Методика описания концептуальной модели государственных данных

(для поставщиков данных в ЕИП НСУД)

Оглавление

[1 Определения, обозначения и сокращения 3](#_Toc52358896)

[2 Общие положения 4](#_Toc52358897)

[3 Методология реализации онтологического подхода к управлению государственными данными 4](#_Toc52358898)

[3.1 Основные сведения 4](#_Toc52358899)

[3.2 Методология концептуального анализа 7](#_Toc52358900)

[3.3 Процедура описания предметной области методами концептуального анализа 11](#_Toc52358901)

[3.4 Перевод онтологии в машиночитаемый формат OWL 20](#_Toc52358902)

[3.4.1 Обзор используемых программных средств 20](#_Toc52358903)

[3.4.2 Используемая терминология 21](#_Toc52358904)

[3.4.3 Ведение разработки онтологий в Protégé 23](#_Toc52358905)

[4 Список использованных источников 25](#_Toc52358906)

Определения, обозначения и сокращения

| **Термин, сокращение** | **Определение** |
| --- | --- |
| OWL | (англ. Web Ontology Language) — язык описания онтологий |
| Государственные данные | Данные, содержащиеся в государственных (муниципальных) информационных системах, государственных (муниципальных) информационных ресурсах или данные, полученные и используемые органом государственной власти (органом местного самоуправления) от иных правообладателей в соответствии с требованиями законодательства |
| Класс (сущность) онтологической модели | Множество объектов, объединенных по ряду признаков. Обычно класс соответствует понятию (концепту), или мысленному образу объекта предметной области |
| КМД | Концептуальная модель данных |
| НСУД | Национальная система управления данными |
| Онтология (в узком смысле) | Описание таксономии (иерархии понятий) предметной области |
| Онтология (в широком смысле) | Раздел философии, изучающий бытие, а также его формы и закономерности |
| Предметная область | Предмет описания онтологии, представляющий собой совокупность понятий, объединенных определенным контекстом |
| Термин | Обозначение понятия, или концепта, на естественном языке |

Общие сведения

## Основные сведения о концептуальной модели данных

Концептуальная модель государственных данных (КМД) описывает онтологию (в узком смысле) предметных областей государственного управления. КМД включает сведения о понятиях (сущностях), описывающих субъекты и объекты государственного управления, такие как «Физическое лицо», «Паспорт гражданина Российской Федерации», «Транспортное средство». Близкие по смыслу понятия объединяются в иерархию. Для каждого понятия КМД возможно описать его атрибуты и связи с иными понятиями КМД.

С учетом целей, задач и специфики НСУД, связанной с общенациональным масштабом проводимой разработки, с наличием значительного количества вовлекаемых в проект государственных информационных систем, а также с инновационным характером разработки, подход к построению КМД осуществляется двумя одновременными взаимосвязанными потоками:

* первый поток — так называемое проектирование КМД «сверху вниз»;
* второй поток — так называемое проектирование КМД «снизу вверх».

Такой способ проектирования предполагает:

* взгляд на предметную область, который понятен онтологу-аналитику и специалистам из предметной области (проектирование «сверху вниз»);
* информационно-технологический взгляд на модель данных, который понятен только ИТ-специалисту; польза состоит в том, что при этом рационально используется весь исторически наработанный массив данных в унаследованных государственных информационных ресурсах и системах (проектирование «снизу вверх»).

## Проектирование концептуальной модели данных «снизу вверх»

Для каждой предметной области (домена) государственного управления может формироваться свой фрагмент КМД, который впоследствии встраивается в общегосударственную (глобальную) КМД.

Собственно, понятия (сущности) КМД образуют глобальный глоссарий государственного управления.

Для увязки каталога данных с КМД осуществляется сопоставление метаданных логического уровня каталога данных (наименований атрибутов или контейнеров) с понятиями или атрибутами понятий КМД. В результате образуется ведомственная карта государственных данных.

Под «метаданными» понимается информация, которая описывает / логически увязывает другие данные, что помогает понять происхождение, структуру, предметную область и контекст данных.

В случае если в КМД отсутствует сущность (класс), необходимая для привязки, рекомендуется создавать новую сущность в статусе «Предложение».

