

Приложение 1
к требованиям по
управлению цифровыми данными

Форма

Паспорт цифровых данных цифровых объектов

1. Общая часть

1) наименование организации:

указывается полное наименование организации;
к примеру: государственное учреждение «Министерство юстиции Республики Казахстан»;

2) наименование цифрового объекта, осуществляющего сбор и обработку цифровых данных:

указывается полное наименование цифрового объекта, осуществляющего сбор и обработку цифровых данных (так, как оно закреплено в соответствующих нормативных правовых актах (далее – НПА), при наличии);

к примеру: национальный реестр индивидуальных идентификационных номеров;

3) уровень межведомственного использования цифровых данных:

указывается уровень межведомственного использования цифровых данных выбором одной из следующих возможных альтернатив: республиканский; местный; цифровые данные на межведомственном уровне не используются;

4) способ ведения:

указывается один из следующих способов ведения: в бумажном виде (указывается для данных, не переведенных в электронный вид); в электронном виде (указывается для данных, представленных в электронном виде); комбинированный (указывается для данных, которые частично переведены в электронный вид);

5) форма цифровых данных:

указываются одна или несколько из следующих форм цифровых данных: первичные цифровые данные; агрегированные цифровые данные;

6) доступность цифровых данных:

указываются уровни доступности цифровых данных: ограниченный доступ (указывается при наличии персональных данных, коммерческой тайны, служебной тайны и данных, охраняемых законами Республики Казахстан); доступ без ограничений (указывается для открытых данных) (при разном уровне доступа к отдельным частям цифровых данных в данной строке проставляются оба перечисленных выше варианта с детализацией сведений);

7) пользователи цифровых данных согласно НПА:

указывается полный перечень наименований пользователей цифровых данных согласно НПА, если круг пользователей определен НПА, получивших интеграции;

8) источник цифровых данных согласно НПА:

указывается согласно НПА полный перечень цифрового объекта.

В отсутствии НПА указывается «Не определено НПА».

9) источники цифровых данных фактические:

указывается фактический перечень цифрового объекта.

Если состав фактических источников цифровых данных совпадает с составом источников цифровых данных согласно НПА, указывается, что перечень фактических источников совпадает с предусмотренным НПА.

10) объект описания:

указываются объекты описания, к которым относятся: физические лица; юридические лица; имущество, включая наличные деньги и документарные ценные бумаги, безналичные денежные средства, бездокументарные ценные бумаги, имущественные права; результаты работ и оказания услуг; охраняемые результаты интеллектуальной деятельности и приравненные к ним средства индивидуализации (интеллектуальная собственность); нематериальные блага;

11) периодичность обновления:

указывается один из следующих вариантов периодичности обновления цифровых данных: периодичность обновления не установлена; обновление по мере необходимости; обновление по мере поступления цифровых данных; непрерывное обновление (поточные цифровые данные); периодичность обновления установлена НПА (указать наименование НПА и периодичность обновления);

12) график обновления:

указываются в зависимости от особенностей обновления цифровых данных: заданный интервал времени с момента последнего обновления; перечень плановых календарных дат; реквизиты НПА, регламентирующего график обновления (при отсутствии графика обновления цифровых данных указывается «Не определен»);

13) порядок обновления:

указываются порядок обновления цифровых данных и его особенности (например, необходимость подписания/согласования НПА) (если порядок установлен НПА, то указывается наименование такого НПА);

14) перечень НПА, на основании которого создан цифровой объект, осуществляющий сбор и обработку цифровых данных:

указывается перечень НПА, регламентирующих сбор и обработку цифровых данных (положения, регламенты информационного взаимодействия, раздел не заполняется, если цифровые данные ведутся в бумажном виде);

15) перечень НПА, регламентирующих сбор и обработку цифровых данных:

указывается перечень НПА, содержащих функции ведения объекта учета или цифровых данных организацией;

16) перечень НПА, регламентирующих внесение и потребление цифровых данных, ограничение и разделение доступа к цифровым данным;

указывается перечень НПА, определяющих правила функционирования, внесения или передачи цифровых данных, правила доступа к цифровым данным и использования этих цифровых данных, правила разграничения доступа к цифровым данным;

17) класс цифровых данных;

18) дата создания – дата создания паспорта цифровых данных.

