ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р ИСО 12620 — 2012

Терминология, другие языковые ресурсы и ресурсы содержания. Спецификация категорий данных и ведение реестра категорий данных для языковых ресурсов

ISO 12620:2009

Terminology and other language and content resources — Specification of data categories and management of a Data Category Registry for language resources
(IDT)

Издание официальное

Москва
Стандартинформ
2014

Предисловие

Цели и принципы стандартизации в Российской Федерации установлены Федеральным законом от 27 декабря 2002 г. № 184-ФЗ «О техническом регулировании», а правила применения национальных стандартов Российской Федерации — ГОСТ Р 1.0—2004 «Стандартизация в Российской Федерации. Основные положения»

Сведения о стандарте

1 ПОДГОТОВЛЕН Закрытым акционерным обществом «Проспект» на основе собственного аутентичного перевода на русский язык стандарта, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 55 «Терминология, элементы данных и документация в бизнес-процессах и электронной торговле»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 20 ноября 2012 г. № 970-ст

4 Настоящий стандарт идентичен международному стандарту ISO 12620:2009 «Терминология, другие языковые ресурсы и ресурсы содержания. Спецификация категорий данных и ведение реестра категорий данных для языковых ресурсов» (ISO 12620:2009 «Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources»).

При применении настоящего стандарта рекомендуется использовать вместо ссылочных международных стандартов соответствующие им национальные стандарты Российской Федерации, сведения о которых приведены в дополнительном приложении ДА.

5 ВВЕДЕН ВПЕРВЫЕ

Информация об изменениях к настоящему стандарту публикуется в ежегодно издаваемом информационном указателе «Национальные стандарты», а текст изменений и поправок — в ежемесячно издаваемых информационных указателях «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ежемесячно издаваемом информационном указателе «Национальные стандарты». Соответствующая информация, уведомления и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет

© Стандартинформ, 2014

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Введение

Идентификация, сбор, администрирование и хранение данных, ассоциируемых с языковыми ресурсами, выполняются в многочисленных разнообразных средах. Элементы данных, входящие в отдельные языковые ресурсы, рассматриваются в настоящем стандарте как *категории данных*, согласно наименованию, общеупотребительному в Техническом комитете ИСО/ТК 37. *Категории данных* в терминологии стандартов ИСО/ТК 37 соответствуют *концепциям элементов данных* стандартов серии ИСО/МЭК 11179, но несколько отличаются от них в отношении определяемых значений. Различия в подходах, используемых для разных типов языковых ресурсов и конкретных систем различного назначения, неизбежно приводят к отличиям в определениях и именах категорий данных. Использование единообразных имен и определений категорий данных для ресурсов одной тематической области (например, для терминологических ресурсов, лексикографических ресурсов, текстовых аннотаций и т.д.) по крайней мере на уровне обмена, способствует согласованности систем и расширяет возможности повторного использования данных. Процедуры определения категорий данных в конкретной тематической области также должны быть единообразными для обеспечения функциональной совместимости категорий данных, которая становится проблематичной, если эти категории данных определяются в разных реестрах.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Терминология, другие языковые ресурсы и ресурсы содержания. Спецификация категорий данных и ведение реестра категорий данных для языковых ресурсов

Terminology and other language and content resources. Specification of data categories and management of a Data Category Registry for language resources

Дата введения – 2014–01–01

1 Область применения

В настоящем стандарте приведены руководящие указания относительно ограничений реализации реестра категорий данных (DCR) для любых типов языковых ресурсов, например, терминологических, лексикографических, основанных на использовании сборников или машинного перевода и т.д. В настоящем стандарте определены механизмы создания, выбора и ведения категорий данных, а также формат обмена для представления этих категорий.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на международные стандарты. В случае ссылок на стандарты, для которых указана дата утверждения, необходимо пользоваться только указанной редакцией. В случае, когда дата утверждения не приведена, следует пользоваться последней редакцией ссылочных стандартов, включая любые поправки и изменения к ним:

ИСО 8601:2004 Элементы данных и форматы обмена. Обмен информацией. Представление дат и времени (ISO 8601:2004, Data

---

Издание официальное

elements and interchange formats – Information interchange – Representation of dates and times)

ИСО/МЭК 11179-1:2004 Информационные технологии. Реестры метаданных (MDR). Часть 1. Структура (ISO/IEC 11179-1:2004, Information technology – Metadata registries (MDR) – Part 1: Framework)

ИСО/МЭК 11179-3:2003 Информационные технологии. Реестры метаданных (MDR). Часть 3. Метамодель системного регистра и основные признаки (ISO/IEC 11179-3:2003, Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes)

3 Термины и определения

В настоящем документе используются термины и определения, приведенные в ИСО/МЭК 11179-1:2004, а также перечисленные ниже термины с соответствующими определениями.

3.1 Элементы данных и категории данных

3.1.1 **элемент данных** (data element): ⟨применительно к языковым ресурсам⟩ единица данных, которая в определенном контексте считается неделимой.

Примечание - В работах по терминологии отдельное поле, например, /term/ (термин), входящее в одну терминологическую запись, рассматривалось как элемент данных и конкретный пример **категории данных** (3.1.3).

3.1.2 **элемент данных** (data element; DE): ⟨применительно к стандартам метаданных⟩ единица данных, для которой определение, идентификация, представление и область значений установлены с помощью набора признаков.

[ИСО/МЭК 11179-1:2004, 3.3.8]

3.1.3 **категория данных** (data category; DC): Результат спецификации конкретного поля данных.

Пример - /partOfSpeech/ (часть речи), /grammaticalGender/ (грамматический род), /grammaticalNumber/ (грамматическое число). Значения, которые ассоциированы с этими элементами (например, /noun/ (существительное), /verb/ (глагол), /feminine/ (женский), /plural/ (множественное) и т.д.), также являются категориями данных согласно настоящему стандарту, но значения данного типа не рассматриваются в качестве концепций элементов данных в стандартах серии ИСО/МЭК 11179.

Примечание 1 - Категория данных – это элементарный дескриптор в лингвистической структуре или схеме аннотации.

Примечание 2 - Категория данных почти соответствует, но не идентична концепции элемента данных в стандартах серии ИСО/МЭК 11179.

Примечание 3 - В текстовых вставках, например используемых в настоящем стандарте, имена категорий данных ограничены символам (косая черта) и набраны курсивом. В некоторых реализациях вместо разделения пробелами составных слов в имени категории данных применяется слитное написание этих слов с использованием букв смешанного регистра ("camel case").

3.1.4 **концепция элемента данных** (data element concept): Концепция, имеющая определение, обозначение и концептуальную область, не зависящую от какого-либо конкретного представления.

[ИСО/МЭК 11179-1:2004, 3.3.9]

3.1.5 **концептуальная область** (conceptual domain): Совокупность допустимых интерпретаций значений.

Примечание 1 – Адаптированное определение из ИСО/МЭК 11179-1:2004.

Примечание 2 – Интерпретации значений концептуальной области могут быть перечислены, уточнены дальнейшими ограничениями или выражены путем описания. Например, категория данных /term/ описывается своим определением и поэтому не может содержать, в частности, контекстную или грамматическую информацию, но перечисление всех значений, ассоциируемых с этой категорией данных, невозможно.

3.1.6 **область значений** (value domain): Совокупность допустимых значений.

[ИСО/МЭК 11179-1:2004, 3.3.38]

3.1.7 **сложная категория данных** (complex data category): Категория данных, имеющая концептуальную область.

3.1.8 **открытая категория данных** (open data category): Сложная категория данных, концептуальная область которой не ограничена перечисленным набором значений.

3.1.9 **открытая концептуальная область** (open conceptual domain): Концептуальная область, ассоциируемая с открытой категорией данных.

3.1.10 **ограниченная категория данных** (constrained data category): Сложная категория данных, концептуальная область которой не представлена перечисленным набором, а сужена ограничением языка или языков на базе конкретной схемы.

3.1.11 **ограниченная концептуальная область** (constrained conceptual domain): Концептуальная область, ассоциируемая с ограниченной категорией данных.

3.1.12 **простая категория данных** (simple data category): Категория данных, не имеющая концептуальной области.

3.1.13 **замкнутая категория данных** (closed data category): Сложная категория данных, концептуальная область которой ограничена набором перечисленных простых категорий данных, составляющих ее область значений.

3.1.14 **замкнутая концептуальная область** (closed conceptual domain): Концептуальная область, ассоциируемая с замкнутой категорией данных.

3.1.15 **схема аннотации** (annotation scheme): Совокупность дескрипторов, их синтаксиса, семантики и условий использования, предназначенная для описаний или интерпретации языкового ресурса

Примечание - TEI ODD (документ "все в одном") – пример схемы аннотации.

3.2 Реестр категорий данных

3.2.1 **реестр категорий данных** (Data Category Registry; DCR): Совокупность категорий данных, используемая в области языковых ресурсов в качестве базы определений лингвистических схем аннотации или любых других форматов.

3.2.2 **спецификация категории данных** (data category specification): Совокупность признаков, полностью описывающих данную концепцию элемента данных.

Примечание- Сокращение "DCS" относится к выборке категорий данных, и его не следует путать со спецификацией категории данных.

3.2.3 **выборка категорий данных** (Data Category Selection; DCS): Набор категорий данных, выбранных из реестра DCR.

Примечание 1 - В выборку DCS могут входить категории данных, используемые в тематической области либо в конкретном приложении или проекте. В последнем случае в DCS могут входить категории данных из нескольких тематических областей.

Примечание 2 - Выборка DCS может быть выражена простым списком категорий данных либо представлена в форме, содержащей все соответствующие спецификации категорий данных и, следовательно, полный набор ограничений, ассоциируемых с DCS. Кроме того, она может быть выражена в обозначениях таких схем, как XML W3C или Relax NG, также включающих список категорий данных и ограничения на эти категории.

3.3 Компоненты спецификации категории данных

3.3.1 **модель данных DCR** (DCR data model): Логическое представление структуры данных и зависимостей в реестре DCR.

Примечание 1 - Модель данных DCR представляется в виде диаграммы класса UML.

Примечание 2 - Приведенное определение основано на ИСО/МЭК 11179-1:2004, в котором «модель данных» определяется как «графическое и/или лексикографическое представление данных с определением их свойств, структуры и взаимосвязей».

3.3.2 **глобальная информация** (Global Information; GI): Техническая или административная информация, применимая ко всей совокупности данных.

[ИСО 16642:2003, п.3.7]

Пример - Название совокупности данных или список ее редакций.

3.3.3 **раздел административной информации** (administration information section): Класс в спецификации категории данных, относящийся к процедурам представления на рассмотрение, регистрации, голосования и утверждения, которые выполняются для спецификаций категории данных, представляемых для включения в реестр DCR и ведения в этом реестре.

3.3.4 **группа регистрации** (registration group): Класс, ассоциируемый с разделом административной информации и содержащий сведения об Органе регистрации (RA), который отвечает за объект администрирования.

3.3.5 **группа представления на рассмотрение** (submission group): Класс, ассоциируемый с разделом административной информации и содержащий сведения о лицах или группах, которые отвечают за представление объекта администрирования на рассмотрение.

3.3.6 **группа принятия решений** (decision group): Класс, ассоциируемый с разделом административной информации и содержащий сведения о процедурах проверки и голосования, связанных с объектом администрирования.

3.3.7 **группа ведения реестра** (stewardship group) Класс, ассоциируемый с разделом административной информации и содержащий сведения о лице или группе лиц, которые отвечают за ведение объекта администрирования в реестре.

3.3.8 **раздел описания** (description section): Класс, относящийся к имени категории данных и к концепции элемента данных, которая документирована в спецификации категории данных.

Примечание - Определения, пояснения и замечания – примеры информации, включаемой в класс описания спецификации категории данных.

3.3.9 **имя элемента данных** (data element name): Класс в спецификации категории данных, в котором перечислены и распределены по категориям допустимые имена, которые можно ассоциировать с категорией данных.

3.3.10 **языковой раздел** (language section): Класс в спецификации категории данных, предоставляющий эквиваленты на рабочем языке для имен категорий данных и других описаний, включенных в спецификацию категории данных.

3.3.11 **лингвистический раздел** (linguistic section): Класс в спецификации категории данных, ограничивающий концептуальную область данного объектного языка.

3.3.12 **рабочий язык** (working language): Язык, используемый для описания объектов.

[ИСО 16642:2003]

3.3.13 **объектный язык** (object language): Описываемый язык. [ИСО 16642:2003]

3.3.14 **раздел имен** (name section): Класс в языковом разделе, содержащий список вариантов имен категории данных, которая описывается в спецификации категории данных.

Примечание - Вариантами имен могут быть эквивалентные имена на другом языке либо имена на том же языке, которые можно использовать в связанных дисциплинах или рабочих средах.