*Пример*: в нескольких видах сведений от ФНС России, описывающих предоставление сведений из ЕГР ЗАГС ([Предоставление из ЕГР ЗАГС по запросу сведений о заключении брака](https://nsud.gosuslugi.ru/pmd/inquire_kinds/731); [Предоставление из ЕГР ЗАГС по запросу сведений о расторжении брака](https://nsud.gosuslugi.ru/pmd/inquire_kinds/733); [Предоставление из ЕГР ЗАГС по запросу сведений о рождении](https://nsud.gosuslugi.ru/pmd/inquire_kinds/734); [Предоставление из ЕГР ЗАГС по запросу сведений о смерти](https://nsud.gosuslugi.ru/pmd/inquire_kinds/735)) имеются атрибуты: «Признак отсутствия сведений о нотариальном засвидетельствовании документа», «Признак отсутствия наименования документа иностранного государства» и другие, которые описывают документ, подтверждающий регистрацию акта гражданского состояния и выданный компетентным органом иностранного государства. В КМД не было подобной сущности для привязки, и было решено создать сущность *«Документ, подтверждающий регистрацию акта гражданского состояния и выданный компетентным органом иностранного государства»* под классом *«Документ, подтверждающий регистрацию акта гражданского состояния»* (родительский класс: *«Документ»*).

В виде сведений от Фонда пенсионного и социального страхования Российской Федерации [Информирование из ЕГИССО по СНИЛС](https://nsud.gosuslugi.ru/pmd/inquire_kinds/321) (и ряде других) содержатся атрибуты «Признак включения в выписку сведений о фактах назначения», «Дата окончания действия назначения» и др., описывающие назначение мер социальной защиты. В КМД не было подходящей сущности для привязки, и было решено создать сущность *«Назначение мер социальной защиты (поддержки)»* под классом *«Действие»*.

Предложения от поставщиков данных в ЕИП НСУД регулярно рассматриваются онтологами Центра компетенций в ЕИП НСУД. Подробнее о том, какие термины могут быть сущностями КМД и как описывать новые термины в КМД, см. в разделе 3.

## Ограничения, накладываемые на описание сущностей КМД

* + 1. Ограничения, накладываемые на описание сущностей КМД в связи с идеологией формата OWL, лежащего в основе модели

Основным ограничением, которое данный формат накладывает на описание сущностей КМД, является следующее: сущность КМД, являющаяся подклассом другой сущности, должна наследовать все атрибуты надкласса и все его связи с другими сущностями КМД. Если она не может наследовать данные признаки, то следует найти для нее другое место в КМД.

Например, у сущности «Физическое лицо» в КМД есть атрибуты: «Фамилия», «Имя», «Отчество», «Пол», «Дата рождения», «Место рождения» и др. Сущность «Ответственное лицо» не может наследовать такие атрибуты, т. к. ответственным лицом может быть не только физическое, но и юридическое лицо. Следовательно, сущность «Ответственное лицо» не может быть подклассом класса «Физическое лицо». В итоге, объект КМД «Ответственное лицо» был помещен под класс «Субъект права», и к нему были добавлены именованные связи:

*Ответственное лицо «может являться» Физическое лицо*

*Ответственное лицо «может являться» Юридическое лицо*

***Примечание***:

Связь между сущностями КМД – это отношение, соединяющее два класса концептуальной модели и образующее триплет (предложение, состощее из трех членов: Субъект + Предикат + Объект). Например, класс «Организация» в КМД имеет связи:

*Организация «может иметь» Офис*

*Организация «может являться» Юридическое лицо*

где «Организация» – субъект, «может иметь» и «может являться» – предикаты, «Офис» и «Юридическое лицо» – объекты.

Связи отображаются в блоке «Связи с объектами КМД» окна концептуальной модели.

* + 1. Ограничения, связанные с редактором концептуальной модели в ЕИП НСУД

Редактор концептуальной модели в ЕИП НСУД основан на принципах программного продукта Protégé (редактор онтологических моделей, распространяемый на условиях свободного программного обеспечения и доступный в форматах настольного приложения и веб-приложения). По состоянию на сентябрь 2023 г., инструмент находится в процессе доработки, поэтому реализован не весь необходимый функционал.