2. Описание жизненного цикла цифровых данных:

1) создание сведений об объекте учета:

указывается порядок первоначального формирования цифровых данных об объекте учета, при этом указывается: являются ли цифровые данные первичными или формируются на основании цифровых данных, предоставляемых из не утвержденных источников (если при формировании записи об объекте учета используются сведения из разных источников информации, то необходимо перечислить их с указанием соответствующих НПА);

2) использование сведений об объекте учета:

указываются сведения о порядке использования цифровых данных об объекте учета, являются ли цифровые данные об объекте учета (все или их часть) первоисточником; существующие механизмы передачи цифровых данных (например, файловая передача цифровых данных; передача цифровых данных посредством интеграционного сервиса); используют ли цифровые данные об объекте учета представители бизнеса (с описанием, кто и для чего использует цифровые данные); уровень доступности цифровых данных (существуют ли правовые ограничения на использование цифровых данных);

3) изменение объекта учета, удаление сведений об объекте учета – указание или описание этапа нахождения жизненного цикла цифровых данных объекта учета:

создание (сбор) цифровых данных;

хранение цифровых данных и их ведение;

использование цифровых данных;

управление цифровыми данными;

архивирование цифровых данных (или их уничтожение).

3. Описание структуры цифровых данных

порядковый номер атрибута данных в таблице (id);

наименование цифрового объекта (information_system_name);

системное наименование базы данных (bd_name);

наименование схемы в базе данных (schema_name);

наименование таблицы базы данных (table_name);

описание назначения таблицы базы данных (table_description);

указание типа таблицы базы данных (table_type);
порядковый номер столбца в таблице базы данных (column_id);
наименование столбца в таблице базы данных (column_name);
тип данных столбца в таблице базы данных (column_type);
назначение столбца в таблице базы данных (column_description);
ссылка на другую таблицу базы данных, при наличии, (reference_tbl_clmn);
ограничение столбца в таблице базы данных (flk);
вид ключа столбца в таблице базы данных (clmn_key_type);
семантический словарь тегов (teg);
мастер данные (isMD);
государственная услуга (isGosService).

Приложение 2
к требованиям
по управлению цифровыми
данными

Форма

Описание неоцифрованных данных

1. Вид объекта неоцифрованных данных (например: субъект, объект, документ).
 2. Наименование сущности субъекта или объекта неоцифрованных данных (например: пассажир, перевозчик, пациент, врач, пенсионер, здание, земельный участок, транспортное средство).
 3. Описание объекта неоцифрованных данных (указать наименование нормативных правовых актов, регламентирующих сбор и обработку цифровых данных).
 4. Наименование атрибута (например: фамилия, имя, отчество (при наличии), дата рождения, номер документа, номер транспорта).
 5. Национальная справочная информация.
-

Приложение 3
к требованиям
по управлению цифровыми
данными

Форма

Паспорт кейса

1. Организация_____.
2. Название кейса_____.
3. Постановка цели_____.
4. Описание_____.
5. На какие вопросы должен отвечать дашборд_____.
6. Показатели (с учетом разрезности, методик/формул формирования, источника):

6.1. Описание показателя:

№ п/п	Показатель	Разрезность	Источник	Периодичность обновления

- 6.2. Методика расчета показателя_____.
 7. Описание необходимой аналитики: (описательный анализ, анализ «что-если», прогнозный анализ)_____.
 8. КРІ дашборда (ключевые показатели эффективности (КРІ), которые необходимо отслеживать)_____.
 9. Планируемые графики для отображения результата (согласованный макет графика)_____.
 10. Согласованные допущения при реализации кейса_____.
 11. Согласование сторон_____.
- _____

Приложение 4
к требованиям
по управлению
цифровыми данными

Форма

**Описание аналитического дашборда на архитектурном портале
цифровых объектов государства**

1. Наименование дашборда: _____;
(краткое и понятное название)
2. Цель/назначение: _____;
(например: мониторинг ключевых показателей региона, контроль исполнения бюджета и так далее)
3. Государственный орган/организация: _____;
(владелец/разработчик дашборда)
4. Ответственное подразделение и контактное лицо: _____;
(фамилия, имя, отчество (при наличии), должность, e-mail, телефон)
5. Адрес сайта или ссылка на ведомственную платформу: _____;
6. Контур: _____;
(единая транспортная среда государственных органов /интернет)
7. Частота обновления данных: _____;
(ежедневно / еженедельно / ежеквартально/ раз в полгода/ раз в год/ по мере поступления данных и так далее)
8. Перечень источников: _____;
(наименование цифровой системы, базы данных)
9. Перечень показателей: _____;
10. Метод обновления данных: _____;
(автоматическое / полуавтоматическое / вручную)
11. Инструмент визуализации
_____ .