3.4 Ведение реестра DCR

3.4.1 **Совет по администрированию реестра категорий данных** (Data Category Registry Board; DCR Board; DCRB): Группа экспертов, которые назначаются членами–участниками Технического комитета и отвечают за поддержание нужного состава и обеспечение согласования реестра категорий данных.

Примечание - Совет DCRB имеет статус Группы валидации (VT) в соответствии с Приложением ST Дополнения ИСО к Директивам ИСО/МЭК [9].

3.4.2 **Орган регистрации** (Registration Authority; RA): Организация, уполномоченная регистрировать элементы данных и/или другие информационные объекты и содержать их в хранилище.

Примечание - Обычно эти информационные объекты включают коды, например языковые коды стандартов серии ИСО 639, категории данных и другие общедоступные идентификаторы. Деятельность Органа регистрации определяется международными стандартами, но само хранилище обычно не описывается в опубликованных стандартах.

3.4.3 **тематическая область** (thematic domain): Класс приложений, сходных по структурам данных, которыми они должны манипулировать.

Пример – Терминология, лексикография, морфосинтаксическая аннотация.

3.4.4 **Группа обслуживания тематической области** (thematic domain group; TDG): Группа экспертов, которые отвечают за выбор и сопровождение категорий данных, относящихся к тематической области.

Примечание - Группа обслуживания тематических областей имеет статус Группы обслуживания (MT) в соответствии с Приложением ST Дополнения ИСО к Директивам ИСО/МЭК [9].

3.4.5 **профиль** (тематической области) (thematic domain profile; profile): Представление спецификации категории данных тематической области, с которым ассоциирована категория данных.

Примечание - У категории данных может быть несколько профилей тематических областей, если она используется в разных тематических областях. До тех пор, пока спецификация категории данных не будет назначена Группе TDG, для профиля устанавливается значение -private- ("частная разработка").

3.5 Роли

3.5.1 **председатель Совета по администрированию реестра категорий данных** (chair of the Data Category Registry Board; chair of the DCR Board; chair of the DCRB): Лицо, назначаемое на пленарном заседании технического комитета ответственным за управление деятельностью Совета DCRB.

3.5.2 **руководитель Группы обслуживания тематической области** (chair of a thematic domain group; TDG chair): Лицо, назначаемое подкомитетом технического комитета, связанным с TDG, которое отвечает за управление работой TDG.

3.5.3 **эксперт** (expert): Лицо, имеющее специальные знания, опыт или заинтересованность иного характера, которое регистрируется для участия в работе DCR.

3.5.4 **эксперт, выносящий решение** (judge): Эксперт, назначаемый председателем Группы обслуживания тематической области для участия в процедуре утверждения любой конкретной

спецификации категории данных или выборки категорий данных, представленной для стандартизации.

3.6 Обмен данными

3.6.1 **формат обмена категориями данных** (Data Category Interchange Format; DCIF): Формат экспорта категорий данных, объединенных в выборку категорий данных, который разработан для удобства использования этих категорий во внешних приложениях.

3.6.2 **снимок текущего состояния** (snapshot): Состояние информационного ресурса, зарегистрированное в определенный момент времени.

Примечание - Информационные ресурсы часто архивируются в виде снимков текущего состояния, которые позже могут быть отождествлены с разными версиями ресурса.

3.6.3 **неизменный идентификатор** (persistent identifier; PID): Уникальный унифицированный идентификатор ресурса (URI), гарантирующий постоянный доступ к цифровому объекту благодаря независимости этого доступа от физического местонахождения или от текущего владельца объекта.

4 Роль категорий данных в управлении языковыми ресурсами

4.1 Обзор

Спецификации категорий данных описывают отдельные информационные блоки, определяющие схему сбора или аннотации данных для конкретного языкового ресурса. Каждая спецификация задает формальное представление категории данных и включает конкретные признаки, описывающие эту категорию (например, ее имя, определение, примеры, комментарии и т.д.). Кроме того, спецификация предоставляет контекст для ее создания и ведения в

реестре DCR. Группы категорий данных, которые выделены в качестве подмножеств их глобального набора, составляющего реестр DCR, образуют выборки категорий данных (DCS). Как указано в ИСО 16642 *Структура терминологической разметки (TMF)*, в DCS наряду с моделью данных должны быть определены различные ограничения, которые применимы к данным информационным структурам или форматам обмена, специфическим для тематической области или приложения.

На рис. 1 показаны возможные варианты использования DCS. В зависимости от приложения выборка DCS может быть просто списком категорий данных с обратной ссылкой на полные спецификации в DCR либо она может быть представлена полным поднабором или даже расширенным набором DCR, состоящим из такого списка с добавленными определениями и ограничениями, связанными с конкретными спецификациями категорий данных.

В более отдаленной перспективе в формальной модели представления категорий данных должен учитываться тот факт, что кроме использования в компьютерах спецификация категории данных может быть предназначена для использования вручную людьми. Например, спецификации такого типа могут составлять ядро DCS, публикуемое в печатном издании, предоставляемое в виде электронного ресурса или определяемое в качестве поднабора DCR технического комитета. Как правило, разработчики конкретного языка разметки или системы управления данными выполняют запросы в DCR с целью создания выборок DCS для своих приложений путем подбора подмножества категорий данных из глобального реестра DCR. В конечном счете точное описание категорий данных, используемых в определенном наборе данных со ссылкой на DCR, позволяет оперативно оценить совместимость этого набора с любым

другим компьютерным приложением и, следовательно, может служить метаданными для указанного набора.

На рис. 1, показана также выборка DCS, в частности, конкретного набора категорий данных из глобального реестра DCR с целью использования в данной тематической области в контексте языковых ресурсов или в отдельном приложении. На схеме приведен пример различных ролей DCS в процессе определения и использования любой схемы лингвистической аннотации. С этой точки зрения выборка DCS главным образом предназначена для пополнения спецификации схемы аннотации DCR, а также модели данных, выражающей общую организацию DCR. Для такой выборки гарантирована определенная степень функциональной совместимости между двумя или большим числом структур данных благодаря упрощению сравнения выбранных категорий данных и накладываемых на них ограничений. Например, это относится к узловым элементам модели данных DCR, соответствующим допустимым вхождениям каждой категории в отдельные структуры данных, в частности, в конкретных приложениях или схемах аннотации. При такой постановке задачи выборка DCS для каждой из рассматриваемых структур может быть выражена, например, в терминах схем Relax NG [12] или W3C XML [11], и для трансляции соответствующих данных из формата обмена категориями данных (DCIF) в альтернативные форматы могут использоваться фильтры XSL (см. 7.9).

Кроме того, DCS можно рассматривать как документированный источник для применяемой схемы лингвистической аннотации. Так как в DCS содержится список всех категорий данных, которые могут использоваться вместе со схемой аннотации, то, вероятно, это лучший источник информации для потенциальных пользователей или

разработчиков, которым необходимо знать, насколько конкретная категория данных соответствует их потребностям.

DCS можно прикрепить к любому процессу передачи данных (или добавить там ссылку на DCS), чтобы у получателя данных была вся информация, необходимая для интерпретации передаваемого содержания. В частности, при этом возможно выражение лингвистических данных в разного рода XML-представлениях для передачи или приема этих данных наиболее понятным способом.

Рисунок 1 – Роль выборок категорий данных в контексте определения схем лингвистической аннотации

4.2 Различные выборки категорий данных (DCS)

На рис. 2 показана взаимосвязь между отдельными спецификациями категорий данных, реестром DCR и любой из возможных выборок DCS, являющейся подмножеством DCR. Внешней окружностью ограничена вся совокупность спецификаций категорий данных в реестре DCR, а внутренними окружностями меньшего

размера – выборки DCS, являющиеся подмножествами DCR. Самые мелкие фигуры с различной плотностью фона соответствуют отдельным спецификациям категорий данных, каждая из которых описывает заданную концепцию категории данных с помощью признаков категорий данных, которые установлены для DCR со ссылкой на признаки, определенные в ИСО/МЭК 11179. Например, одна небольшая фигура может представлять спецификацию категории данных для категории данных */term/ (термин)*.

Некоторые категории данных, включенные DCR, относятся лишь к одной тематической области более широкого поля терминологии, других языковых ресурсов и ресурсов содержания. Например, категория */conceptIdentifier/ (идентификатор понятия)*, вероятно, уникальна для терминологических ресурсов (хотя это и не обязательно), а категория */senseNumber/ (номер значения)*, возможно, характерна для лексикографических ресурсов. С другой стороны, многие категории данных, часто жестко фиксированной лингвистической природы, например, */partOfSpeech/ (часть речи)*, */grammaticalGender/ (грамматический род)*, */grammaticalNumber/ (грамматическое число)* и т.д., являются общими для самых разнообразных ресурсов. Разумеется, назначение таких категорий не всегда одинаково в разных тематических областях, тем не менее они представляют по существу то же самое понятие в различных видах ресурсов. Таким образом, все собственные категории данных каждой тематической области должны быть внесены в глобальный реестр DCR в виде спецификаций категорий данных, а все категории данных, разделяемые ей с другими видами ресурсов, должны быть в ней идентифицированы. Подмножество категорий данных и их спецификаций, используемое в тематической области, должно составлять специфическую для этой области выборку DCS из реестра DCR.

Как указывалось выше, такие подмножества DCS представлены на рисунке окружностями. Для конкретного приложения или среды коллективной работы можно затем составить подмножество из выборок DCS одной или нескольких тематических областей. Восьмиугольник на рис. 2 – пример подмножества для конкретного приложения. Это подмножество целиком содержится в выборке DCS для терминологических категорий данных и предназначено для терминологического приложения, хотя некоторые содержащиеся в нем спецификации категорий данных являются общими для языковых ресурсов разного рода. В связи с этим следует отметить, что некоторые приложения действительно могут заимствовать категории данных из DCS нескольких тематических областей. Кроме того, если входящие в DCS категории в настоящее время не являются частью DCR, теоретически возможно, что эта выборка частично будет надмножеством категорий. В таких случаях разработчикам и пользователям рекомендуется регистрировать новые категории данных в DCR.

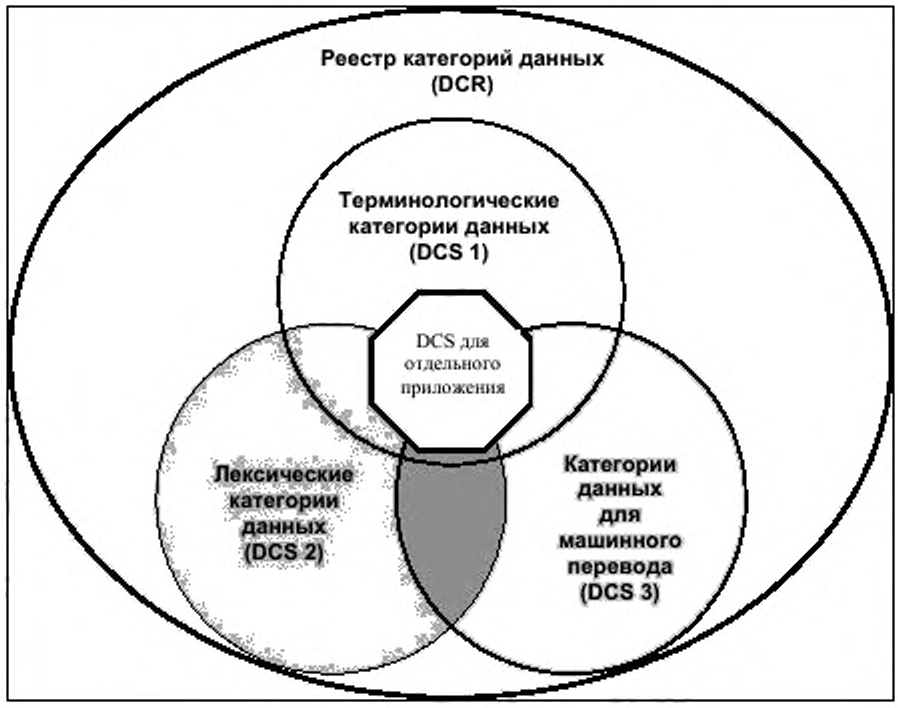


Рисунок 2– DCS для конкретного приложения в контексте различных DCS Группы TDG

5 Требования к реализации реестра DCR для языковых ресурсов

В данном разделе описаны основные требования к реестру DCR, необходимые для поддержки мероприятий по стандартизации, выполняемых техническим комитетом.

