

Структура базы данных для оценки долговременной безопасности хранения радиоактивных отходов

А. М. Скоробогатов, С. В. Панченко, О. Н. Апанасюк, Т. А. Буланцева

Федеральное государственное бюджетное учреждение науки Институт проблем безопасного развития атомной энергетики РАН, Москва, Россия

Аннотация. В работе рассмотрена структура базы данных, независимой от предметно-ориентированной области знаний. При разработке архитектуры базы данных учитывался опыт эксплуатации региональной базы радиэкологических данных. Описана структура баз данных, оптимизированная под специфику долговременного хранения данных, которая базируется на сериях национальных стандартов в области информационных технологий (Метаданные) и государственной системы обеспечения единства измерений. База данных с заданной структурой может сохранять широкий класс данных и при необходимости неограниченно? как предметно, так и физически расширена до возможностей Big data без изменения структуры в долговременной перспективе. Пределом является размер памяти физических носителей. Стандартизация форматов таблиц данных и соединительных таблиц взаимодействия с метаданными позволяет упростить и повысить надежность процедур конверсии и миграции информации.

Ключевые слова: структура базы данных, долговременное хранение данных, метаданные.

DOI 10.14357/20718632220206

Введение

Отличительной особенностью информации о состоянии пунктов хранения радиоактивных отходов (далее – ПХРО) и районов их размещения является ее разнообразие, выходящее за пределы какой-либо определенной предметной области знаний, и необходимость сохранения детальных знаний на всех этапах жизненного цикла ПХРО, который может многократно превышать продолжительность жизни человека.

Действующие нормативные документы Ростехнадзора по безопасности при использовании атомной энергии (РБ-003-21 [1] и др.) определяют требования к оценке долговременной безопасности ПХРО с использованием зна-

чительного объема исходной информации различного характера. Это:

- характеристика площадки ПХРО и района его размещения;
- информация о населении, проживающем в районе размещения ПХРО;
- характеристика окружающей среды;
- сведения об инженерной части ПХРО;
- данные о вмещающих и (или) несущих породах ПХРО;
- сведения о радиоактивных отходах (далее – РАО) и упаковках РАО.

Исходную информацию необходимо сохранять на длительном промежутке времени и иметь возможность управлять ею в целях про-

ведения оценок безопасности ПХРО и прогноза изменения ситуации в горизонте событий сотни и тысячи лет с использованием различных вычислительных моделей.

Очевидным решением является создание комплекса предметно-ориентированных баз данных (ПОБД). Примером подобных решений является база данных (БД) по полигону твердых бытовых отходов «Каскад» (Воронежская область), в которой каждое тематическое направление представлено своей, уникальной по набору атрибутов, таблицей [2]. Предметная ориентация информации влечет за собой собственный специфический набор описания сведений и, соответственно, уникальную структуру БД. Как правило, каждая ПОБД имеет собственную модель данных, уникальный поименованный набор объектов (таблиц), структуру таблиц и формат описателей сведений. В последующем, как показывает практика, автоматизация процедур извлечения и обработки информации в ПОБД сопровождается разработкой специального программного обеспечения (СПО), ориентированного на конкретную структуру базы данных.

Обеспечение долговременной сохранности информации (определяя термин по ГОСТ Р 54989-2012/ISO/TR 18492:2005¹) обуславливает потребность своевременно осуществлять конвертацию и миграцию сведений. С каждым последующим этапом жизненного цикла ПХРО степень детализации сведений об объекте, состоянии барьеров безопасности и окружающей среды района размещения ПХРО нарастает в зависимости от предметно-ориентированной информации. Образование массивов информации происходит неравномерно и практически независимо по каждому из компонентов комплекса ПОБД. Некоторые массивы сведений, локализованные по району размещения ПХРО (например: гидрометеорологическая информация), могут перейти в стадию больших данных (big data) с соответствующими технологиями масштабирования, обработки и анализа в различные моменты времени. Из этого следует нарастающая сложность условий конвертации

и миграции различно структурированных и разнообразных форматов компонентов комплекса ПОБД, которая может повлечь за собой потерю аутентичности информации и нарушит требование воспроизводимости сведений.

Проблемы, например, многоформатного представления информации Государственного фонда данных о состоянии окружающей природной среды (Госфонда) [3] и ее несвоевременной миграции описаны в работе [4]. Значимость решения возникших проблем подтверждается фактом того, что развитие Госфонда стало важной частью «Стратегии деятельности в области гидрометеорологии и смежных с ней областях на период до 2030 г. (с учетом аспектов изменения климата)», утвержденной Правительством России [5]. Таким образом, возникает необходимость нахождения таких решений, которые, по возможности, должны исключить зависимость структуры БД от предметной ориентации.

1. Структура базы данных

Долговременная сохранность информации зависит от следующих составляющих [6–8]:

- *независимости* данных (от способа хранения и обработки данных);
- *интерпретируемости* (читаемости) данных (независимость от формата данных);
- сохранения *аутентичности* (неизменности) данных в процессе хранения (например, при смене эксплуатируемой информационной системы);
- *надежности* хранения цифровых данных (увязана с невозможностью существования цифровых данных вне цифровых программно-технических средств);
- сохранения *семантики* хранимых данных (сохранения понимания, что означают данные);
- *устойчивости программно-технических средств* хранения к внешним воздействиям.

Для решения проблемы долговременной сохранности информации требуется разработка структуры БД, инвариантной к предметной области знаний и оптимизация объема сохраняемой информации. В соответствии с законодательством, информация – это сведения (сообщения, данные) независимые от формы их представления. Согласно стандартам по ин-

¹ ГОСТ Р 54989-2012/ISO/TR 18492:2005. Обеспечение долговременной сохранности электронных документов. М.: Стандартинформ, 2013. 18 с.

формационным технологиям, данные – это многократно интерпретируемое представление информации, пригодное для передачи и обработки формализованным образом (ГОСТ Р ИСО/МЭК 11179-1—2010², ГОСТ 33707-2016 (ISO/IEC 2382:2015)³, ГОСТ Р ИСО 8000-2—2019⁴). Уточняя определение понятия «Данные», сформулируем: Данные – это формализованный результат наблюдений, измерений, вычислений или иных сведений, документированный и находящийся на физическом носителе, при извлечении и обработки которых используются программно-технические способы.