Приложение 5
к требованиям
по управлению
цифровыми данными

Форма

Паспорт библиотеки данных

1. Общие положения

- 1) назначение документа;
- 2) область применения (обучение модели, дообучение модели, поиск по базе, тестирование, оценка качества, эксплуатация);
- 3) термины и определения;
- 4) наименование и версия библиотеки данных;
- 5) наименование владельца библиотеки данных;
- 6) наименование поставщика библиотеки данных и его структурного подразделения, фактически формирующего библиотеку данных;
- 7) наименование организации, размещающей библиотеку данных;
- 8) наименование организации, осуществляющей разработку цифрового объекта, взаимодействующего с библиотекой данных;
- 9) основание формирования библиотеки данных (наименование и реквизиты нормативного правового акта, поручения, договора, протокола, письма);
- 10) сведения об ответственных лицах со стороны владельца библиотеки данных, включающие:
 - фамилию, имя, отчество (при наличии) менеджера по данным;
 - фамилию, имя, отчество (при наличии) ответственного за качество данных;
 - фамилию, имя, отчество (при наличии) ответственного за кибербезопасность;
 - фамилию, имя, отчество (при наличии), должность, подразделение, контакты сотрудника, реализующего извлечение, интеграцию данных;
- 11) сведения о лицах, осуществляющих согласование паспорта библиотеки данных со стороны владельца библиотеки данных, включающие:
 - фамилию, имя, отчество (при наличии);
 - должность;
 - дату согласования;
 - реквизиты документа согласования.

2. Цели и сценарии использования библиотеки данных

1) наименование компонента, использующего библиотеку данных (система искусственного интеллекта, платформенный программный продукт национальной платформы искусственного интеллекта, модель искусственного интеллекта, модуль формирования базы знаний, модуль классификации, модуль генерации документов, аналитический модуль и тому подобное);

2) цели использования библиотеки данных (обучение, дообучение, тестирование, оценка, эксплуатация, формирование базы знаний, формирование эталонного набора данных);

3) ключевые сценарии использования библиотеки данных (не менее 3 (три), не более 10 (десять)).

3. Состав данных и классификация

1) типы данных (структурированные данные (таблицы или витрины данных), текстовые данные; неструктурированные данные (сканированные документы, изображения, аудио), национальная справочная информация, эталонный набор данных);

2) объем и период охвата формирования библиотеки данных;

3) язык документов библиотеки данных.

4. Источники и происхождение данных

1) перечень источников данных, включая сведения:

наименование цифровой системы, реестра, хранилища, архива;

наименование владельца источника данных;

правовое основание владения данными;

тип данных;

способ извлечения данных (прикладной программный интерфейс (далее - API), пакетная загрузка данных (далее - ETL), потоковая передача изменений данных (далее- CDC), выгрузка);

периодичность обновления;

2) ограничения по источникам, связанные с запретом на использование данных, выгрузке данных, исключению атрибутов или их обезличиванию;

3) наименование первичного эталонного источника;

4) ожидаемый корректный результат для каждого элемента эталонного набора данных.

5. Форматы данных

1) перечень форматов текстовых данных;

2) перечень форматов неструктурированных данных;

3) перечень форматов структурированных данных.

