Едином регистре недвижимого имущества, прав на него и сделок с ним);

– характеристика правильности, то есть, соответствия данных реальности. Очевидно, что неправильные данные нельзя распространять пользователям;

– характеристика объективности, то есть уверенности, что данные не содержат субъективных оценок;

– характеристика точности измерения и фиксации данных. В зависимости от процесса и целей анализа показатели необходимо фиксировать с точностью до дня, часа, минуты или секунды. Либо вес товара измерять с точностью до тонны или грамма. Либо точность измерения границ земельного участка должна быть не хуже заданного значения;

– характеристика своевременности, которая характеризует время после сбора данных, по прошествии которого они становятся доступными клиентам. Время должно соответствовать скорости жизненного цикла соответствующего процесса. Корректные, но устаревшие, то есть уже не актуальные данные, опасны для принятия решений;

– характеристика целостности. Целостность предполагает, что данные не описывают ситуацию, которая невозможна;

– характеристика согласованности, которая указывает на соответствие одних данных другим. Отсутствие согласованности – свидетельство ошибок в данных;

– характеристика когерентности, то есть согласованности с другими источниками данных.

и другие.

Модель должна содержать правила определения значения каждой характеристики, то есть, элемента показателя качества. Именно метрика качества информирует потребителей о характеристиках качества и позволяет оценивать степень пригодности данных к использованию. Параметры качества должны быть измеримыми, ожидаемые результаты иметь количественное определение. Параметры должны иметь значимость для потребителя данных каждого информационного ресурса. Метрика должна интересовать потребителей данных, а за несоответствие данных декларированным показателям метрики ее качества должен отвечать собственник соответствующих информационных систем. Выход значения измеряемого параметра за пределы установленного должен инициировать процедуру улучшения данных. Для достижения указанных целей рекомендуется принять стандарт с обязательным применением «СТБ. Информационные технологии. Качество данных» и «СТБ. Информационные технологии. Пространственные данные. Качество данных». Такими стандартами могут стать мо-

модифицированные международные стандарты «ISO/TS 8000.Data quality», «ISO/TS 19157. SpatialData. Dataquality».

Решение второй из указанных проблем предлагается начать принятием на национальном уровне концепции качества данных для аналитики и обучения ИИ, аналитики больших данных (BigDataAnalytics, BDA) в организациях. Концепция должна содержать руководство по процессу управления качеством данных для машинного обучения с учителем, без учителя, с частичным привлечением учителя, структуру жизненного цикла данных для ИИ, требования и рекомендации по управлению качеством данных с учетом их жизненного цикла в цепочках поставок, обязанности органа, создающего ИИ.

Для достижения указанной цели рекомендуется принять стандарт с обязательным использованием «СТБ. Искусственный интеллект. Качество данных для аналитики и машинного обучения». Его основой могут стать модифицированные стандарты «ISO/IEC 5259. Artificialintelligence for analytics and machine learning (ML)– Data quality»; «ISO/IEC 24668.Artificialintelligence – Processmanagementframeworkforbigdataanalytics».

Решение третьей из указанных проблем обеспечивается предстоящей передачей в эксплуатацию автоматизированной информационной системы национальной платформы интероперабельности и правил ее постепенного развития (жизненных циклов). Такая платформа создана в 2025 году и с 2026 года начинаются ее жизненные циклы. Для установления правил управления такими жизненными циклами рекомендуется принять стандарты «СТБ. Информационные технологии. Интероперабельность. Основные положения» и «СТБ. Информационные технологии. Управление жизненными циклами платформы интероперабельности».

Предлагаемые рекомендации сформированы в рамках исполнения НИР «Выработка научно обоснованных подходов по повышению качества государственных данных», выполненного по заказу Министерства связи и информатизации Республики Беларусь.

ЛИТЕРАТУРА

1. Минцифры. Национальный проект «Экономика данных и цифровая трансформация государства» [Электронный ресурс]. URL: <https://digital.gov.ru/target/naczionalnyj-proekt-ekonomika-dannyh-icifrovaya-transformacziya-gosudarstva>(дата обращения: 9.01.2026).

2. Большие данные (BigData)[Электронный ресурс]. URL: <https://www.tadviser.ru/index.php>(дата обращения: 9.01.2026).