Реестр DCR технического комитета должен:

— быть справочным хранилищем категорий данных и связанной информации для всех существующих и будущих стандартов технического комитета, относящихся к моделированию данных или обмену данными;

— быть бесплатно доступным в интерактивном режиме (online);

— обеспечивать регистрацию существующих практических схем путем включения каждой категории данных и способа, которым она реализована в конкретных проектах или инициативах. Сюда может входить регистрация различных типов кодировок, от основных кодов (например, 'f' для термина «женский» в морфосинтаксических описаниях EAGLES [15]) до действительных представлений XML;

— предоставлять имена и определения на различных языках;

— описывать использование каждой категории данных на разнообразных объектных и рабочих языках. Сюда могут входить особые определения (например, если прикладная область категории данных слегка отличается), определенные замечания по использованию, примеры или списки значений (например, область значений категории */gender/ (род)* есть {*/masculine/*, */feminine/*} во французском и {*/masculine/*, */feminine/*, */neuter/*} в немецком языке);

— описывать использование категорий данных в разнообразных ресурсах, если это необходимо;

— связывать административную информацию с каждой категорией данных, что позволит отслеживать представление, одобрение или пересмотр категорий данных;

— связывать каждую категорию данных с одним или несколькими профилями, соответствующими тематическим областям, к которым относится эта категория (например, */partOfSpeech/ (часть речи)* относится как к терминологии, так и к лексикографии);

— предоставлять Группе обслуживания тематической области технического комитета механизм для представления на рассмотрение групп категорий данных, относящихся к сфере ее деятельности;

— предоставлять частные рабочие области, которые могут использоваться отдельными лицами или рабочими группами, не входящими в технический комитет, для создания или загрузки

собственных спецификаций и выборок категорий данных, для их публикации и передачи в общее пользование и, по желанию, представления категорий данных на регистрацию и утверждение;

— регулярно обновляться путем включения, в соответствии с установленными правилами предложений, полученных от экспертов в данной области;

— отвечать основным принципам, установленным для стандартов серии ИСО/МЭК 11179;

— быть всегда доступным во всем мире благодаря распределению копий на сайтах-зеркалах;

— предусматривать долгосрочные идентификаторы благодаря системе твердой привязки ссылок на отдельные спецификации данных;

— поддерживать безопасные передовые практические методы архивирования;

— предусматривать предоставление техническому комитету периодических снимков текущего состояния для категорий данных, которые были утверждены согласно процедурам, описанным в Приложении ST Дополнения ИСО к Директивам ИСО/МЭК и в разделе 8 настоящего стандарта, с учетом того, что технический комитет должен обеспечивать доступность данного подмножества реестра DCR с использованием стандартных практических методов и с сохранением доступности этих элементов в самой среде реестра DCR.

6 Орган регистрации реестра DCR ИСО/ТК 37

Как определено в Директивах ИСО/МЭК, реестр DCR технического комитета должен быть реализован под руководством Органа регистрации, обязанностью которого является регистрация

категорий данных в соответствии с правилами, установленными в настоящем стандарте.

Техническим комитетом должен быть назначен Орган регистрации реестра DCR.

Орган регистрации должен организовать ведение реестра DCR в качестве веб-службы. Все функции для пользователей и пользовательская поддержка должны быть доступны на определенном web-сайте.

7 Представление категорий данных, используемых в языковых ресурсах

7.1 Введение

В данном разделе описывается модель данных реестра DCR для технического комитета. Она формулируется на унифицированном языке моделирования (UML), по необходимости расширяемом дополнительными ограничениями на языке объектных ограничений (OCL). Полная схема модели данных показана на рис. 3, разбитом на три части на рис. 4, 5 и 6, связующим классом между которыми служит категория данных.

Реестр DCR должен включать два главных класса:

— класс глобальной информации (GI);

— класс для одной или нескольких спецификаций категорий данных (DC).

Каждая спецификация категории данных должна включать два обязательных класса:

— класс, предназначенный для администрирования и идентификации категории данных (раздел административной информации – Administration Information);

— класс, предназначенный для документирования категории данных на одном или нескольких рабочих языках, возможно, с помощью нескольких имен в согласовании с конкретной базой данных, форматом или приложением (раздел описания – Description).

В зависимости от точного типа категории данных возможны дальнейшие расширения ее спецификации. К таким расширениям относятся следующие классы:

— один или несколько классов, описывающих концептуальную область категории данных (класс Conceptual Domain и его подклассы);

— один или несколько классов, описывающих концептуальную область и/или использование категории данных в контексте конкретного объектного языка (раздел Linguistic и его подклассы).

Другие классы, связанные с указанными главными классами, подробнее описаны в других подразделах.

7.2 Класс глобальной информации

В реестре DCR должны быть следующие элементы, относящиеся к глобальной информации (класс Global Information):

— название и полный набор контактных данных Органа регистрации (точный адрес, номера телефонов и факсов, контактные адреса e-mail);

— дата назначения Органа регистрации (далее – RA) и другие важные даты, касающиеся изменений в этом Органе;

— имя главного ответственного за администрирование RA;

— хронологическая информация о предыдущих RA, если обязанности RA передавались другой организации;

— краткие наименования или сокращения, используемые в реестре DCR при ссылках на действующий RA и все предыдущие RA, которые могли отвечать за реестр;

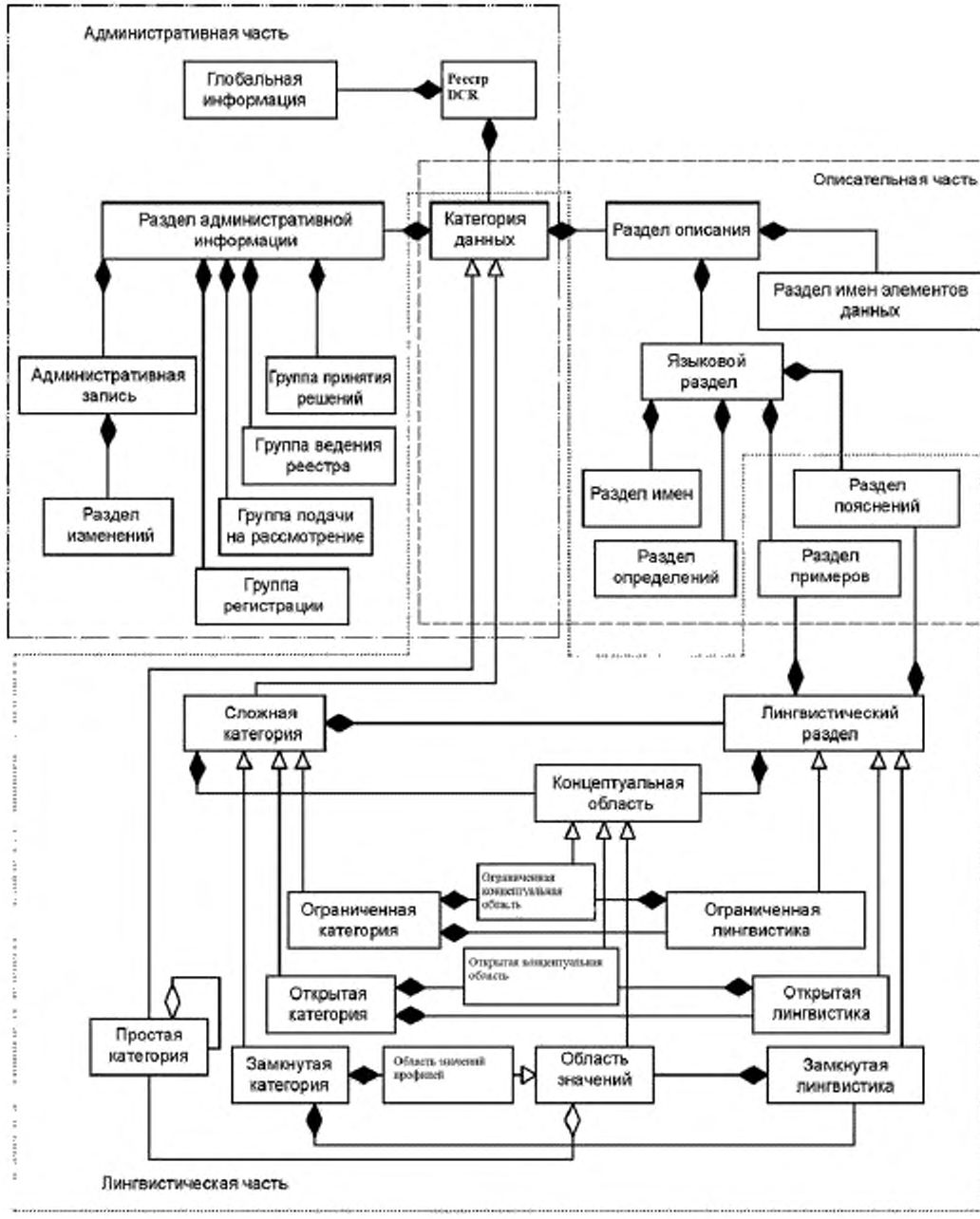
— имена и аффилированность членов Совета DCRB;

— сведения о соглашении между ИСО и RA;

— заявление о миссии DCR;

— заявление об юридической ответственности и юридических ограничениях.

В настоящем стандарте не накладываются какие-либо дополнительные специальные ограничения на данный класс. Ответственность за предоставление надлежащей технической информации возлагается на RA и DCRB.

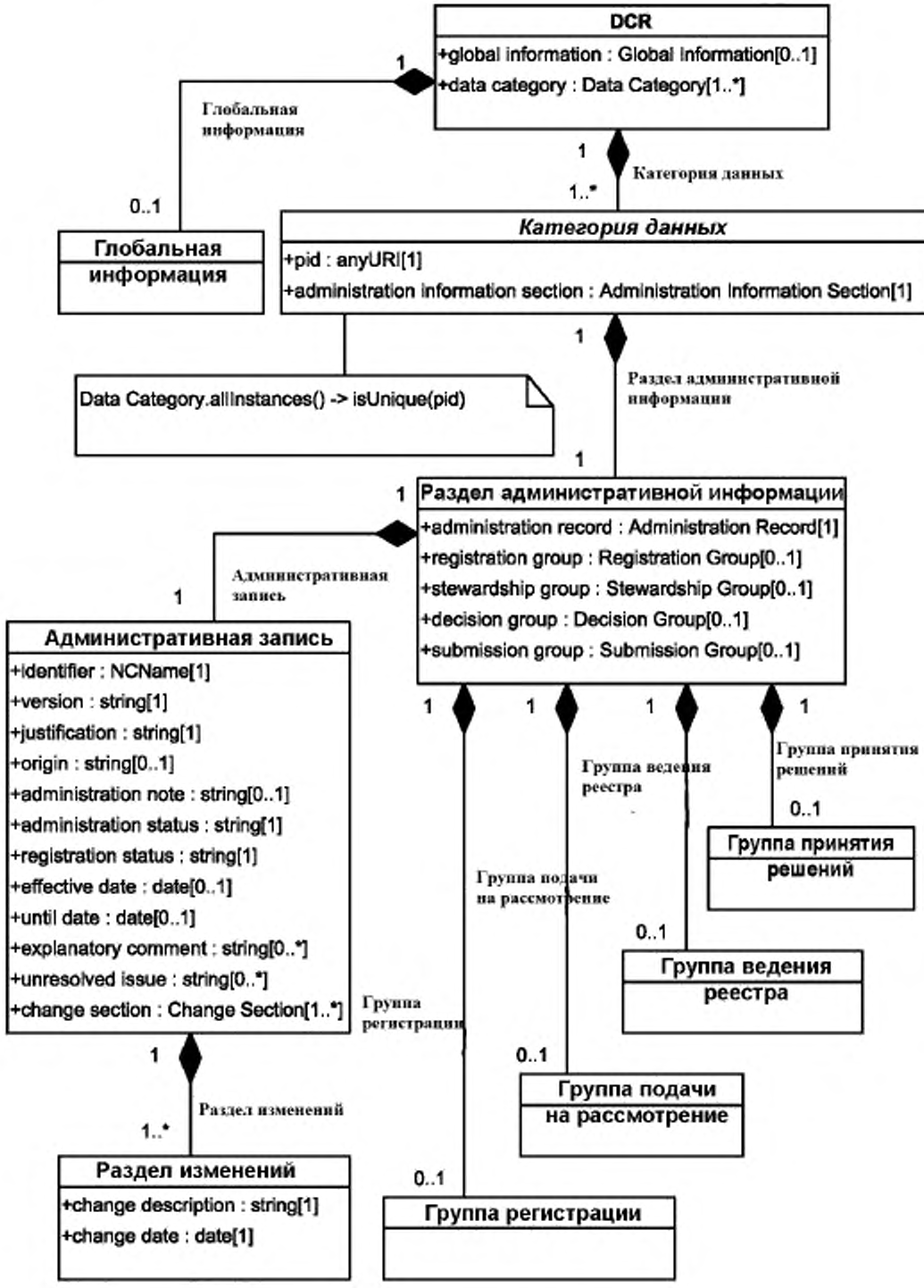


△ – отношение наследования между подклассом и его базовым классом;

♦ – отношение композиции между двумя классами;

◊ – отношение агрегации между двумя классами.

Рисунок 3 – Обзорная схема модели данных реестра DCR



△– отношение наследования между подклассом и его базовым классом;

♦ – отношение композиции между двумя классами;

◊ – отношение агрегации между двумя классами.