Альтернативным способом структурной организации БД при инвариантности от областей знаний в сфере экологии и охраны окружающей среды явилось предложение [9], в котором за сущность БД принято «значение размера величины» (R_v), характеризующее в пределе 14-мерным пространством метаданных – величины, времени, источника информации, исследуемой среды, фактора, местоположения (4 параметра), организации, проводящей исследования, и т. д.

База региональных радиэкологических данных (БРРД) [10] и БД системы экологического обеспечения эксплуатации космодрома «Восточный» [11] используют подобную структуру. В результате эксплуатации БРРД были выявлены недостатки из-за неполноты 14-мерного пространства метаданных, требующих увеличения размерности пространства метаданных для описания R_v . Запись R_v в БД, первоначально являющая вектором (набором указателей) в линейном 14-мерном пространстве метаданных, может неограниченно удлиняться от вынужденного увеличения размерности пространства метаданных (добавления новых описателей данных). Нестандартность записи R_v является угрозой нарушения семантики данных. Поэтому, основной задачей является обеспечение стандартизации записи R_v на длительную временную перспективу.

² ГОСТ Р ИСО/МЭК 11179-1-2010. Информационная технология. Регистры метаданных (РМД). Часть 1. Основные положения). М.: Стандартинформ, 2012, 24 с.

³ ГОСТ 33707-2016 (ISO/IEC 2382:2015). Информационные технологии. Словарь. М.: Стандартинформ, 2017, 206 с.

⁴ ГОСТ Р ИСО 8000-2-2019. Качество данных. Часть 2. Словарь. М.: Стандартинформ, 2019, 16 с.

Дальнейшее развитие альтернативного способа структурной организации БД осуществлялось с учетом применения требований серии национальных стандартов (ГОСТ Р ИСО/МЭК 11179 «Информационные технологии. Регистры метаданных (РМД)». Часть 1. Основные положения). М.: Стандартинформ, 2012. 24 с.) и использование идеи Grid-технологии для компьютерной инфраструктуры [12], основанной на создании стандартизованных служб безопасного доступа к данным и вычислительным ресурсам.

Введем понятие Grid-пространства метаданных (аналог служб безопасного доступа к данным и вычислительным ресурсам): Grid-пространство метаданных – это независимый набор множеств метаданных, объединенных по признаку целевого назначения характеристик описания данных. Зададим Grid-пространства метаданных:

- базовые (необходимые) определители данных – Grid-пространство B ;
- местоположение данных (место проведения наблюдений, измерений) – Grid-пространство W ;
- характеристика наблюдений или измерений (Grid-пространство H);
- условия проведения наблюдений или измерений (Grid-пространство C).

Grid-пространства не пересекаются. Адресация в каждом из Grid-пространств определяется соответствующими множествами $\{B_i\}$, $\{W_j\}$, $\{H_k\}$ и $\{C_l\}$, где $i, j, k, l \geq 1$ и может $i \neq j \neq k \neq l$. Размерность B -пространства постоянно $i = \text{const}$, а размеры пространств j, k и l ограничены физической памятью носителей.

Семантика R_v определяется совокупностью элементов множеств $\{B_i\}$, $\{W_j\}$, $\{H_k\}$ и $\{C_l\}$ (Рис. 1).

Оптимизация и стандартизация структуры записи данных R_v требует детализации категорий данных. Определим, что данные, исключая сетевые/графовые и потоковые данные [13], могут представляться в следующих категориях:

Q_Res – результат наблюдений/условий измерений, описанный в виде неструктурированной или слабоструктурированной документированной информации (документ - document-based information), подразумевает типы данных

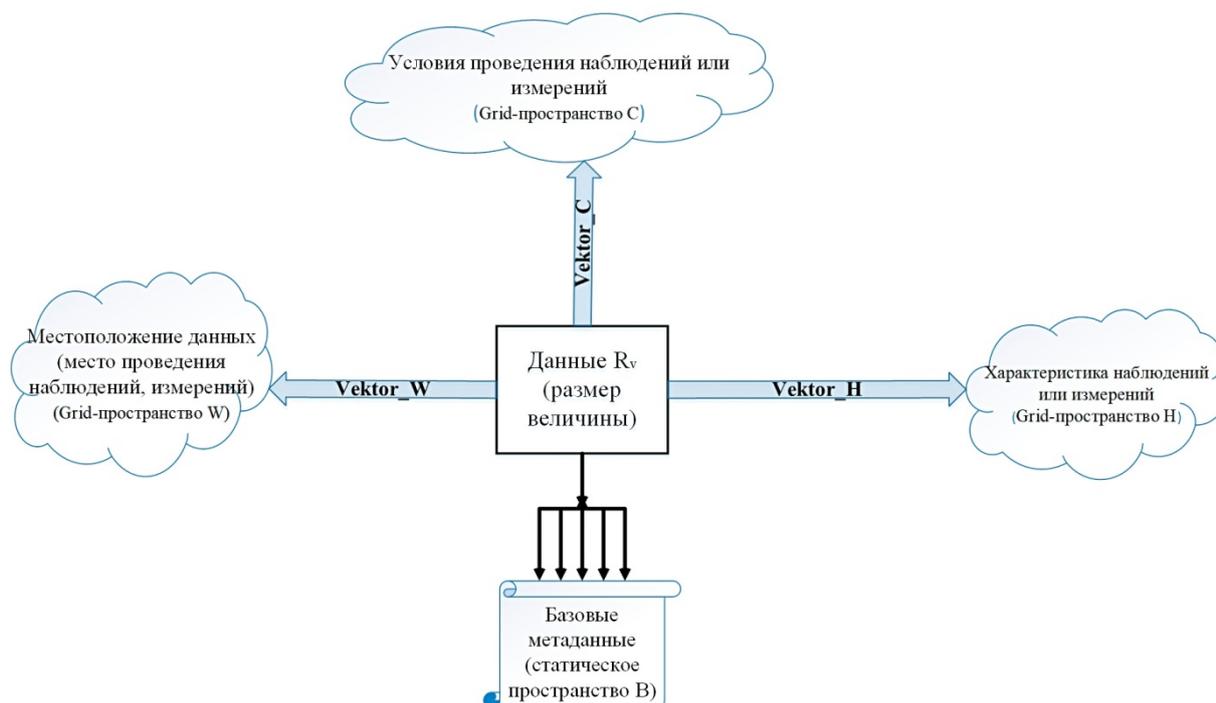


Рис. 1. Образ БД

- текст, указатель (адрес) файлов или IP-адрес электронного архива;

N_Res – числовые значения (оценок) результатов наблюдений, измерений и вычислений;

U_Res – числовые значения оценок неопределенности, интервала охвата или диапазона результатов наблюдений, измерений и вычислений;

C_Res – числовые значения условий, при которых проводились наблюдения, измерения и вычисления;

SP_Res – числовые значения, утвержденные/принятые уполномоченными органами как справочные или нормативные, или регулирующие уровни величин, или применяемые в вычислительных моделях параметры (коэффициенты).