В частности, не реализованы следующие функции:

* наследование атрибутов и связей надклассов – подклассами.

Например, класс «Крупногабаритное транспортное средство» является подклассом класса «Транспортное средство». У класса «Транспортное средство» есть атрибуты: *Государственный регистрационный номер, Идентификационный номер (VIN), Год выпуска, Марка, Модель* и др. Все эти атрибуты должны наследоваться классом «Крупногабаритное транспортное средство». Поскольку в настоящее время механизм наследования в ЕИП НСУД не реализован, наследуемые атрибуты подклассов онтологи заводят вручную. Тем не менее, данный механизм в перспективе будет существовать, и его необходимо учитывать. Иными словами, прежде чем помещать новую сущность под тот или иной класс, необходимо проверить, какие атрибуты / связи есть у данного класса или его надклассов;

* множественное наследование подклассов (когда один и тот же класс может являться «потомком» нескольких классов иерархии). В ЕИП НСУД множественное наследование реализовано исключительно посредством именованных связей.

Например, класс «Свидетельство о рождении» является подклассом класса «Свидетельство» (подкласс сущности «Документ»). Но в то же время «Свидетельство о рождении» является документом, удостоверяющим личность физического лица до получения им паспорта. В силу ограничений онтологического редактора в ЕИП НСУД, невозможно создать класс «Свидетельство о рождении» в двух частях КМД: под классом «Свидетельство» и под классом «Документ, удостоверяющий личность». Поэтому класс «Свидетельство о рождении» был помещен под класс «Свидетельство», и для него была создана именованная связь: *Свидетельство о рождении «является» Документ, удостоверяющий личность*.

Обращаем внимание на то, что механизм двойного наследования в онтологическом редакторе ЕИП НСУД останется неизменным и после доработки инструмента, т. е. для отражения множественного наследования рекомендуется использовать описанный выше способ.

# Создание сущностей концептуальной модели данных (КМД): основные рекомендации

В процессе создания сущностей (объектов) КМД в ЕИП НСУД необходимо учитывать следующее:

1. Какие термины могут быть объектами КМД:
	1. прежде всего, имеющие нормативные определения. Проверку нормативных определений рекомендуется осуществлять по «Словарю нормативных определений» в правовой системе «Консультант Плюс»;
	2. ключевые термины предметной области, не имеющие нормативных определений (например, термины информационных технологий, важные в контексте ряда видов сведений, связанных с ИТ-тематикой: «Файл», «Артефакт (ИТ)» и др.).
2. Одно и то же понятие может обозначаться разными терминами (синонимия). Разные термины, обозначающие одно и то же понятие, являются *одной* сущностью КМД, а не двумя, тремя и т. д. Это означает следующее:
	1. прежде чем добавлять новую сущность в КМД, необходимо *проверить* наличие в КМД других обозначений данного понятия (синонимичных терминов). Проверка осуществляется при помощи полнотекстового поиска.

Например, в ведомственной ИС или витрине данных используется термин *«Мера социальной поддержки»*, и в КМД есть сущность, описывающая данное понятие, но наименование сущности отличается от термина ведомственной ИС или ВД: *«Мер****ы*** *социальной защиты* ***(поддержки)****»*. Это означает, что создавать новый объект в КМД не нужно.

В ведомственной ИС или витрине данных используется термин *«Организация спортивной подготовки»*, и в КМД есть класс, описывающий данное понятие, но его наименование также отличается от термина ведомственной ИС или ВД: *«Организация,* ***осуществляющая*** *спортивную подготовку»*. Это означает, что и в данном случае создавать новый объект в КМД не нужно.

Если аналитик считает, что в КМД должен быть отражен и другой термин, обозначающий данное понятие, то заводить отдельный класс для второго термина не нужно – достаточно добавить вариант названия к уже имеющемуся термину в КМД. См. пример: *«Паспорт Гражданина РФ, удостоверяющий его личность за пределами территории Российской Федерации (Заграничный паспорт, Загранпаспорт))»*. В данном наименовании сущности КМД отражены: официальное наименование документа из НПА (*Паспорт Гражданина РФ, удостоверяющий его личность за пределами территории Российской Федерации*) и неофициальные, обиходные наименования того же типа документа в скобках (*Заграничный паспорт, Загранпаспорт*) – для удобства пользователя, который, скорее всего, будет искать в КМД более привычное ему неофициальное наименование.