Значения кратности: 0..1 = 0 или 1; 0..* = 0 или более; 1 = 1 и только 1; 1..* = 1 или большему числу случаев.

Рисунок 4 –Административная часть модели данных реестра DCR

7.3 Классы категории данных

7.3.1 Подтипы категорий данных

В реестре DCR содержатся категории данных двух следующих примитивных подтипов:

a) сложные категории данных, например, */term/ (термин)* или */grammaticalGender/ (грамматический род)*;

b) простые категории данных, например, */masculine/ (мужской), /feminine/ (женский)* и т.д.

Эти два базовых типа категорий данных представлены в модели DCR двумя подклассами категории данных.

В категорию данных входит следующий признак:

+pid [1]: неизменный идентификатор (PID) категории данных. Для стандартизованных категорий данных RA должен стремиться к обеспечению разрешимости этих идентификаторов до категорий, даже если сами категории исключаются из рекомендованных. При ссылке на категорию данных в библиографическом или техническом контексте следует использовать **+pid (**см. также 7.8).

7.3.2 Сложная категория данных

Категория данных этого типа может иметь концептуальную область. В модели данных DCR поддерживаются концептуальные области трех типов: открытые (open), ограниченные (constrained) и замкнутые (closed). Для отдельных объектных языков концептуальная область может быть ограничена еще в большей мере. Моделирование областей и ограничений описано в 7.6 и 7.7.

7.3.3 Простая категория данных

Простые категории данных служат для представления значений, ассоциируемых с областями значений сложных категорий данных. В отличие от сложных категорий данных эти категории могут войти в иерархию значений после ассоциации ‘is a’ (‘является’). Например, пользователь может объявить, что */properNoun/ (имя собственное)* является */noun/ (именем существительным)*. Данный тип ассоциации недопустим для сложных категорий данных, так как описание расширенных концептуальных иерархий выходит за рамки реестра DCR.

+is a [0..1]: признак, используемый для указания на простую категорию данных более общего характера, с которой ассоциируется рассматриваемая простая категория данных.

Данная ассоциация ограничена в том смысле, что две простых категории данных, для которых она задается, должны быть включены

как минимум в один совпадающий профиль. Кроме того, граф, который строится в результате ассоциаций, должен быть ациклическим, т.е. простая категория данных не может явно или неявно являться обобщенной простой категорией данных для себя самой.

7.4 Раздел административной информации

7.4.1 Связанные классы

Раздел административной информации может быть разбит на пять связанных классов:

— Administration Record (административная запись), в которой объединены данные, связанные с глобальным управлением объектом администрирования;

— Registration Group (группа регистрации) с информацией, относящейся к RA, а также к DCRB;

— Submission Group (группа представления на рассмотрение) с информацией, которая связана с субъектом, представившим категорию данных для ее включения в реестр. Этим субъектом может быть либо группа обслуживания тематической области, избравшая категорию данных, и/или эксперт либо группа экспертов, инициировавших представление на рассмотрение;

— Stewardship Group (группа ведения реестра) с информацией о группе обслуживания тематической области, отвечающей за ведение объекта администрирования и выступающей в качестве группы обслуживания согласно терминологии Приложения ST из Дополнения ИСО к Директивам ИСО/МЭК [9];

— Decision Group (группа принятия решений) с информацией, которая связана с процедурами принятия решений, используемыми

для оценки пригодности категории данных соответствующей группой обслуживания тематической области и для валидации этой оценки Советом DCRB.

7.4.2 Информация, представляемая в разделе административной записи

В разделе административной записи Administration Record содержится информация об идентификации и обслуживании объекта администрирования. С этим разделом ассоциированы следующие признаки:

— **+identifier** [1; ИСО/МЭК 11179-3] (идентификатор): условная строка для ссылок на категорию данных.

Согласно ИСО/МЭК 11179-3 идентификатор представляется в виде буквенно-цифровой символьной строки. Для удобочитаемости могут использоваться последовательности английских слов, которые отражают смысл идентификатора (например, */term/ (термин)*, */normativeAuthorization/ (нормативная авторизация)*, */preferredTerm/ (предпочтительный термин)*), но это соглашение не должно препятствовать использованию дополнительных имен категорий данных на английском или каком-либо ином языке. Следует также отметить, что элементы идентификатора должны ограничиваться в нем буквами разного регистра.

Чтобы идентификатор категории данных можно было использовать в словарях XML, он должен быть действительной локальной частью классифицированного имени, как это определено для XML документов, соответствующих рекомендации по пространствам имен XML.

При ссылке на категорию данных в библиографическом или техническом контексте следует использовать **+pid**, а не **+identifier** (см. также 7.8);

— **+version** [1] (версия): используется для уточнения **+identifier** и указания версии категории данных;

— **+administration note** [0..1; ИСО/МЭК 11179-3] (замечание по администрированию): любое общее замечание об объекте администрирования;

— **+administration status** [1] (административный статус): наименование статуса в процессе администрирования для обработки запросов на регистрацию под руководством DCRB. Для указания административного статуса могут использоваться следующие значения:

-private- (частная разработка): спецификация категории данных используется только в частной рабочей области эксперта или используется совместно членами закрытой группы, но не была (и, возможно, никогда не будет) представлена для стандартизации,

-submission- (подано на рассмотрение): спецификация категории данных была подана отдельным экспертом или группой экспертов (согласно информации о них в разделе представления на рассмотрение) на рассмотрение данной группе TDG (указанной в разделе группы ведения реестра), в результате чего был инициирован процесс выбора и стандартизации категории данных, который иллюстрируется на рис. 9,

-pre-evaluation- (предварительная оценка): руководители DCRB и TDG утвердили предложение и инициировали для него этап оценки, передав группе обслуживания тематической области,

-evaluation- (оценка): возможность принятия спецификации категории данных оценивается группой обслуживания тематической области,

-rejected-TDG- (отвергнуто TDG): спецификация категории данных была отвергнута TDG,

-accepted-TDG- (принято TDG): спецификация категории данных была принята TDG, что отражено в разделе Resolution of Acceptance (решение о принятии) (см. рисунок 9),

-pre-validation- (предварительная валидация): подготовка к валидации председателем DCRB,

-validation- (валидация): спецификация категории данных была утверждена TDG, подготовлена руководителями TDG и DCRB, а затем направлена в DCRB для окончательной валидации, т. е. для рассмотрения и утверждения,

-accepted- (принято DCRB): спецификация прошла валидацию и спецификация была принята DCRB для включения в реестр DCR,

-rejected-DCRB- (отвергнуто DCRB): спецификация категории данных была отвергнута DCRB;

— **+registration status** [1; ИСО/МЭК 11179-3] (статус регистрации): наименование статуса в цикле регистрации объекта администрирования.

Для **+registration status** могут использоваться следующие значения (по ИСО/МЭК 11179-6:2005):

-candidate- (кандидат): спецификация категории данных была предложена для обработки по этапам процедуры регистрации реестра DCR.

Примечание – Значение статуса регистрации спецификации категории остается -candidate-, пока этот административный статус не будет окончательно зафиксирован как -accepted- (и в этом случае статус регистрации становится -standard-).

-standard- (стандартный): Советом DCRB было подтверждено качество спецификации категории данных и то,

что она представляет интерес для широкого круга пользователей в сообществе пользователей реестра DCR;

-deprecated- (не рекомендовано): Советом DCRB было подтверждено, что спецификация категории данных сейчас или в дальнейшем не рекомендована для применения в сообществе пользователей реестра;

-superseded- (заменено): Советом DCRB было подтверждено, что спецификация категории данных более не рекомендована для применения в сообществе пользователей реестра и для предпочтительного применения Советом была указана заменяющая ее спецификация.

— **+effective date** [0..1] (дата вступления в силу): день, в который спецификация категории данных стала/станет доступной пользователям реестра DCR, в формате YYYY-MM-DD согласно ИСО 8601:2004 и ИСО/МЭК 11179-3;

— **+change section** [1..*] (раздел изменений): сведения о том, когда спецификация категории данных была подвергнута последнему изменению, включая сведения о времени создания спецификации категории данных (например, в частной рабочей области эксперта); в реестре DCR должны быть зарегистрированы все изменения.

— **+explanatory comment** [0..*; ИСО/МЭК 11179-3] (пояснительные комментарии): описательные замечания о спецификации категории данных;

— **+origin** [0..1; ИСО/МЭК 11179-3] (происхождение): первоисточник (документ, проект, дисциплина или модель) спецификации категории данных;

— **+justification** [1] (обоснование): краткое описание того, почему категория данных должна быть включена в реестр;

— **+unresolved issue** [0..*; ИСО/МЭК 11179-3] (неразрешенная проблема): вопрос, остающийся открытым в отношении надлежащего документирования спецификации категории данных;

— **+until date** [0..1; ИСО/МЭК 11179-3] (конечная дата): день (в формате YYYY-MM-DD согласно ИСО 8601:2004), в который спецификация категории данных утратит силу в реестре. Устанавливается, когда статус регистрации категории меняется на -deprecated- или -superseded-.

В классе раздела изменений Change Section, используемом в разделе **+change section**, указывается следующая информация:

+change date [1] (дата изменения): день (в формате YYYY-MM-DD согласно ИСО 8601:2004), в который было внесено изменение в спецификацию категории данных;

+change description [1] (описание изменения): описание изменения спецификации категории данных, представленное в произвольной форме (например, “обновлено определение…”).

7.4.3 Информация, представляемая в классе группы регистрации

В настоящем стандарте не накладываются какие-либо явные ограничения на эту компоненту. Решение о предоставлении замечания технического характера с пояснением реализации данной компоненты при ее использовании на практике остается за Органом регистрации.

7.4.4 Информация, представляемая в классе группы представления на рассмотрение

В настоящем стандарте не накладываются какие-либо явные ограничения на эту компоненту. Решение о предоставлении замечания технического характера с пояснением реализации данной

компоненты при ее использовании на практике остается за Органом регистрации.

7.4.5 Информация, представляемая в классе группы ведения реестра

В настоящем стандарте не накладываются какие-либо явные ограничения на эту компоненту. Решение о предоставлении замечания технического характера с пояснением реализации данной компоненты при ее использовании на практике остается за Органом регистрации.

7.4.6 Информация, представляемая в классе группы принятия решений

В настоящем стандарте не накладываются какие-либо явные ограничения на эту компоненту. Решение о предоставлении замечания технического характера с пояснением реализации данной компоненты при ее использовании на практике остается за Органом регистрации.

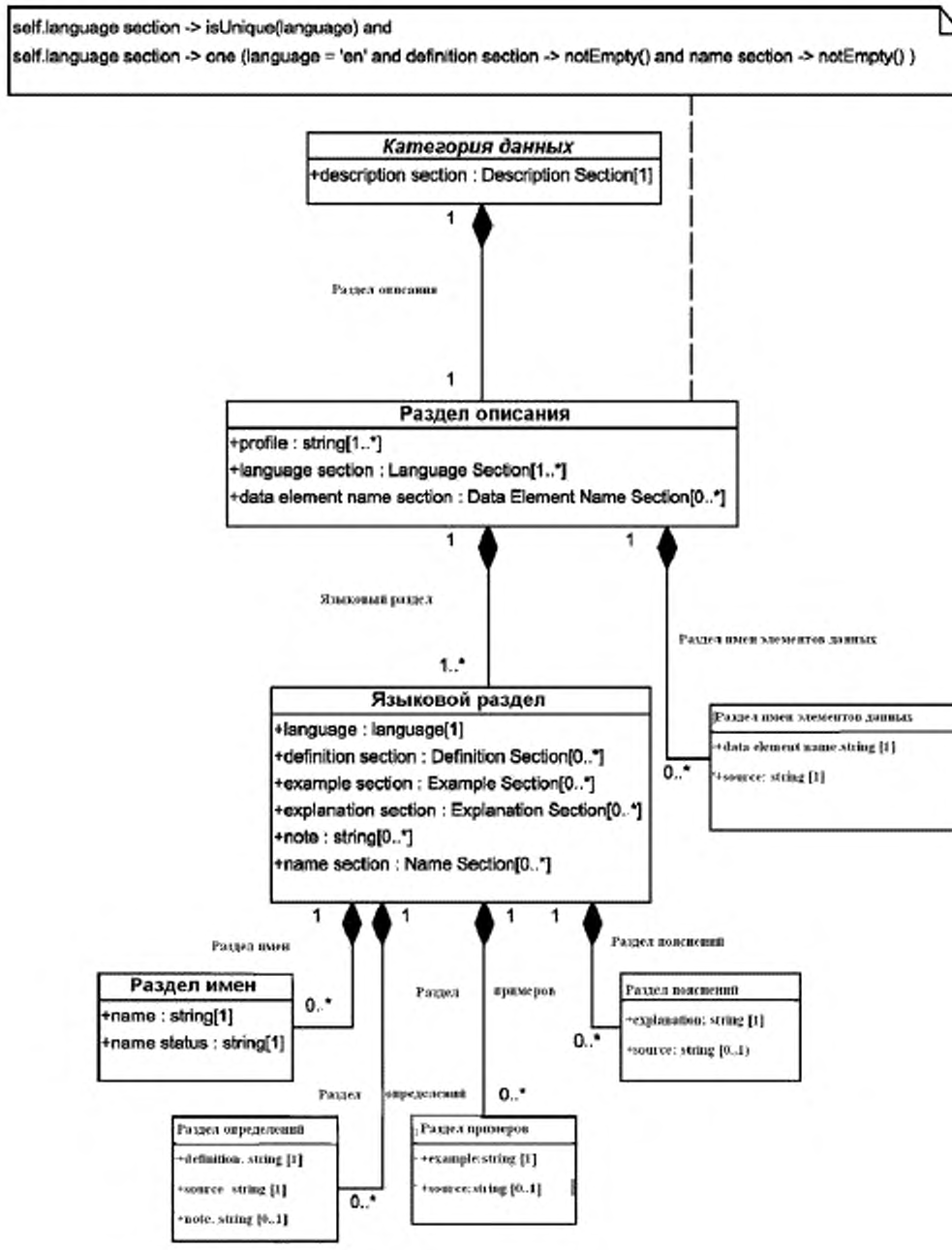
7.5 Документирование категорий данных

7.5.1 Информация, представляемая в разделе описания

Раздел описания спецификации категории данных можно рассматривать в качестве совокупности одного или нескольких языковых разделов и, возможно, нескольких классов имен элементов данных. В языковых разделах приводится документация по категории данных для конкретного рабочего языка. Для каждой категории данных всегда должен быть раздел английского языка, включающий не менее одного раздела определения и одного раздела имен. В