Под понятием «наблюдения» в ряде случаев понимаются, например, сведения, представленные статистическими службами.

Выделение в отдельную категорию данных **U_Res** (неопределенности) является оптимизацией и определяется тем фактом, что значительная часть результатов наблюдений, измерений и вычислений **N_Res** публикуется без указания оценок неопределенностей измерений/наблюдений/вычислений.

Категория данных **C_Res** (числовые значения условий, при которых проводились наблюдения, измерения и вычисления) является расширением аналитических возможностей БД и дает возможность систематизировать **N_Res** в зависимости от тех или иных условий.

Вынужденное добавление категории данных **SP_Res** обусловлено встречающимися в публикациях сведениями, выраженными в относительных единицах при сравнении с нормативными или лимитирующими уровнями, например, в единицах предельно допустимых концентраций или единицах допустимых выбросов/сбросов и т. д. Также **SP_Res** предназначено для сохранения значений параметров, используемых в различных вычислительных моделях, в целях обеспечения воспроизводимости вычислений.

Реализация усовершенствованного альтернативного подхода к структуре БД приведена на схеме базы данных (Рис. 2). В-пространство представлено стационарными (существующими все время жизни БД) непересекающимися множествами в виде таблиц **SP_FACTOR**, **SP_UNITS**, **TYPE_DATA**, **KLASS_DATA**, **TIMEDATE** и **ISX**, которые, соответственно,

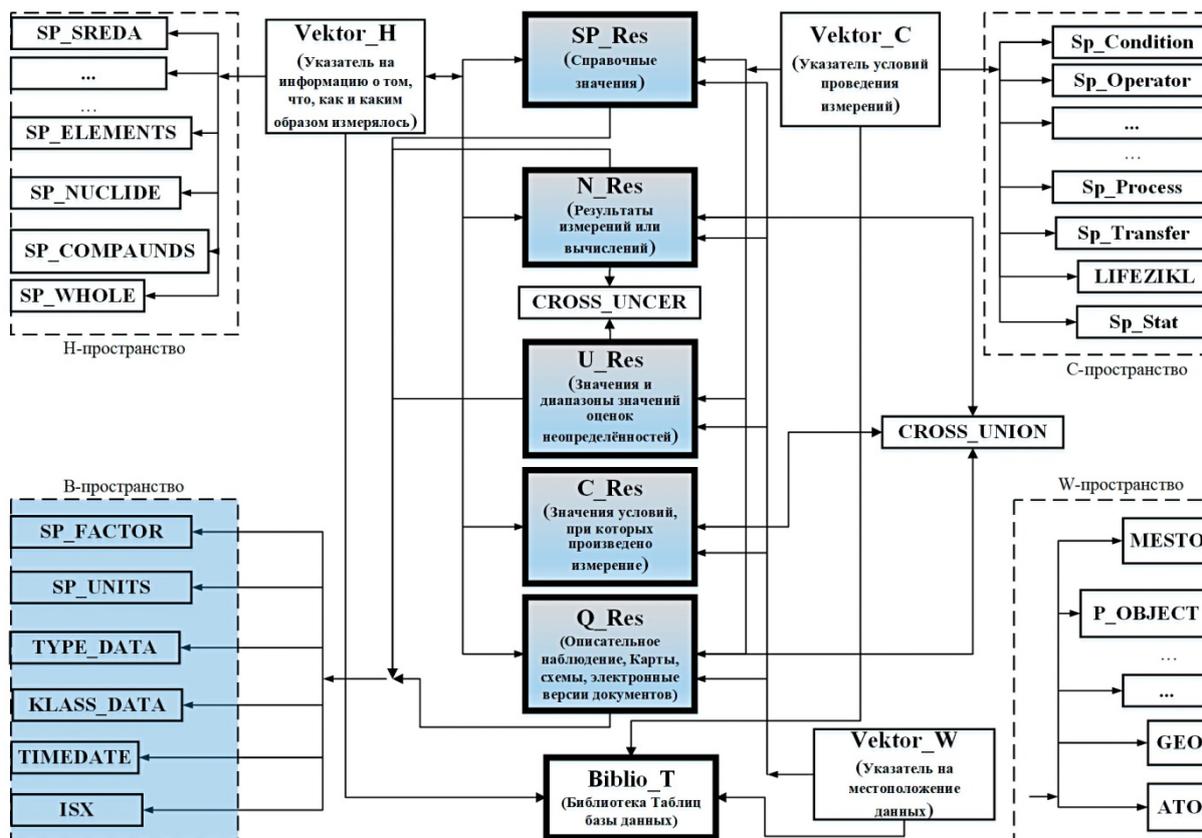


Рис. 2. Схема инвариантной предметно-ориентированной БД

являются справочниками величин, единиц измерения, типов данных, класса данных (групп), времени (моменты, интервалы времени) и источников информации.

Справочник типов данных TYPE_DATA содержит понятия: эксперимент лабораторный, эксперимент в натурных условиях (in-situ), статистические наблюдения, мониторинг, контроль, расчет (теоретический), расчет (по фактическим данным – интерполирование, экстраполирование), аудио-, фото-, видеонаблюдение, 3-D обследование и т. д.