1. Все сущности КМД наследуют атрибуты и связи своих надклассов. Соответственно, нельзя добавлять к сущности подкласс, если он не может наследовать атрибуты или связи данного класса. Например, не всякий «Балансодержатель» имеет ФИО, место и дату рождения (атрибуты физического лица), т. к. «Балансодержателем» может быть как физическое, так и юридическое лицо. Следовательно, «Балансодержатель» может быть подклассом класса «Субъект права», а не класса «Физическое лицо».
2. Каждая сущность, добавляемая в КМД, должна обязательно иметь:
* *наименование на русском языке,*
* *определение (описание объекта КМД)*,
* *источник определения* (*документ-основание* либо *внешний источник*).

Остальные поля в карточке добавляемой сущности (*Наименование объекта КМД на английском, Техническое наименование объекта КМД*) являются факультативными. Теги в настоящий момент не присваиваются.

***Некоторые разъяснения по общей структуре КМД***

На самом верхнем уровне сущности КМД подразделяются на «Объекты», «Явления» и «Понятия». Строго говоря, каждая сущность КМД отражает «понятие» – ветвь иерархии «Понятие» условна и создана для абстрактных сущностей, которые невозможно отнести ни к объектам, ни к явлениям.

Далее класс «Объект» подразделяется на большие подклассы: «Актор» (деятель), «Материальный объект» и «Нематериальный объект» (именно под данным классом лежит значимый верхнеуровневый класс «Организация»).

Класс «Явление» подразделяется на большие подклассы: «Действие», «Деятельность», «Обстоятельство», «Правовое состояние», «Причина», «Событие» и др.

Каждый из больших подклассов, в свою очередь, также имеет подклассы, образуя отдельную ветвь иерархии.

Перед началом работы с КМД рекомендуется изучить структуру верхнеуровневых классов, ознакомиться с их определениями, атрибутами и связями, т. к. эти атрибуты и связи будут наследоваться подклассами данных классов.

В случае возникновения вопросов рекомендуется обращаться в ЦК НСУД, к ответственным специалистам ФГАУ НИИ «Восход».

# Список использованных источников

При составлении методики использовались следующие источники:

1. Федеральный закон от 27.07.2006 № 149-ФЗ (ред. от 02.12.2019) «Об информации, информационных технологиях и о защите информации»
2. Постановление Правительства Российской Федерации от 03.06.2019 № 710 «О проведении эксперимента по повышению качества и связанности данных, содержащихся в государственных информационных ресурсах».
3. Распоряжение Правительства Российской Федерации от 03.06.2019 № 1189-р «Об утверждении Концепции создания и функционирования национальной системы управления данными и плана мероприятий («дорожной карты») по созданию национальной системы управления данными на 2019 - 2021 годы».
4. Методические рекомендации «О Регламенте (стандарте) управления данными государственных информационных систем, создаваемых, развиваемых и эксплуатируемых на Единой цифровой платформе Российской Федерации «ГосТех»» (Проект).
5. ГОСТ Р 56272-2014, часть 8 «Практические методы интеграции распределенных систем: практическая реализация сетевого языка онтологий OWL»
6. Горшков С. Введение в онтологическое моделирование. Ревизия 2.3. Москва: «Тринидата», 2016
7. ISO/IEC DIS 21838-2 Information technology. Top-level ontologies. Part 1: Requirements; Part 2: Basic Formal Ontology (Проект)
8. OWL 2 Web Ontology Language. Manchester Syntax (Second Edition) / W3C Working Group Note 11 December 2012. URL: http://www.w3.org/TR/owl2-manchester-syntax.
9. OWL 2 Web Ontology Language. Profiles (Second Edition) / W3C Recommendation 11 December 2012. URL: http://www.w3.org/TR/owl2-profiles.
10. OWL 2 Web Ontology Language. Structural Specification and Functional-Style Syntax (Second Edition). W3C Recommendation 11 December 2012.