разделах имен элементов данных указываются имена категории данных в одной или нескольких базах данных, в одном или нескольких форматах или в одном или нескольких приложениях. С разделом описания ассоциирован следующий признак:

+profile [1..*] (профиль): используется для соотнесения рассматриваемой спецификации категории данных с одной или несколькими тематическими областями, которыми занимается технический комитет (например, морфосинтаксис, синтаксис, метаданные, описания языков и т.д.). При создании спецификации категории данных значение +profile по умолчанию устанавливается равным -private-, если (или до тех пор, пока) пользователь не выберет один или несколько профилей тематических областей. Для подачи заявки на стандартизацию необходим выбор как минимум одного профиля тематической области, так как за сопровождение стандартизованных спецификаций категорий данных отвечает соответствующая группа TDG.


self.language section -> isUnique(language) and
self.language section -> one (language = 'en' and definition section -> notEmpty() and name section -> notEmpty())
Категория данных
+description section : Description Section[1]
1
Раздел описания
1
Раздел описания
+profile : string[1..*]
+language section : Language Section[1..*]
+data element name section : Data Element Name Section[0..*]
1
1
Языковый раздел
Раздел имен элементов данных
1..*
Языковой раздел
+language : language[1]
+definition section : Definition Section[0..*]
+example section : Example Section[0..*]
+explanation section : Explanation Section[0..*]
+note : string[0..*]
+name section : Name Section[0..*]
0..*
Раздел имен элементов данных
+data element name:string [1]
+source: string [1]
1
1
1
1
Раздел имен
Раздел пояснений
Раздел имен
+name : string[1]
+name status : string[1]
0..*
Раздел
примеров
Раздел пояснений
+explanation: string [1]
+source: string [0..1]
0..*
Раздел
определений
0..*
Раздел определений
+definition. string [1]
+source string [1]
+note. string [0..1]
0..*
Раздел примеров
+example:string [1]
+source:string [0..1]

△ – отношение наследования между подклассом и его базовым классом;

♦ – отношение композиции между двумя классами;

◊ – отношение агрегации между двумя классами.

Значения кратности: 0..1 = 0 или 1; 0..* = 0 или более; 1 = 1 и только 1; 1..* = 1 или большему числу случаев.

Рисунок 5 – Описательная часть модели данных реестра DCR

7.5.2 Информация, представляемая в языковом разделе

В языковом разделе Language Section описывается концепция категории данных в контексте указанного рабочего языка.

— **+language [1]** (язык): рабочий язык. Содержимое **+language** должно соответствовать IETF BCP 47, RFC 5646;

— **+note** [0..*] (замечание): любые дополнительные сведения о категории данных, исключая техническую информацию, которая обычно вошла бы в **+explanation**;

+name section [0..*] (раздел имен): регистрирует возможное имя категории данных на конкретном языке. Этот раздел может повторяться в языковом разделе. С ним связаны следующие описательные признаки:

+name [1] (имя): блок из одного или нескольких слов, используемый для ссылок на категорию данных при использовании соответствующего рабочего языка. Имена, присвоенные категории данных, не должны использоваться для ее идентификации (см. **+identifier**);

+name status [1] (статус имени): указание на доступность и применение. Возможные значения данного признака следующие:

-standardized name- (стандартизованное имя): имя было утверждено государственным, региональным или международным органом стандартизации;

-preferred name- (предпочтительное имя): в случае нескольких имен – имя, определенное как наиболее подходящее либо уполномоченным органом, либо для конкретной среды или приложения;

-admitted name- (допустимое имя): в случае нескольких имен – имя, определенное как приемлемое либо уполномоченным органом, либо для конкретной среды или приложения;

-deprecated name- (нерекомендуемое имя): имя, отвергнутое либо уполномоченным органом, либо для конкретной среды или приложения;

-superseded name- (имя, замененное на другое): имя, утвержденное Советом DCRB как более не рекомендуемое для применения в сообществе пользователей реестра DCR и для которого в качестве более предпочтительного было указано новое имя.

— **+definition section** [0..*] (раздел определения): определение концепции категории данных, которая ассоциирована с данной категорией, приведенное на языке языкового раздела. В разделах английского языка, обязательных для любых спецификаций категорий данных, должны быть определения. Они приводятся в разделе определений, содержащем следующие признаки:

+definition [1] (определение): безусловная формулировка, которая должна быть достаточно общей для ее применимости ко всем тематическим областям и реализациям категории данных;

+source [1] (источник): источник, из которого было взято или адаптировано определение;

+note [0..1] (замечание): любые дополнительные сведения об определении;

— **+explanation section** [0..*] (раздел пояснений); добавочная информация о концепции категории данных. Пояснения приводятся в разделе пояснений, содержащем следующие признаки:

+explanation [1] (пояснение): любые дополнительные сведения о категории данных, которые были бы неуместны в ее определении (например, более точный лингвистический контекст использования категории данных);

+source [0..1] (источник): источник, из которого было взято или адаптировано пояснение;

— **+example section** [0..*] (раздел примеров): пример образца, иллюстрирующий категорию данных. Следует включать только примеры, в которых приводится общая иллюстрация категории данных, но не примеры использования конкретного языка, которые следует документировать в языковом разделе. Примеры приводятся в классе примеров Example Class, содержащем следующие признаки:

+example [1] (пример): пример, иллюстрирующий категорию в целом, а не ее использование для конкретного языка;

+source [0..1] (источник): источник, из которого был взят или адаптирован пример.

7.5.3 Информация, представляемая в разделе имени элемента данных

Раздел имени элемента данных должен использоваться для записи одного имени категории данных, применяемого для конкретной базы данных, формата или приложения. Экземпляры этого раздела могут повторяться в разделе описания для разных имен, используемых в разных приложениях. С каждым разделом ассоциированы следующие атрибуты:

— **+data element name** [1] (имя элемента данных): один идентификатор [слово, блок из нескольких слов или (буквенно) цифровое представление], используемый для ссылок на категорию данных в конкретной базе данных, формате или приложении. Имена элементов данных недопустимо использовать для идентификации категории данных вне конкретной базы данных, в другом формате или приложении (см. **+identifier**);

— **+source** [1] (источник): информация о том, в какой базе данных, каком формате или каком приложении используется имя элемента данных.

7.6 Классы концептуальных областей

7.6.1 Отличия концептуальных областей

В реестре DCR поддерживаются три типа концептуальных областей: открытые, ограниченные и замкнутые. Для открытой концептуальной области нет никаких ограничений на значения, образующие ее, как определено в классе Open Conceptual Domain. Замкнутая концептуальная область состоит из перечисленных допустимых значений, которые выбраны из реестра DCR. Для ограниченной концептуальной области указывается набор допустимых значений, который не может быть выражен в терминах замкнутой концептуальной области, например: *даты всех лет после 1965 г.* Во всех данных классах имеется следующий признак:

+data type [1] (тип данных): тип данных, определенный для XML-схемы W3C этой сложной категории данных, по умолчанию – *string (строка).*

7.6.2 Информация, представляемая в открытой концептуальной области

Для открытой концептуальной области (класс Open Conceptual Domain) допустимы все возможные значения, ассоциируемые с конкретным типом данных.

7.6.3 Информация, представляемая в правиле для концептуальной области

Иногда на лингвистические ресурсы с помощью различных схем накладываются дополнительные ограничения. Правило для концептуальной области (класс Conceptual Domain Rule) позволяет задавать ограничения на возможные значения в концептуальной области для конкретного типа данных на языке правил, подходящем для рассматриваемой схемы. Одно и то же ограничение может быть выражено на нескольких языках. Ответственная группа TDG должна подтвердить эквивалентность на этапе оценки в процессе стандартизации.

— **+rule type** [1] (тип правила): язык, на котором изложено правило, например, язык XML-схем W3C или язык объектных ограничений ;

— **+rule** [1..*]: ограничение, выраженное на языке правил.

7.6.4 Информация, представляемая в области значений

В классе области значений Value Domain перечисляются допустимые значения, представленные простыми категориями данных.

— **+value** [1..*] (значение): ссылка на простую категорию данных, которая описывает один элемент из множества значений, допустимых для сложной категории данных.

Пример - В качестве области значений категории */grammaticalGender/ (грамматический род)* можно было бы задать {*/masculine/*, */feminine/*, */neuter/*} *(мужской, женский и средний)*.

7.6.5 Информация, представляемая в классе области значений профиля

Сложная категория данных может быть ассоциирована с несколькими профилями. Подкласс области значений профиля Profile Value Domain позволяет связать конкретную область значений с конкретным профилем.

— **+profile** [1] (профиль): профиль, с которым ассоциируется данная область значений.

Следует отметить, что ограничение, связанное с замкнутой категорией данных и выраженное на языке объектных ограничений OMG, приводит к следующим обязательным условиям для областей значений профиля:

a) у каждого профиля может быть только одна область значений;
b) у каждого профиля, членом которого является сложная категория данных, должна быть область значений;
c) области значений могут быть только у профилей, членом которых является сложная категория данных;
d) только простые категории данных, ассоциированные с тем же профилем, что и профиль сложной категории данных, могут быть представлены в ее области значений;
e) область значений замкнутой категории данных не может быть расширена лингвистическими разделами.

7.7 Классы лингвистического раздела

Класс лингвистического раздела Linguistic Section используется для задания характеристик сложной категории данных на конкретном

объектном языке. В базовом классе Linguistic Section приводится следующая информация:

— **+language** [1] (язык): описываемый язык (т.е. *объектный язык*). Значение атрибута **+language** должно соответствовать IETF BCP 47, RFC 5646;

— **+conceptual domain** [0..*] (концептуальная область): дополнительные условия, например, дополнительные ограничения или подмножество области значений, задаваемые для концептуальной области, которая объявлена для сложной категории данных, и относящиеся к описанному в лингвистическом разделе объектному языку;

— **+example section** [0..*] (раздел примеров): примеры использования категории данных для выбранного объектного языка. Здесь используется тот же класс Example Section, который был описан для класса Language Section;