Справочник классификации данных KCLASS_DATA содержит следующие классы:

- данные, характеризующие опасный или потенциально опасный объект;
- данные, характеризующие инженерные системы (барьеры) безопасности опасного или потенциально опасного объекта;
- данные, характеризующие конструкционные материалы инженерных систем (барьеров) безопасности опасного или потенциально опасного объекта;

– данные, характеризующие геологию и гидрогеологию опасного или потенциально опасного объекта;

– данные, характеризующие внешние условия вблизи опасного или потенциально опасного объекта, включая климат и метеорологию;

– результаты мониторинга окружающей среды и здоровья человека;

– данные, оценивающие уровень воздействия на окружающую среду и человека в результате функционирования опасного или потенциально опасного объекта.

Справочник времени TIMEDATE содержит время, которое может быть либо моментом времени, непрерывным интервалом или кусочно-дробными отрезками времени, а также геологическим временем (геохронология), временем экспозиции и т. д.

W-пространство (Рис.2) поддерживается таблицами-справочниками, определяющими местоположение относительно которого получены значения тех или иных величин. В настоящее время W-пространство представлено

таблицами АТО, GEO, P_ОБЪЕКТ и MESTO, которые являются, соответственно, таблицами-справочниками административно-территориальных образований (с точностью до населенного пункта), наименований географических объектов, промышленных объектов, места проведения измерений (точки). W-пространство неограниченно расширяемо за счет включения, по мере необходимости, новых таблиц-справочников, например, водных объектов, лесов, сельхозугодий, природных зон, кадастровых участков и т. д. В таблице MESTO может сохраняться как точка измерений с трехмерными координатами (широта, долгота и высота/глубина) так и как некое поименованное пространство, например, «Санитарно-защитная зона», «Зона наблюдения», к которым привязаны сведения, содержащиеся в отчетах организаций по безопасности.

Н-пространство (Рис. 2) наполняется таблицами-справочниками, которые должны дать ответы на вопрос, в какой среде что измерялось и каким образом. На данный момент Н-пространство содержит следующие таблицы-справочники: SP_SREDA, SP_ELEMENT, SP_NUCLIDE, SP_COMPAUNDS и SP_WHOLE, в которых, соответственно, содержатся объекты наблюдений (воздух, вода, почва, выброс, сброс и т. д.), химические элементы, радионуклиды и их группы, химические вещества/соединения или иные поименованные группы соединений (пыль, зола, сажа и т. д.), компоненты объекта исследований.

Поясним на примере как это работает. Предположим, при проведении исследований изучались вопросы вертикальной миграции радионуклида ^{137}Cs в почве, по результатам которых получены удельные активности радионуклида по слоям почвы с шагом 5 см на глубину до 1 метра. При записи в БД значений величин (удельная активность) будут выбраны: указатели на запись «почва» - в таблице-справочнике SP_SREDA; ^{137}Cs - в таблице-справочнике SP_NUCLIDE; почвенный слой, например, 0–5 см или 10–15 см - в таблице SP_WHOLE. Аналогично и в воде, тогда при описании значения величин (концентрации) будут задействованы: указатели на запись «вода» - в таблице-справочнике SP_SREDA; за-

пись конкретного вещества (например, карбонат кальция CaCO_3) - в таблице-справочнике SP_COMPAUNDS; в таблице SP_WHOLE - «фильтрат» или «осадок», в зависимости от значения на запись.

Н-пространство неограниченно расширяемо за счет включения, по мере потребности, новых таблиц-справочников, например, типов почв, методик выполнения измерений, справочников материалов и тканей, типов грунта, строительных материалов, биологических объектов (растений, животных), пород древесной продукции, продукции сельского хозяйства, конструктивных элементов и т. д.

С-пространство (Рис. 2) содержит сведения об условиях получения значений величин и состоит из таблиц-справочников Sp_Condition, Sp_Operator, Sp_Process, Sp_Transfer, LIFEZIKL и Sp_stat, которые включают в себя следующие параметры:

- состояния/состав (газ; жидкость; твердое тело; молекулярная форма; водная форма; водная форма, связанная в организме; органически связанная форма; сумма водной и органически связанной форм и т. д.);
- организации, выполняющие или представляющие результаты наблюдений;
- условия (внешнее облучение, внутреннее облучение, суммарное облучение, аэробные условия, анаэробные условия, открытая местность, помещение и т. д.);
- виды излучений (фотонное, альфа-излучение, бета-излучение и т. д.);
- стадии жизненного цикла (проектирование, эксплуатация, вывод из эксплуатации, завершение эксплуатации, возраст или возрастные группы, биологические стадии роста/развития и т. д.);
- условия статистического представления значений величин (среднее, среднеарифметическое, среднегеометрическое, медиана и т. д.).

С-пространство неограниченно расширяемо за счет включения, по мере потребности, новых таблиц-справочников, например, профессий или профессиональных групп, типов лесорастительных условий, категорий лесов, категорий земель сельскохозяйственного назначения, климатических зон и т. д.

Выбор точек в многомерных C-, H-, W-пространствах метаданных осуществляется посредством соединительных таблиц Vektor_C, Vektor_H и Vektor_W (Рис. 2 и Табл. 1), имеющих типовой (по порядку, наименованию, типу и размеру) набор полей-указателей.

Соединительные таблицы Vektor_C, Vektor_H и Vektor_W поддерживаются библиотекой таблиц Biblio_T, в которой хранятся наименования таблиц БД, описание их назначения, указатель (путь имен) актуального местонахождения таблиц на физических носителях памяти и, соответственно, присвоенный им уникальный идентификатор.

В соединительных таблицах Vektor_C, Vektor_H и Vektor_W в целях обеспечения долговременной аутентичности значений идентификаторов Data_id, Basetable_id, Space_id и Spacetable_id дублируются в виде символьного образа идентификатора в соответствующих полях CData_id, CBasetable_id, CSpace_id и CSpacetable_id. Поля Arxiv и Kod_language служат для поддержки операций миграции и конверсии БД.

Идентификатор данных Data_id выделяет запись значения величины, находящуюся в таблице данных, определяемую указателем Basetable_id на библиотеку таблиц Biblio_T. Аналогично, идентификатор метаданных Space_id выделяет запись метаданных, находящуюся в

таблице метаданных, определяемую указателем Spacetable_id на библиотеку таблиц Biblio_T.

Состав полей таблиц данных Q_Res, N_Res, C_Res и SP_Res (Рис. 2) имеет типовую (по наименованиям, типу и размеру) структуру (Табл. 2).