6. Описание структуры данных и метаданные

1) схема библиотеки данных включает:

наименование таблицы;

назначение таблицы;

вид таблицы (основная, справочная, журнал событий, витрина данных);

тип таблицы (содержит первичный или внешний ключ);

периодичность обновления;

2) описание структуры таблицы включает:

наименование таблицы;

наименование поля таблицы;

тип базы данных;

признак типа формата данных (числовой, текстовый, категория, дата и время, логический, географический);

обязательность наличия данных в поле таблицы (ограничение Not Null);

принимаемые допустимые значения;

наименование ограничения, определяющего первичный или внешний ключ;

разрешение на дальнейшее использование данных оператором национальной платформы искусственного интеллекта для целей формирования иных библиотек данных (да или нет);

3) описание структуры метаданных включает:

уникальный идентификатор документа;

наименование источника данных;

вид документа;

язык документа;

заголовок страницы, описывающей контент документа;

дату создания или актуализации документа;

статус документа (утвержденный, утратил силу, проект);

контрольную сумму документа;

сведения о наличии персональных данных в документе или сведений конфиденциального характера;

текстовую информацию документа.

7. Описание порядка подготовки данных (ETL, CDC, очистка, преобразование)

1) описание маршрута данных в последовательности:

источник данных;

способ извлечения;

зона приема данных (зона для временного хранения данных перед их обработкой или нулевая витрина данных);

способ преобразования, очистки;

наименование витрины или каталога;

наименование компонента, использующего библиотеку данных;

2) описание правил преобразования (нормализация форматов дат или кодировок, правила удаления дублей, правила обработки пропусков, правила сопоставления с национальной справочной информацией, правила исключения запрещенных атрибутов);

3) описание механизма определения изменения данных.

8. Качество и механизмы контроля качества данных

1) описание перечня применимых критериев качества (точность, полнота, взаимосвязанность, сопоставимость, актуальность, уникальность, достоверность, своевременность);

2) описание проверок для реализации форматно-логического контроля данных (заполненность обязательных атрибутов по ограничению Not Null, совпадение типов данных, диапазонов значений данных, применяемых масок, отсутствие дублирующих значений ключей и идентификаторов, корректность связей между таблицами по значениям первичного и внешних ключей, корректность допустимых значений в национальной справочной информации, соблюдение сроков обновлений);

3) описание результатов проверки качества данных, включающее:

объем данных;

долю пропусков по обязательным полям (в процентах);

долю выявленных дублированных данных (в процентах);

количество нарушений форматно-логического контроля;

перечень критичных дефектов;

4) описание порядка обработки дефектов качества, включающее:

способ фиксации инцидентов;

сроки реакции на выявленный инцидент;

фамилию, имя, отчество (при наличии) ответственных за обработку инцидента;

порядок повторной передачи исправленных данных для формирования библиотеки данных.

9. Эталонный набор данных для тестирования библиотеки данных

1) описание эталонного набора данных, включающее:

назначение эталонного набора данных;

сценарии использования эталонного набора данных;

объем эталонного набора данных;

покрытие;

2) описание формата эталонного набора данных, включающее:

идентификатор эталонного набора данных;

сценарий использования;
 вид входных данных (запрос, документ, набор полей);
 ожидаемый корректный результат;
 правила прохождения;
 источник эталонного набора данных.

10. Описание механизмов защиты данных

- 1) описание режима доступа к данным и ролей доступа (открытый или ограниченный);
- 2) описание наличия конфиденциальных и (или) персональных данных, иных ограничений, связанных с их обработкой;
- 3) описание способа обезличивания данных (при необходимости), включающее:
 - наименование атрибута, подлежащего обезличиванию;
 - описание метода (маскирование, хэширование, агрегация, удаление);
 - места обезличивания;
 - способы и результаты контроля корректности обезличивания;
- 4) описание журналирования событий, включающее:
 - тип события, подлежащего журналированию (доступ, изменение);
 - срок и место хранения журналов событий.

11. Порядок предоставления и формат поставки данных для формирования библиотеки данных

- 1) описание способа предоставления данных (API, ETL, CDC, защищенная файловая передача, база данных со статусом «только для чтения»);
 - 2) описание комплекта поставки данных, включающее сведения о:
 - объеме и формате данных;
 - наличии метаданных;
 - наличии описания результатов проверок для реализации форматно-логического контроля данных эталонного набора данных (при наличии требований к оценке качества);
 - контрольных суммах переданных файлов с данными или выгрузок данных;
 - 3) описание периодичности поставки данных и параметры SLA, включающие:
 - частоту (периодичность) поставки данных;
 - допустимую задержку поставки данных,
 - время окна обслуживания.
-