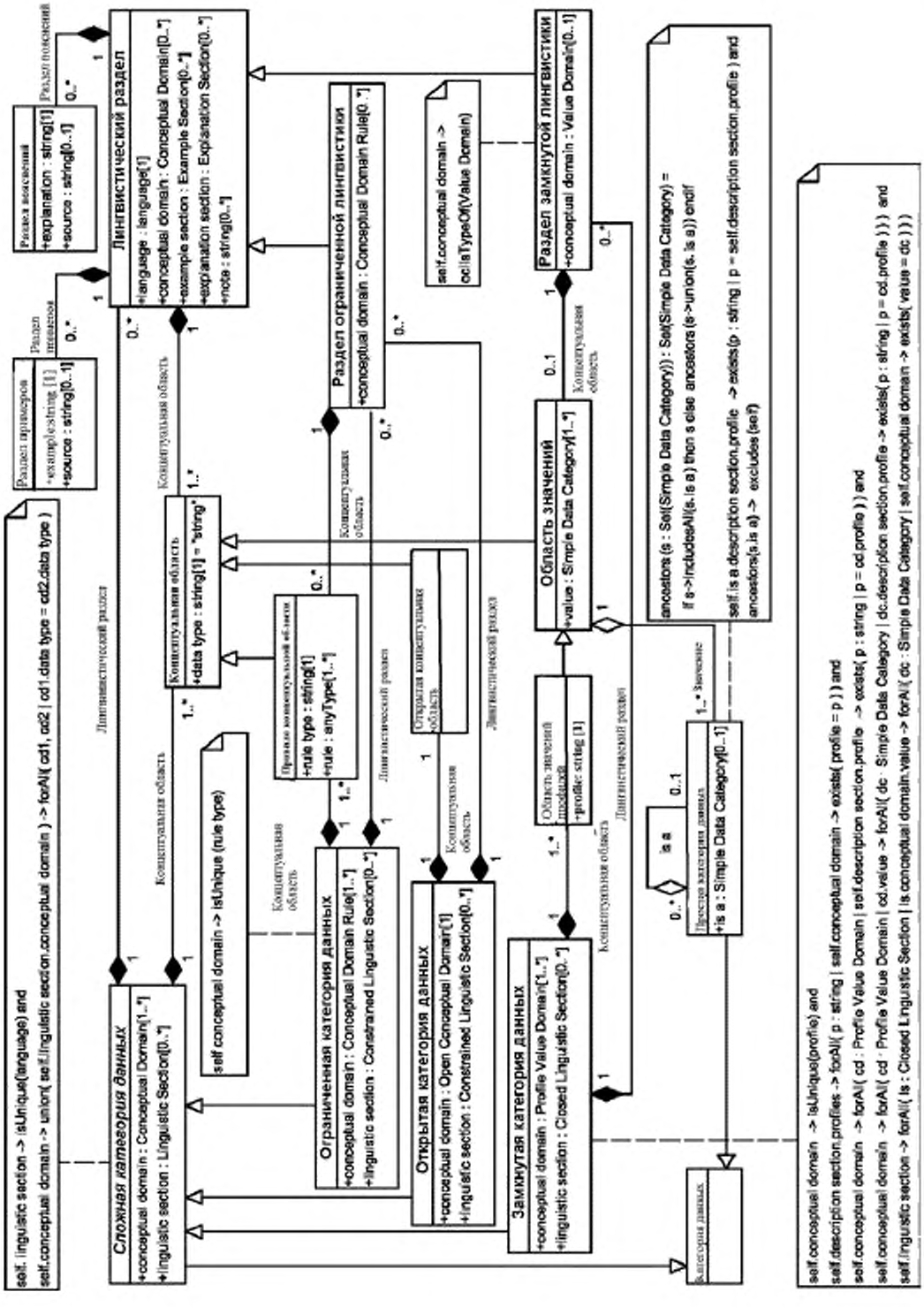
— **+explanation section** [0..*] (раздел пояснений): дополнительные пояснения по использованию категории данных с выбранным объектным языком. Здесь используется тот же класс Explanation Section, который был описан для класса Language Section;

— **+note** [0..*] (замечание): любые дополнительные сведения о категории данных, исключая техническую информацию, которая обычно вошла бы в **+explanation**.

Подклассы класса Linguistic Section позволяют пользователям сужать концептуальные области сложных категорий данных. В разделе ограниченной лингвистики Constrained Linguistic Section можно накладывать зависящие от объектного языка дополнительные ограничения на ограниченную или открытую категорию данных, используя одну или несколько схемно-зависимых областей. Такие ограничения должны выделять подмножества концептуальных

областей сложных категорий данных, то есть расширение множества допустимых значений при этом запрещено.

Область значений замкнутой категории данных может быть сужена в разделе замкнутой лингвистики Closed Linguistic Section. Например, для замкнутой категории данных */grammaticalGender/ (грамматический род)* в таком разделе для французского языка French Closed Linguistic Section область значений сужается до {*/masculine/*, */feminine/*} (мужской, женский). Если для замкнутой категории данных используются области значений профиля, то область значений категории данных для конкретной комбинации профиля и объектного языка определяется пересечением соответствующих областей значений.


self. linguistic section -> isUnique(language) and
self.conceptual domain -> union(self.linguistic section.conceptual domain) -> forAll(cd1, cd2 | cd1.data type = cd2.data type)
Сложная категория данных
+conceptual domain : Conceptual Domain[1..*]
+linguistic section : Linguistic Section[0..*]
Ограниченная категория данных
+conceptual domain : Conceptual Domain Rule[1..*]
+linguistic section : Constrained Linguistic Section[0..*]
Открытая категория данных
+conceptual domain : Open Conceptual Domain[1]
+linguistic section : Constrained Linguistic Section[0..*]
Замкнутая категория данных
+conceptual domain : Profile Value Domain[1..*]
+linguistic section : Closed Linguistic Section[0..*]
Область значений
+value : Simple Data Category[1..*]
Область значений
+profile: string [1]
Простая категория данных
+is a : Simple Data Category[0..1]
Лингвистический раздел
Концептуальная область
Раздел примеров
+example:string [1]
+source : string[0..1]
Раздел пояснений
+explanation : string[1]
+source : string[0..1]
Раздел пояснений
Лингвистический раздел
+language : language[1]
+conceptual domain : Conceptual Domain[0..*]
+example section : Example Section[0..*]
+explanation section : Explanation Section[0..*]
+note : string[0..*]
Раздел ограниченной лингвистики
+conceptual domain : Conceptual Domain Rule[0..*]
Раздел замкнутой лингвистики
+conceptual domain : Value Domain[0..1]
self.conceptual domain -> oclIsTypeOf(Value Domain)
Концептуальная область
+data type : string[1] = "string"
Правило концептуальной области
+rule type : string[1]
+rule : anyType[1..*]
Открытая концептуальная область
Лингвистический раздел
self conceptual domain -> isUnique (rule type)
Концептуальная область
Лингвистический раздел
Концептуальная область
Концептуальная область
Концептуальная область
Концептуальная область
Лингвистический раздел
Концептуальная область
Концептуальная область
Лингвистический раздел
Категория данных
ancestors (s : Set(Simple Data Category)) : Set(Simple Data Category) =
if s->includesAll(s.is a) then s else ancestors(s->union(s.is a)) endif
self.is a.description section.profile -> exists(p : string | p = self.description section.profile) and
ancestors(s.is a) -> excludes (self)
self.conceptual domain -> isUnique(profile) and
self.description section.profiles -> forAll(p : string | self.conceptual domain -> exists(profile = p)) and
self.conceptual domain -> forAll(cd : Profile Value Domain | self.description section.profile -> exists(p : string | p = cd.profile)) and
self.conceptual domain -> forAll(cd : Profile Value Domain | cd.value -> forAll(dc : Simple Data Category | dc.description section.profile -> exists(p : string | p = cd.profile))) and
self.linguistic section -> forAll(ls : Closed Linguistic Section | ls.conceptual domain.value -> forAll(dc : Simple Data Category | self.conceptual domain -> exists(value = dc)))

△ – отношение наследования между подклассом и его базовым классом;

♦ – отношение композиции между двумя классами;

◊ – отношение агрегации между двумя классами;

Значения кратности: 0..1 = 0 или 1; 0..* = 0 или более; 1 = 1 и только 1; 1..* = 1 или большему числу случаев.

Рисунок 6 – Лингвистическая часть модели данных реестра DCR

7.8 Ссылки на категорию данных

Явные ссылки на категорию данных из какого-либо ресурса должны выполняться путем вставки в этот ресурс неизменного идентификатора (PID) спецификации этой категории. Идентификатор PID назначается в реестре DCR автоматически.

В некоторых языках описания схем изначально имеются теги для вставки этих PID. В спецификациях, выраженных на языке ODD "Инициативы по кодированию текстов", следует использовать тег <equiv>. Например, следующий код указывает на то, что вводимый элемент (<pos>) имеет смысл, определенный для */partOfSpeech/* в реестре DCR:

```
<elementSpec ident="pos">
      <equiv   name="partOfSpeech"   uri="http://www.isocat.org/datcat/ISO-DC-
1345"/>
      <!-- additional specifications here -->
</elementSpec>
```

В языках описания схем, не предусматривающих такой возможности, но все же основанных на словаре XML, для вставки PID категорий данных может использоваться небольшой справочный словарь категорий данных. В частности, можно обеспечить эквивалентность элемента *POS* и категории */partOfSpeech/* из реестра DCR путем вставки атрибута dcr:datcat в подходящее место схемы Relax NG:

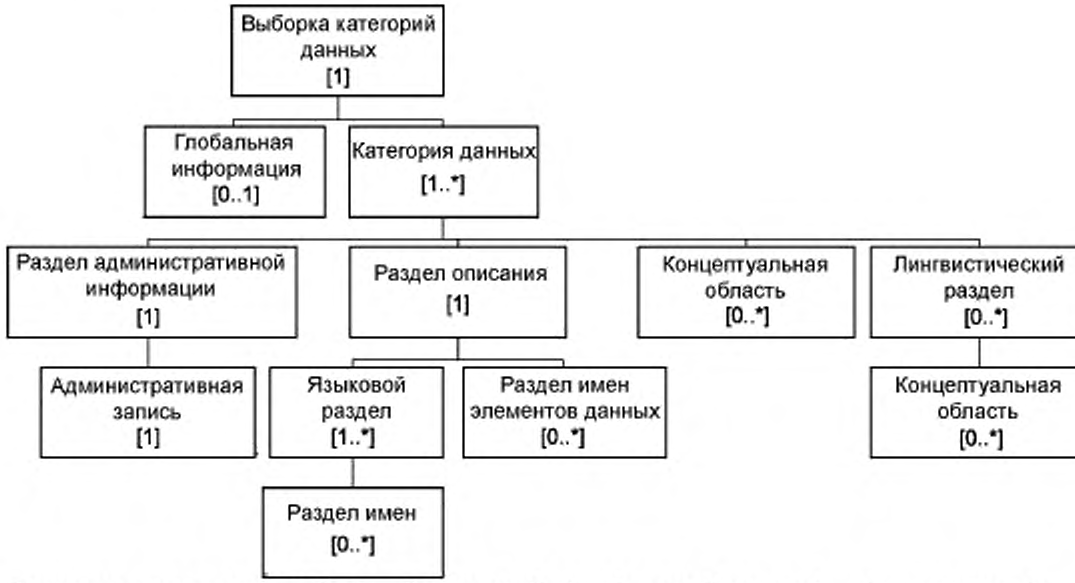
```
<rng:element    name="POS"    dcr:datcat="http://www.isocat.org/datcat/ISO-DC-
1345">
      <!-- additional specifications here -->
</rng:element>
```

Образец схемы Relax NG для использования в таких случаях приведен в приложении А.

7.9 Формат обмена категориями данных

В этом подразделе описывается формат обмена категориями данных (DCIF), который должен использоваться в техническом комитете для архивирования всего реестра DCR или его части, для обмена этими данными, а также в случаях, когда отдельным лицам необходима обработка и передача их собственных категорий данных, относящихся к языковым ресурсам.

Модель DCIF на рис. 7 тесно связана с диаграммой классов, описывающей модель данных DCR (см. рисунки 3 – 6). Однако поскольку диаграмма классов достаточно сложна, то для формата обмена реализован упрощенный вариант этой модели. Модель DCIF описывается в качестве иерархической компонентной модели. Компоненты соотносятся с главными классами модели данных. Упрощение достигается путем включения информации из некоторых ассоциированных классов в главные классы. При необходимости для полученных в результате компонент приводятся аннотации, содержащие признак типа, который указывает на исходный подкласс, с целью обеспечения полной или почти полной сохранности после прямых и обратных преобразований документов DCIF. Тем не менее схема DCIF расширяет множество документов, разрешенных в модели данных DCR, например, она допускает вхождение сложных категорий данных в область значений. При экспорте DCS из DCR в формате DCIF всегда создается XML-документ, соответствующий как модели данных DCR, так и схеме DCIF. В связи с меньшей определенностью схемы DCIF такие XML-документы всегда проверяются на соответствие модели данных DCR во время их импорта.



Значения кратности: 0..1 = 0 или 1; 0..* = 0 или более; 1 = 1 и только 1; 1..* = 1 или большему числу случаев.

Рисунок 7 – Модель данных в основе DCIF

В некоторых случаях (архивирование данных, переход на новую систему и т.д.) выборка DCS в формате DCIF будет содержать все спецификации категорий данных DCR. Однако в большинстве случаев любой данный результат экспорта в DCIF будет относиться к выборке DCS для Группы TDG, для одного эксперта или для группы экспертов, каждая выборка может включать DCS для приложения, если это необходимо.

Кроме упрощения структуры модели данных DCR, в DCIF имеются и следующие упрощения:

— в истории изменений, сохраняемых в разделе описания, включаются только описания первоначального и самого последнего изменения;

— исключена внутренняя административная информация DCR, в том числе классы Registration Group, Submission Group,

Stewardship Group и Decision Group, а также признаки +administration status и +administration note из раздела описаний.