В случае, если значение размера величины (результат наблюдения, измерений и т. д.) CRes в опубликованном виде превышает 2048 символа или это не символьные сведения, то тогда опубликованные значения размещают отдельным самостоятельным документом в удобном для чтения формате на физическом носителе памяти, а в CRes размещают указатель на созданный документ.

Содержимое полей таблицы данных в целях обеспечения долговременной аутентичности информации дублируется чередованием числовых значений и их символьных образов. Поля Arxiv и Kod_language служат для поддержки операций миграции и конверсии БД.

Если результаты измерений/наблюдений являются совместными (в терминах РМГ 29-2013 ГСОЕИ. Метрология. Основные термины и определения [14]), то тогда задействуется соединительная таблица CROSS_UNION (Рис. 2 и Табл. 3), имеющая аналогичный представленному в Табл. 1 порядок полей, но отличающийся по их наименованию.

Табл. 1. Состав полей-указателей соединительных таблиц Vektor_C, Vektor_H и Vektor_W

| № поля п/п | Наименование поля | Тип поля | Размер, минимальная единица памяти | Описание |
|------------|-------------------|------------|------------------------------------|--|
| 1 | Data_id | Числовой | 16 | Идентификатор данных |
| 2 | CData_id | Символьный | 20 | Идентификатор данных в таблицах данных, выраженный в символьном виде (эквивалент поля 1) |
| 3 | Basetable_id | Числовой | 4 | Идентификатор таблиц данных |
| 4 | CBasetable_id | Символьный | 10 | Идентификатор таблицы данных, выраженный в символьном виде (эквивалент поля 3) |
| 5 | Space_id | Числовой | 16 | Идентификатор метаданных |
| 6 | CSpace_id | Символьный | 20 | Идентификатор метаданных в таблицах метаданных, выраженный в символьном виде (эквивалент поля 5) |
| 7 | Spacetable_id | Числовой | 4 | Идентификатор таблиц метаданных |
| 8 | CSpacetable_id | Символьный | 10 | Идентификатор таблицы метаданных, выраженный в символьном виде (эквивалент поля 7) |
| 9 | Arxiv | Символьный | 254 | Код/указатель архива в символьном виде |
| 10 | Kod_language | Символьный | 10 | Кодировочный код символов (например, ASCII или UNICODE и т. д.) |

Табл. 2. Структура полей данных Q_Res, N_Res, C_Res и SP_Res

| Наименование поля | Тип поля | Размер, минимальная единица памяти | Описание |
|-------------------|------------|------------------------------------|---|
| Res_id | Числовой | 16 | Идентификатор данных |
| CRes_id | Символьный | 128 | Идентификатор данных, выраженный в символьном виде (эквивалент поля Res_id) |
| Ist_id | Числовой | 4 | Идентификатор источника данных (указатель на запись в таблице ISX) |
| CIst_id | Символьный | 10 | Идентификатор источника данных, выраженный в символьном виде (эквивалент поля Ist_id) |
| Factor_id | Числовой | 4 | Идентификатор величины (указатель на запись в таблице SP_FACTOR) |
| CFactor_id | Символьный | 10 | Идентификатор величины, выраженный в символьном виде (эквивалент поля Factor_id) |
| Unit_id | Числовой | 4 | Идентификатор единицы измерения (указатель на запись в таблице SP_UNITS) |
| CUnit_id | Символьный | 10 | Идентификатор единицы измерения, выраженный в символьном виде (эквивалент поля Unit_id) |
| Time_id | Числовой | 4 | Идентификатор времени, к которому относятся данные (указатель на запись в таблице TIMEDATE) |
| CTime_id | Символьный | 10 | Идентификатор времени, выраженный в символьном виде (эквивалент поля Time_id) |
| Tdata_id | Числовой | 4 | Идентификатор типа данных (указатель на запись в таблице TYPE_DATA) |
| CTdata_id | Символьный | 10 | Идентификатор типа данных, выраженный в символьном виде (эквивалент поля Tdata_id) |
| Kdata_id | Числовой | 4 | Идентификатор класса данных (указатель на запись в таблице KCLASS_DATA) |
| CKdata_id | Символьный | 10 | Идентификатор класса данных, выраженный в символьном виде (эквивалент поля Kdata_id) |
| Res | Числовой | 16 | Числовое значение размера величины (Числовой эквивалент поля CRes) |
| CRes | Символьный | 2048 | Значение размера величины, выраженное в символьном виде аутентично опубликованным сведениям |
| Arxiv | Символьный | 254 | Код/указатель архива в символьном виде |
| Kod_language | Символьный | 10 | Кодировочный код символов или файла (например, ASCII, UNICODE, PDF/A и т. д.) |

При регистрации совместных измерений/наблюдений (например, в пробе измеряют удельную активность ^{137}Cs , ^{90}Sr , ^{40}K и $^{239+240}\text{Pu}$) за базовое значение принимают первое по порядку значение записи в БД, то есть заполняют поля соединительной таблицы CROSS_UNION – Data_id и Basetable_id., а все последующие значения размещают в полях Cordata_id и Cortable_id соединительной таблицы CROSS_UNION.

Особое место в таблицах данных занимает таблица U_Res, в которой сохраняются результаты оценок неопределенностей измерений как в

оценках отклонений от среднего (Тип А) или оценок, основанных на определенных ранее знаниях (Тип В), так и в виде минимакс диапазона значений или интервала охвата с заданным уровнем доверия. Вследствие данного факта, структура полей таблицы U_Res несколько отличается от таблиц данных Q_Res, N_Res, C_Res и SP_Res. Таблица U_Res взаимодействует с таблицей N_Res посредством соединительной таблицы CROSS_UNCER (Рис. 2 и Табл. 4). Подключение соединительной таблицы CROSS_UNCER обусловлено тем, что не во всех случаях числовые значения измерений/наблюдений имеют оценки

Табл. 3. Состав полей соединительной таблицы CROSS_UNION

| № поля п/п | Наименование поля | Тип поля | Размер, минимальная единица памяти | Описание |
|------------|-------------------|------------|------------------------------------|--|
| 1 | Data_id | Числовой | 16 | Идентификатор данных |
| 2 | CData_id | Символьный | 20 | Идентификатор данных в таблицах данных, выраженный в символьном виде (эквивалент поля 1) |
| 3 | Basetable_id | Числовой | 4 | Идентификатор таблиц данных |
| 4 | CBasetable_id | Символьный | 10 | Идентификатор таблицы данных, выраженный в символьном виде (эквивалент поля 3) |
| 5 | Cordata_id | Числовой | 16 | Идентификатор данных |
| 6 | CCordata_id | Символьный | 20 | Идентификатор данных в таблицах данных, выраженный в символьном виде (эквивалент поля 5) |
| 7 | Cortable_id | Числовой | 4 | Идентификатор таблиц данных |
| 8 | CCortable_id | Символьный | 10 | Идентификатор таблицы данных, выраженный в символьном виде (эквивалент поля 7) |
| 9 | Arxiv | Символьный | 254 | Код/указатель архива в символьном виде |
| 10 | Kod_language | Символьный | 10 | Кодировочный код символов (например, ASCII или UNICODE и т. д.) |

Табл. 4. Состав полей соединительной таблицы CROSS_UNCER

| № поля п/п | Наименование поля | Тип поля | Размер, минимальная единица памяти | Описание |
|------------|-------------------|------------|------------------------------------|--|
| 1 | Data_id | Числовой | 16 | Идентификатор данных, указатель на таблицу N_Res |
| 2 | CData_id | Символьный | 20 | Идентификатор данных в таблице N_Res, выраженный в символьном виде (эквивалент поля 1) |
| 3 | Uncdata_id | Числовой | 16 | Идентификатор неопределенности данных, указатель на таблицу U_Res |
| 4 | CUncdata_id | Символьный | 20 | Идентификатор неопределенности данных в таблицах U_Res, выраженный в символьном виде (эквивалент поля 3) |
| 5 | Arxiv | Символьный | 10 | Код архива в символьном виде |
| 6 | Kod_language | Символьный | 10 | Кодировочный код символов (например: ASCII или UNICODE и т. д.) |

неопределенности измерений. Примером могут служить сведения, приводимые в формах федерального или отраслевого статистических наблюдений.

В настоящее время БД реализована на основе СУБД Oracle 11g и по состоянию на 01.01.2022 г. в таблице N_RES содержится 510081 записей (значений). Запрос на выгрузку, например, сведений по району размещения Белоярской АЭС (32111 данных) выполняется за 0,224 с.

2. Требования к обеспечению долговременной сохранности данных

Поля Arxiv, Kod_language должны быть во всех таблицах БД.

В процессе всего жизненного цикла БД записи, относящиеся к метаданным (В-, С-, Н- и W-пространства метаданных), не подлежат удалению (принцип сохранения целостности и семантики данных).

Справочники базы данных (метаданные) должны содержать, в случае их предметной идентичности, идентификаторы или кодификаторы, классификаторы официальных отечественных/международных регистров или реестров. Например, справочник веществ SP_COMPAUNDS имеет дополнительные поля (атрибуты), содержащие актуальные коды CAS, RTECS, EC и регистрационный номер РПОХВ, обеспечивающие связь с Российским регистром потенциально опасных химических и биологических веществ/международными регистрами веществ.

Процессы разбиения на части больших данных (разбиение на страницы или тома), конверсии и миграции выполняются в следующих случаях:

- по истечении определенного регламентного времени;
- по факту достижения заданного объема записей в какой-либо таблице БД;
- изменения производительности БД (снижение времени доступа к данным при типовом запросе) на 10 и более процентов;
- изменение требований к кодировочным таблицам символического представления информации;
- появление новых носителей физической памяти.

Актуальная версия таблиц БД в поле Arxiv имеет значение "пусто" или NULL. Поле Arxiv заполняется обозначениями с определенным порядком.

Процедуры конверсии и миграции выполняются в следующем порядке: сначала актуальные версии таблиц БД (значения поля Arxiv – "пусто" или NULL) и проверкой работоспособности, а затем последующие по порядку убывания времени создания.

Конверсию данных осуществляют последовательно в начале метаданные, затем соединительные таблицы и напоследок - таблицы данных. Приоритет имеют меньшие по объему таблицы.

Заключение

В статье описана разработанная структура БД, способная охватить широкий класс документированной информации, в том числе не-

структурированной или слабоструктурированной при использовании поля C_RES как ссылочного указателя на документ (Табл.2). БД содержит таблицы, имеющие стандартизованный состав полей.

Следствием наличия в БД таблиц со стандартизованным составом полей является отсутствие необходимости применения СПО из-за универсальности процедур доступа к данным. СПО применяют на этапе анализа данных при выгрузке из систем хранения данных.

В каждой таблице БД содержатся символические поля, поддерживающие режим аутентичности представляемых данных.

БД может быть неограниченно масштабирована и расширена за счет включения новых массивов различных категорий метаданных без внесения изменений в структуру БД. Ограничением является объем памяти физического носителя.

Литература

1. Руководство по безопасности при использовании атомной энергии Оценка долговременной безопасности пунктов глубинного захоронения радиоактивных отходов. РБ-003-21: утв. пост. Ростехнадзором 19 марта 2021 г. № 101, введ. с 19 марта 2021 г. М.: ФБУ «НТЦ ЯРБ», 2021.
2. Корабельников Н. А., Зинюков Ю. М. Структура базы данных для объектов экологического мониторинга – полигонов твердых бытовых отходов // Вестник Воронежского государственного университета. Серия: гео-логия. 2019. С. 118–124.
3. Единый Государственный Фонд Данных. [Электронный ресурс] URL: <http://meteo.ru/egfd?ysclid=11ufpb8ut1> (дата обращения: 11.04.2022).
4. Шаймарданов В. М. Создание программно-аппаратного комплекса долговременного хранения и использования гидрометеорологической информации // Ученые записки Российского государственного гидрометеорологического университета. 2011. № 21. С. 118–125.
5. Распоряжение Правительства РФ от 3 сентября 2010 г. N 1458-р. Стратегии деятельности в области гидрометеорологии и смежных с ней областях на период до 2030 года (с учетом аспектов изменения климата). URL: <http://static.government.ru/media/files/AAOMfY6lO8AaNaZQVTTLUrYCTRrhNva.pdf> (дата обращения: 11.04.2022).
6. Соловьев А. В., Баканова Н. Б., Проблемы долговременной сохранности больших данных // Информационные технологии и вычислительные системы. 2019. № 2. С. 44–53. DOI: 10.14357/20718632190205.
7. Соловьев А. В. Алгоритмическое решение проблемы сохранения аутентичности цифровых данных // Информа-