Пример документа DCIF приведен в приложении В. Документы DCIF должны соответствовать компактному варианту схемы, приведенному в приложении С, в соответствии с ИСО 19757-2:2003.

8 Процедуры ведения реестра DCR

8.1 Общая организация

Несмотря на то, что реестр DCR является централизованным ресурсом, его можно разбить на подмножества категорий данных по тематическим областям. Каждое из этих подмножеств включает выборку категорий данных (DCS), за администрирование которой отвечает Группа обслуживания тематической области (TDG), состоящая из назначенных экспертов. Такие выборки DCS предоставляют логически обоснованную схему для представления на рассмотрение и сопровождения новых категорий данных. В каждой спецификации категории данных имеется признак, который указывает на профиль, определяющий тематическую область или области, с которыми данная категория ассоциирована. В этом отношении ведение реестра не в полной мере централизовано, а базируется на структуре, заимствующей необходимые экспертные знания из подобластей лингвистических ресурсов, представленных в техническом комитете. На рис. 8 иллюстрируется распределение групп обслуживания тематических областей в рамках административной организации реестра DCR.

Рисунок 8 – Выборки DCS для отдельных тематических областей и частные категории данных, обслуживаемые отдельными экспертами

На рис. 8 видно, что такая категория данных, как */masculine/ (мужской)*, являющаяся простой категорией данных и входящая в область значений замкнутой категории данных */grammaticalGender/ (грамматический род)*, "принадлежит" морфосинтаксической группе TDG, но, очевидно, что она используется в выборках DCS разнообразных групп TDG. Внешним овалом ограничены категории данных *частной* области, то есть категории, созданные в реестре DCR отдельными экспертами, но не представленные для включения в стандартизованную часть выборки (см. 8.4.3).

8.2 Роли и ответственность

Реестр DCR должен быть общедоступным он-лайн на сайте технического комитета для публичных консультаций, чтобы практикующие специалисты по языкам и конструкторы систем могли пользоваться им в любой ситуации. Лицам, желающим принять участие в работе с реестром DCR, следует заявить об этом, заполнив он-лайн форму, после чего они могут быть назначены *экспертами*.

Независимым экспертам выделяются частные рабочие области реестра, в которых они могут выбирать существующие категории данных для включения в свои выборки DCS и создавать новые спецификации категорий данных для личного пользования. Кроме того, они могут подавать категории данных на рассмотрение для включения в глобальный набор спецификаций категорий данных, как описано ниже, или предлагать изменения существующих категорий данных.

8.3 Группы обслуживания тематических областей

Независимо от работ отдельных экспертов, описанных в 8.2, работы ведутся различными группами TDG технического комитета, ответственными за представление и выбор спецификаций категорий данных, которые разработаны для потребностей документирования данных в конкретных тематических областях. Рассматриваемые в виде набора, эти категории данных составляют специальную выборку DCS конкретной группы TDG. Для определения такой выборки DCS необходимы следующие условия для формирования группы TDG:

a) потребность определения дополнительной выборки DCS в новой тематической области, возникшая в ходе разработки текущих стандартов Подкомитетом (ПК) или Техническим комитетом (ТК) либо после появления новой описательной тематической области, которая существенна для ПК или ТК;

b) за три месяца до пленарного заседания ПК или ТК, на котором будут утверждаться новые предложения, ПК или ТК должны представить документ (“Предложение о создании новой группы обслуживания тематической области”) с определением цели и области деятельности новой группы, а также ее возможных связей с существующими группами TDG, назначенными для реестра DCR;

c) в случае указанной выше потребности ПК или ТК может, если сочтет уместным, принять резолюцию об образовании группы TDG.

В группу TDG должны входить следующие лица, которые будут выполнять обязанности экспертов, выносящих решение, и членов группы обслуживания, отвечающих за оценку спецификаций категорий данных, подаваемых на рассмотрение в TDG:

— руководитель группы, назначаемый ПК во время образования группы, на место которого по необходимости могут быть впоследствии назначены другие;

— группа экспертов, состав которой определяют члены–участники ПК;

— группа подходящих экспертов из других организаций, создание которой предложено членами или руководителем TDG при необходимости усиления состава группы TDG для достижения нужных рабочих показателей. Суммарное число экспертов из других организаций не должно превышать 50% всего числа экспертов в группе TDG.

После формирования группы TDG ей также назначается значение профиля, которое затем вносится в реестр и указывает на выборку DCS, необходимую для данной тематической области. Значение или значения признака *profile* вносятся в соответствующие спецификации категорий данных, и эта информация используется для представления DCS в виде списка или множества спецификаций категорий данных.

В некоторых особых случаях, например при формировании TDG в связи с разработкой какого-либо стандарта, могут быть допущены определенные отклонения от вышеупомянутых процедур. В частности, может быть так, что действующий Орган регистрации для ИСО 639, существующие процедуры и структура могут быть назначены TDG для

описания языка в связи с проработкой новой части ИСО 639. В других случаях мероприятия по стандартизации будут проводиться в тесном сотрудничестве с другими техническими комитетами.

8.4 Порядок работ

8.4.1 Процесс принятия решения по категории данных

Как указано в 7.4.2 (подпункт 3, +administration status), процесс, приводящий к включению категории данных в стандартизованную часть DCR или к пересмотру категории из этой части, должен состоять из четырех шагов, при этом процесс стандартизации начинается с представления категории данных на рассмотрение в TDG:

0) *процесс создания*, в который входит создание категории данных отдельным экспертом и который сам по себе не является процессом стандартизации, так как категории данных остаются в частной рабочей области эксперта до их возможного окончательного представления на рассмотрение для стандартизации;

1) *процесс представления на рассмотрение*, который инициирует процесс стандартизации: категория данных представляется на рассмотрение в качестве заявки на изменение (CR) в группу TDG для включения в стандартизованную часть реестра DCR;

2) *процесс выбора*, во время которого группа TDG определяет те категории данных, которые уместны для конкретной сферы приложений в техническом комитете; данный процесс происходит параллельно *оценке* согласно Приложению ST, как показано на рис. 9;

3) *процесс гармонизации* под наблюдением DCRB, гарантирующий согласование новых предложений с областью

применения реестра DCR и категорий данных, уже содержащихся в нем; данный процесс, предусматривающий окончательное утверждение новых категорий Советом DCRB, происходит параллельно *валидации* согласно Приложению ST, как показано на рис. 9.

8.4.2 Создание новой категории данных

По собственной инициативе или по поручению TDG эксперты создают спецификации категорий данных в своих частных рабочих областях и могут либо оставить их в своих личных выборках DCS, либо передать на рассмотрение для включения в общедоступную, стандартизованную часть DCR. Как отмечено выше, все вновь создаваемые категории автоматически приписываются к области *Private* и остаются вне процесса стандартизации до представления на рассмотрение группе TDG.

8.4.3 Подача заявки на изменение (CR)

8.4.3.1 Подача CR для новой категории данных

Заявка CR для новой категории данных, подаваемая в группу TDG, должна быть основана на описании, которое соответствует модели данных DCR, и включать: информацию, ассоциируемую с разделом описания Description Section и его подчиненными классами и признаками, которая должна быть документирована в *процессе представления на рассмотрение,* в том числе:

-по крайней мере одно определение в разделе английского языка;

-источники определений, если определения взяты из известных источников;

-по крайней мере одно значение профиля тематической области, ассоциируемое со спецификацией категории данных;

-по крайней мере один раздел английского языка для категории данных;

-краткое описание с обоснованием актуальности категорий данных для сферы языковых ресурсов;

-обязательную административную информацию, автоматически генерируемую программным обеспечением DCR;

-дополнительные сведения, например, имена и определения на других языках, если они уместны и имеются в наличии.

8.4.3.2 Подача CR с предложением изменения существующей спецификации категории данных

Любой эксперт или группа TDG может предложить модификацию категории данных. Запрос на модификацию должен рассматриваться в качестве CR и включать следующие данные:

— категория данных, однозначно определяемая по ее неизменному идентификатору;

— только признаки, для которых предлагается изменение (модификация либо добавление информации);

— краткое описание с обоснованием предлагаемого изменения.

8.4.3.3 Подача CR для включения существующей категории данных в DCS тематической области

Любой эксперт или группа TDG может предложить добавление категории данных в дополнительную тематическую область путем включения соответствующего значения профиля в спецификацию этой категории. Такой запрос должен рассматриваться в качестве CR и включать следующие данные:

— категория данных, однозначно определяемая по ее неизменному идентификатору;

— краткое описание с обоснованием важности включения категории данных в тематическую область;

— имя тематической области, в которую предлагается включить категорию данных.

8.4.3.4 Подача CR для выбора категорий данных, составляющих DCS тематической области

Полный набор категорий данных, составляющих выборку DCS, которая ассоциируется с конкретной тематической областью, определяется в результате выполнения совокупности вышеупомянутых процедур создания новых категорий данных, модификации существующих категорий данных и включения последних в тематические области.

8.4.3.5 Подача CR в связи с мероприятиями по гармонизации

Гармонизация определений категорий данных и использования этих категорий в различных группах TDG должна всегда являться конечной целью описанных процедур и предметом курирования DCRB. Запросы на гармонизацию рассматриваются как заявки CR в части стандартной схемы мероприятий, которая показана на рис. 9. При необходимости гармонизации существующих определений категорий данных она происходит в контексте заявки CR на изменение.

8.5 Совет по администрированию реестра категорий данных (DCRB)

8.5.1 Организация Совета DCRB

Обязанностью Совета DCRB является обеспечение поддержания состава и согласованности реестра. Совет должен играть гармонизирующую роль в отношении предложений, представляемых экспертами и группами обслуживания тематических областей.

В Совет DCRB должны входить:

— группа экспертов, назначаемых членами–участниками технического комитета;

— председатель, назначаемый на пленарном заседании ТК сроком на два года, который может быть продлен один раз.

8.5.2 Порядок работ

Совет DCRB совместно с Органом регистрации отвечает за валидацию представленных заявок и за публикацию стандартизованного раздела реестра DCR.

8.5.2.1 Валидация предложений о включении категорий данных

Любая категория данных, представленная на рассмотрение, должна как минимум соответствовать критериям, которые приведены в 8.4.3.

Категория данных должна быть подвергнута оценке в ответственной группе TDG, указанной в признаке +administration status, который описан в 7.4.2, должна быть утверждена, отклонена либо передана на рассмотрение другой группе TDG.

Категория должна пройти процедуру валидации на уровне DCRB.

Положительным результатом голосования при валидации считается более 70% голосов, после чего статус категории данных повышается до “стандартизовано”. Если набрано менее 70% голосов, то категории данных должен быть присвоен статус “отвергнуто” и подателю заявки должна быть передана информация о причинах такого решения. Если отвергнутая категория данных после ее изменения представляется на рассмотрение повторно, то повторное рассмотрение должно следовать процедуре, используемой для новой спецификации категории данных. При этом должны быть в наличии замечания о предыдущем представлении на рассмотрение.

8.5.2.2 Публикация справочной версии реестра

Каждые шесть месяцев Орган регистрации, отвечающий за ведение реестра DCR, должен выпускать обновленную версию реестра DCR со всеми стандартизованными спецификациями категорий данных, которые были утверждены Советом DCRB. Этой версии, отражающей текущее состояние реестра, должна быть присвоена дата и номер, и в течение следующего шестимесячного периода она должна считаться официальной стандартной версией. Совет DCRB должен передавать извещения членам ТК, связанным с ними организациям, секретарю ПК и руководителям групп TDG об изменениях, которые произошли после предыдущего выпуска (о добавлениях, модификациях и исключениях). Спецификации категорий данных, которые "публикуются" таким образом, должны составлять "стандарт в формате базе данных" для утвержденных категорий данных, которые при этом должны оставаться доступными для всех пользователей в условиях динамичного развития реестра DCR.