- ционные технологии и вычислительные системы. 2021. № 1. С. 12–19. DOI: 10.14357/20718632210102.
8. Соловьев А. В. Решение проблемы интерпретации цифровых данных долговременного хранения // Труды ИСА РАН. 2021. Т. 71. № 2. С. 43–49. DOI: 10.14357/20790279210206.
 9. Скоробогатов А. М., Корсак М. Н., Кроленко М. И. Об организации инвариантной базы данных в сфере экологии и охраны окружающей среды // В сборнике: Наука, образование, производство в решении экологических проблем (Экология-2020). Матер. XVI Межд. науч.-техн. конф., в 2-х томах, посвящается 75-летию Победы в Великой Отечественной войне. Уфа: УГАТУ. 2020. Т.1. С. 155–159.
 10. Свидетельство о государственной регистрации базы данных № 2018621930. Региональная база радиэкологических данных для количественной оценки негативных эффектов и рисков от радиационного и химических факторов техногенного воздействия на население и окружающую среду Свердловской области. Версия 2018.01. ИБРАЭ РАН (RU). Дата регистрации в Реестре баз данных 03 декабря 2018 г.
 11. Самброс В. В., Ловцкая О. В., Петрова М. В., Савеленок А. Н. Структура базы данных системы экологического обеспечения эксплуатации космодрома «Восточный» // Материалы II Всероссийской научно-практической конференции «Космодром «Восточный» – будущее космической отрасли России», Благовещенский государственный педагогический университет. 2013. С. 102–108.
 12. Кореньков В. В., Нечаевский А. В., Трофимов В. В. Моделирование распределенной системы сбора, передачи и обработки данных для крупных научных проектов (мегапроект НИКА) // Информационные технологии и вычислительные системы. 2013. № 4. С. 37–44.
 13. Davy Cielien, Arno D. B. Meysman, and Mohamed Ali. *Introducing Data Science: Big Data, Machine Learning, and more, using Python tools*. Manning Publ. 2016. 320 p. ISBN: 9781633430037.
 14. РМГ 29-2013. Рекомендации по межгосударственной стандартизации. Государственная система обеспечения единства измерений. Метрология. Основные термины и определения. М.: Стандартинформ, 2014. 60 с.

Скоробогатов Анатолий Михайлович. Федеральное государственное бюджетное учреждение науки Институт проблем безопасного развития атомной энергетики Российской академии наук, Москва, Россия. Научный сотрудник. Количество печатных работ: 25. Область научных интересов: радиэкология, радиационная безопасность, информационные технологии, методы обработки информации, системы управления базами данных. E-mail: sam@ibrae.ac.ru

Панченко Сергей Владимирович. Федеральное государственное бюджетное учреждение науки Институт проблем безопасного развития атомной энергетики Российской академии наук, Москва, Россия. Заведующий лабораторией. Количество печатных работ: 95. Область научных интересов: радиэкология, радиационная безопасность, информационные технологии. E-mail: ranch@ibrae.ac.ru

Апанасюк Олег Николаевич. Федеральное государственное бюджетное учреждение науки Институт проблем безопасного развития атомной энергетики Российской академии наук, Москва, Россия. Главный специалист по мониторингу комплексной безопасности. Количество печатных работ: 46. Область научных интересов: радиационная безопасность, информационные технологии, проектирование информационных систем. E-mail: aon@ibrae.ac.ru

Буланцева Татьяна Анатольевна. Федеральное государственное бюджетное учреждение науки Институт проблем безопасного развития атомной энергетики Российской академии наук, Москва, Россия. Инженер. Количество печатных работ: 7. Область научных интересов: радиационная безопасность, информационные технологии. E-mail: tabu@ibrae.ac.ru

Structure of the Database for Assessing the Long-Term Safety of Radioactive Waste Storage Sites

A. M. Skorobogatov, S. V. Panchenko, O. N. Apanasyuk, T. A. Bulantseva

Nuclear Safety Institute of the Russian Academy of Sciences, Moscow, Russia

Abstract. The paper considers the structure of the database, independent of the subject-oriented field of knowledge. When developing the architecture of the database, the experience of operating a regional database of radioecological data was taken into account. The structure of the database, optimized for the specifics of long-term data storage, which is based on a series of national standards in the field of information technology (Metadata) and the state-impact system for ensuring the unity of measurements, is described. A database with a given structure can store a wide class of data and, if necessary, is infinitely expanded both substantively and physically to the capabilities of Big Data without chang-

ing the structure in the long term. The limit is the memory size of the physical media. Standardizing the formats of data tables and connecting tables of interaction with metadata allows you to simplify and increase the reliability of information conversion and migration procedures.

Keywords: database structure, long-term data storage, metadata.

DOI 10.14357/20718632220206

References

1. RB-003-21. 2021. Rukovodstvo po bezopasnosti pri ispol'zovanii atomnoy energii. Otsenka dol - govremennoy bezopasnosti punktov glubinnogo zakhroneniya radioaktivnykh otkhodov [Safety Guidelines for Atomic Energy Use Long-term safety assessment of deep disposal facilities for radioactive waste]. Utv. post. Rostekhnadzor from 19.03.2021 No. 101, put into effect. 19.03.2021. Moscow: FSI "NTC YARB" (in Russian).
2. Korabelnikov, N. A., Yu. M. Zinyukov. 2019. Struktura bazy dannykh dlya ob"ektov ekologicheskogo monitoringa – poligonov tverdykh bytovykh otkhodov [Database structure for environmental monitoring facilities – solid waste landfills]. Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: geologiya [Vestnik Voronezhskogo gosudarstvennogo universiteta. Series: Geology]. 118–124 (in Russian).
3. Edinyi Gosudarstvennyi Fond Dannykh [Unified State Data Fund]. Available at: <http://meteo.ru/egfd?ysclid=11ufpb8ut1> (accessed April 11, 2022) (in Russian).
4. Shaimardanov, V. M. 2011. Sozдание programmno-apparatnogo kompleksa dolgovremennogo khraneniya i ispol'zovaniya gidrometeorologicheskoi informatsii [Creation of a software and hardware complex for long-term storage and use of hydrometeorological information]. Uchenye zapiski Rossiiskogo gosudarstvennogo gidrometeorologicheskogo universiteta [Scientific notes of the Russian State Hydrometeorological University]. 21:118–125 (in Russian).
5. Rasporyazhenie Pravitel'stva RF ot 3 sentyabrya 2010 g. N 1458-r Strategii deyatel'nosti v oblasti gidrometeorologii i smezhnykh s nei oblastyakh na period do 2030 goda (s uchetom aspektov izmeneniya klimata). [Order of the Government of the Russian Federation dated September 3, 2010, no. 1458-r On the strategy of activities in the field of hydrometeorology and related areas for the period up to 2030 year (taking into account aspects of climate change)]. Available at: <http://static.government.ru/media/files/AAOMfY6IO8AaNaZQVTTLUrYCTRrhNva.pdf> (accessed April 11, 2022) (in Russian).
6. Solovyev, A. V., N. B. Bakanova. 2019. Problems of long-term preservation of big data. Informatsionnye tekhnologii i vychislitel'nye sistemy [Journal of Information Technologies and Computing Systems]. 2:44–53. doi: 10.14357/20718632190205 (in Russian).
7. Solovyev, A. V. 2021. Algorithmic Solution to the Problem of Authenticity of Digital Data. Informatsionnye tekhnologii i vychislitel'nye sistemy [Journal of Information Technologies and Computing Systems]. 1:12–19. doi: 10.14357/20718632210102 (in Russian).
8. Solovyev, A. V. 2021. Solving the problem of interpreting digital data for long-term keeping. Trudy ISA RAN [Proceedings of Institute for Systems Analysis]. 71(2):43–49. doi: 10.14357/20790279210206 (in Russian).
9. Skorobogatov, A. M., M. N. Korsak, M. I. Krolenko. 2020. About the organization of an invariant database in the field of ecology and environmental protection. V sbornike: Nauka, obrazovanie, proizvodstvo v reshenii ekologicheskikh problem (Ekologiya-2020). Materialy XVI Mezhdunarodnoi nauchno-tekhnicheskoi konferentsii, v 2-kh tomakh, posvyashchaetsya 75-letiyu Pobedy v Velikoi Otechestvennoi voine [In the collection: Science, education, production in solving environmental problems (Ecology-2020). Materials of the XVI International Scientific and Technical Conference, in 2 volumes, dedicated to the 75th anniversary of the Victory in the Great Patriotic War]. Ufa: UGATU. 1:155–159 (in Russian).
10. Svidetel'stvo o gosudarstvennoi registratsii bazy dannykh [Certificate of state registration of the database] No. 2018621930. 2018. Regional'naya baza radioekologicheskikh dannykh dlya kolichestvennoi otsenki negativnykh effektiv i riskov ot radiatsionnogo i khimicheskikh faktorov tekhnogennoy vozdeistviya na naselenie i okruzhayushchuyu sredu Sverdlovskoi oblasti. Versiya 2018.01. IBRAE RAN (RU). Data registratsii v Reestre baz dannykh 03 dekabrya 2018 g. [Regional database of radioecological data for the quantitative assessment of negative effects and risks from radiation and chemical factors of technogenic impact on the population and the environment of the Sverdlovsk region. Version 2018.01. IBRAE RAS (RU). Date of registration in the Database Register 03 December 2018] (in Russian).
11. Sambros, V. V., O. V. Lovtskaya, M. V. Petrova, A. N. 2013. Savelenok. Struktura bazy dannykh sistemy jekologicheskogo obespecheniya jekspluatatsii kosmodroma "Vostochnyj" [Structure of the database of the environmental support system for the operation of the Vostochny cosmodrome]. Materialy II Vserossiiskoy nauchno-prakticheskoy konferentsii «Kosmodrom «Vostochnyj» – budushhee kosmicheskoy otrasli Rossii», Blagoveshchenskij gosudarstvennyj pedagogicheskij universitet [Materials of the II All-Russian Scientific and Practical Conference "Vostochny Cosmodrome – the Future of the Space Industry of Russia", Blagoveshchensk State Pedagogical University]. 102–108 (in Russian).
12. Korenkov, V. V., A. V. Nechaevskii, V. V. Trofimov. 2013. Data acquisition simulation for NICA experiment. Informatsionnye tekhnologii i vychislitel'nye sistemy [Journal of Information Technologies and Computing Systems]. (4):37–44 (in Russian).
13. Cielen D., Arno D. B. Meysman, and Mohamed Ali. Introducing Data Science: Big Data, Machine Learning, and

- more, using Python tools. Manning Publications Co, 2016, 320 p. ISBN: 9781633430037.
14. RMG 29-2013. 2014. Rekomendatsii po mezhgosudarstvennoi standartizatsii. Gosudarstvennaya sistema obespecheniya edinstva izmerenii. Metrologiya. Osnovnye terminy i opredeleniya [Recommendations on interstate standardization. State system for ensuring the uniformity of measurements. Metrology. Basic terms and definitions]. Moscow: Standartinform Publ. 60 p. (in Russian).

Skorobogatov A. M. Nuclear Safety Institute of the Russian Academy of Sciences. Russian Federation, 115191, Moscow, Bolshaya Tul'skaya St., 52; e-mail: sam@ibrae.ac.ru

Panchenko S. V. Nuclear Safety Institute of the Russian Academy of Sciences; Russian Federation, 115191, Moscow, Bolshaya Tul'skaya St., 52; email: panch@ibrae.ac.ru

Apanasyuk O. N. Nuclear Safety Institute of the Russian Academy of Sciences; Russian Federation, 115191, Moscow, Bolshaya Tul'skaya St., 52; e-mail: aon@ibrae.ac.ru

Bulantseva T. A. Nuclear Safety Institute of the Russian Academy of Sciences; Russian Federation, 115191, Moscow, Bolshaya Tul'skaya St., 52; e-mail: tabu@ibrae.ac.ru