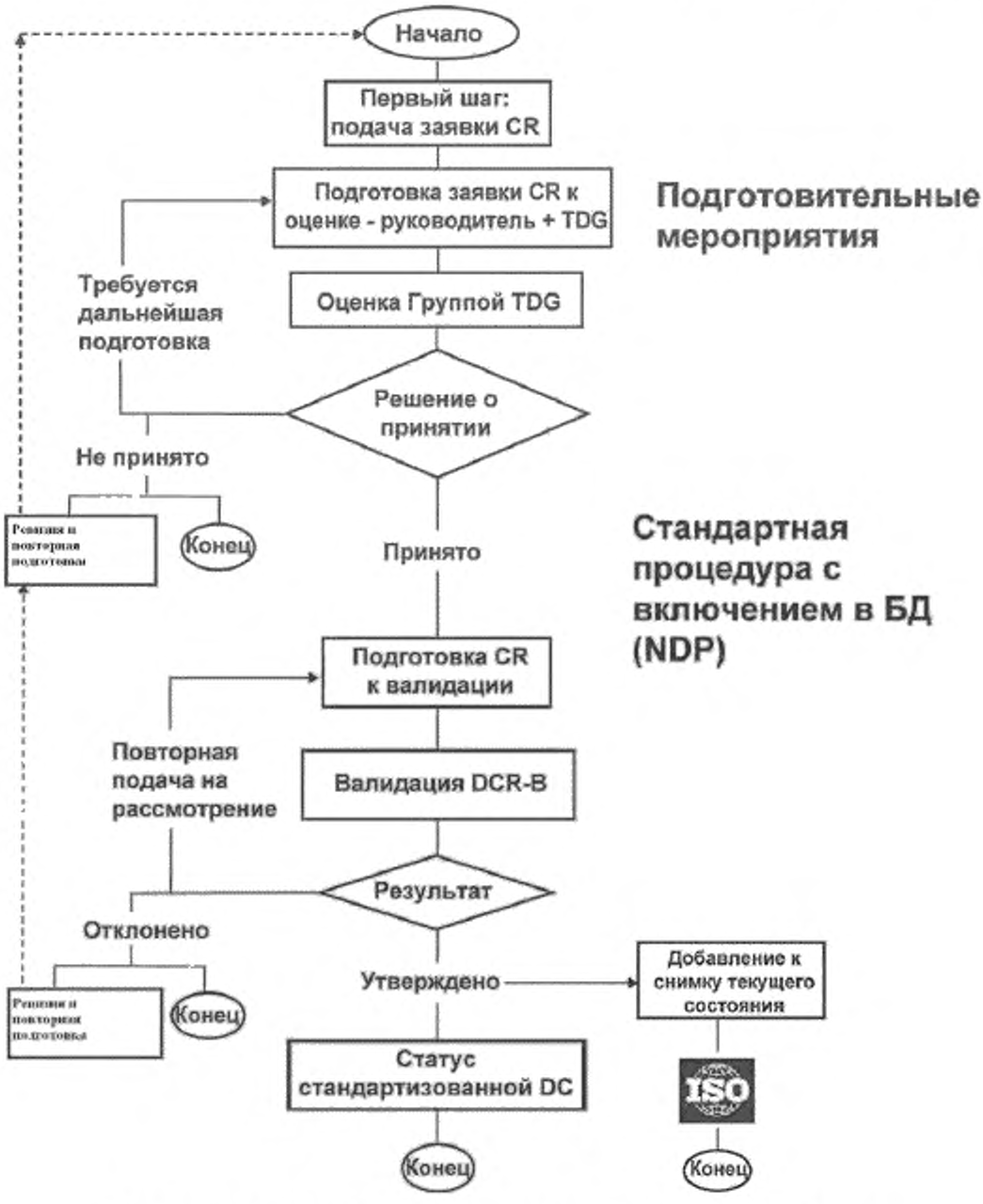


Рисунок 9 – Схема последовательности мероприятий

Приложение А
(обязательное)
Компактная схема RELAX NG для ссылок на категории данных

```
default namespace dcr = "http://www.isocat.org/ns/dcr"

start = any

any = dcr_elements | element * - dcr:* { content }

content = ( dcr_attributes | attribute * - dcr:* { text } | any | text )*

dcr_attributes = dcr_attribute_datcat | dcr_attribute_any

dcr_attribute_datcat = attribute datcat { xsd:anyURI }

dcr_attribute_any = attribute dcr:* - dcr:datcat { text }

dcr_elements = dcr_element_datcat | dcr_element_any

dcr_element_datcat = element datcat { attribute pid { xsd:anyURI } }

dcr_element_any = element dcr:* { content }
```

Примечание – Данная схема открыта в пространстве имен http://www.isocat.org/ns/dcr; см. шаблоны с именами dcr_attribute_any и dcr_element_any. Новые атрибуты и элементы, вводимые при использовании этой схемы, предназначены для применения с целью представления информации, относящейся только к реестру DCR.

Приложение В
(справочное)
Пример представления DCIF

Ниже приведен пример XML представления DCIF для основной информации, которая ассоциирована с категорией данных на основе принципов, описанных в настоящем документе:

```
<?xml version="1.0"?>
<dataCategorySelection xmlns="http://www.isocat.org/ns/dcif" dcif-version="1.0">
   <globalInformation>
      Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
   </globalInformation>
   <dataCategory pid="http://www.isocat.org/datcat/ISO-DC-1234" type="complex">
      <administrationInformationSection>
         <administrationRecord>
            <identifier>grammaticalGender</identifier>
            <version>1.0.0</version>
            <registrationStatus>standard</registrationStatus>
            <origin>ISO 12620:1999</origin>
            <justification>
               In many, but not all, Indo-European languages grammatical gender
               describes the classification of nouns according to their behaviour
               with respect to inflection, pronoun agreement, semantics, morphology,
               and linguistic convention.
            </justification>
            <creation>
               <creationDate>1999-01-01</creationDate>
                  <changeDescription xml:lang="en">
                     Initial creation of the /grammatical gender/ data
                     Category
                  </changeDescription>
            </creation>
         </administrationRecord>
      </administrationInformationSection>
      <descriptionSection>
         <profile>terminology</profile>
         <languageSection>
            <language>en</language>
            <definitionSection>
               <definition xml:lang="en">
                  A grammatical category that indicates grammatical relationships
                  between words in sentences.
               </definition>
               <source>ISO 12620:1999</source>
               <note>
                  The concept of gender varies from language to language and is not a
                  universal feature of all languages.
               </note>
            </definitionSection>
            <exampleSection>
               <example xml:lang="en">
                  In French, vie [life] is feminine and is used with feminine articles
                  such as la, the feminine pronoun elle, and feminine adjective
                  endings, for example, une vie longue.
               </example>
            </exampleSection>
            <nameSection>
               <name>grammatical gender</name>
               <nameStatus>standardized name</nameStatus>
            </nameSection>
         </languageSection>
      </descriptionSection>
      <conceptualDomain type="closed">
         <dataType>string</dataType>
         <profile>terminology</profile>
         <value pid="http://www.isocat.org/datcat/ISO-DC-1111">
            <!-- PID reference to the /masculine/ simple DC -->
```

```
        </value>
        <value pid="http://www.isocat.org/datcat/ISO-DC-1112">
          <!-- PID reference to the /feminine/ simple DC -->
        </value>
        <value pid="http://www.isocat.org/datcat/ISO-DC-1113">
          <!-- PID reference to the /neuter/ simple DC -->
        </value>
        <value pid="http://www.isocat.org/datcat/ISO-DC-1114">
          <!-- PID reference to the /other/ simple DC -->
        </value>
      </conceptualDomain>
      <linguisticSection type="closed">
        <language>fr</language>
        <conceptualDomain type="closed">
          <dataType>string</dataType>
          <value pid="http://www.isocat.org/datcat/ISO-DC-1111">
            <!-- PID reference to the /masculine/ simple DC -->
          </value>
          <value pid="http://www.isocat.org/datcat/ISO-DC-1112">
            <!-- PID reference to the /feminine/ simple DC -->
          </value>
        </conceptualDomain>
      </linguisticSection>
    </dataCategory>
</dataCategorySelection>
```

Приложение С
(обязательное)
Компактная схема DCIF RELAX NG

```
default namespace dcif = "http://www.isocat.org/ns/dcif"

start = dataCategorySelection

dataCategorySelection =
   element dataCategorySelection
   {
      attribute dcif:version { "1.0" },
      attribute name { text }?,
      attribute owner { text }?,
      ( globalInformation? & dataCategory+ )
   }

globalInformation = element globalInformation { ( text   any )+ }

dataCategory =
   element dataCategory
   {
      attribute pid { xsd:anyURI },
      dataCategorySummary?,
      dataCategoryLineage?,
      (
         ( attribute type { "complex" | "simple" } )
       | ( complexDataCategory | simpleDataCategory )
       | (
            administrationInformationSection
          & descriptionSection
          & ( complexDataCategory | simpleDataCategory )
         )
      )
   }

dataCategorySummary =
   attribute identifier { xsd:string },
   attribute owner { xsd:string },
   attribute version { xsd:string },
   attribute name { xsd:string }?,
   attribute definition { xsd:string }?

dataCategoryLineage =
   attribute first { xsd:anyURI }?,
   attribute previous { xsd:anyURI }?,
   attribute next { xsd:anyURI }?,
   attribute latest { xsd:anyURI }?

administrationInformationSection =
   element administrationInformationSection { administrationRecord }

administrationRecord =
   element administrationRecord
   {
      element identifier { xsd:NCName }
    & element version { xsd:string }
    & element registrationStatus
      {
         "candidate" | "standard" | "deprecated"  . "superseded"
      }
```

```
    & element origin { xsd:string }?
    & element justification { xsd:string }?
    & element explanatoryComment
      {
         xsd:string,
         attribute xml:lang { xsd:language }?
      }*
    & element unresolvedIssue
      {
         xsd:string,
         attribute xml:lang { xsd:language }?
      }*
    & element effectiveDate { xsd:date }?
    & element untilDate { xsd:date }?
    & element creation
      {
         element creationDate { xsd:date }
       & element changeDescription
         {
            xsd:string,
            attribute xml:lang { xsd:language }?
         }
      }
    & element lastChange
      {
         element lastChangeDate { xsd:date }
       & element changeDescription
         {
            xsd:string,
            attribute xml:lang { xsd:language }?
         }
      }?
   }

descriptionSection =
   element descriptionSection
   {
      element profile { xsd:string }+
    & languageSection+
    & dataElementNameSection*
   }

languageSection =
   element languageSection
   {
      element language { xsd:language }
    & element definitionSection
      {
         element definition
         {
            xsd:string,
            attribute xml:lang { xsd:language }?
         }
       & element source { xsd:string }
       & element note { xsd:string, attribute xml:lang { xsd:language }? }?
      }*
    & exampleSection*
    & explanationSection*
    & element note { xsd:string, attribute xml:lang { xsd:language }? }*
    & nameSection*
   }
```

```
exampleSection =
   element exampleSection
   {
      element example { xsd:string, attribute xml:lang { xsd:language }? }
    & element source { xsd:string }?
   }

explanationSection =
   element explanationSection
   {
      element explanation { xsd:string, attribute xml:lang { xsd:language }?
}
    & element source { xsd:string }?
   }

nameSection =
   element nameSection
   {
      element name { xsd:string }
    & element nameStatus
      {
         "standardized name"
       | "preferred name"
       | "admitted name"
       | "deprecated name"
       | "superseded name"
      }
   }

dataElementNameSection =
   element dataElementNameSection
   {
      element dataElementName { xsd:string } & element source { xsd:string }
   }

complexDataCategory =
   attribute type { "complex" },
   ( openDataCategory | closedDataCategory | constrainedDataCategory )

openDataCategory = openConceptualDomain & constrainedLinguisticSection*

openConceptualDomain =
   element conceptualDomain
   {
      attribute type { "open" },
      element dataType { xsd:string }
   }

closedDataCategory = profileValueDomain & closedLinguisticSection*

profileValueDomain =
   element conceptualDomain
   {
      attribute type { "closed" },
      (
         element dataType { xsd:string }
       & element profile { xsd:string }
       & element value { attribute pid { xsd:anyURI }, dataCategorySummary?
}+
      )
   }
```

```
closedLinguisticSection =
   element linguisticSection
   {
      attribute type { "closed" },
      ( linguisticSection & valueDomain? )
   }

linguisticSection =
   element language { xsd:language }
 & exampleSection*
 & explanationSection*
 & element note { xsd:string, attribute xml:lang { xsd:language }? }*

valueDomain =
   element conceptualDomain
   {
      attribute type { "closed" },
      (
         element dataType { xsd:string }
       & element value { attribute pid { xsd:anyURI }, dataCategorySummary?
}+
      )
   }

constrainedDataCategory = conceptualDomainRule* &
constrainedLinguisticSection*

conceptualDomainRule =
   element conceptualDomain
   {
      attribute type { "constrained" },
      (
         element dataType { xsd:string }
       & element ruleType { xsd:string }
       & element rule { ( text | any )+ }+
      )
   }

constrainedLinguisticSection =
   element linguisticSection
   {
      attribute type { "constrained" },
      ( linguisticSection & conceptualDomainRule* )
   }

simpleDataCategory =
   attribute type { "simple" },
   element isA { attribute pid { xsd:anyURI }, dataCategorySummary? }?

any =
   element * - dcif:*
   {
      attribute * - dcif:* { text }*,
      ( text | any )*
   }
```

Приложение D
(справочное)
Алфавитный список определений

DC	**DC**	3.1.3
DCIF	**DCIF**	3.6.1
DCR	**DCR**	3.2.1
DCRB	**DCRB**	3.4.1
DCS	**DCS**	3.2.3
DE	**DE**	3.1.2
GI	**GI**	3.3.2
PID	**PID**	3.6.3
RA	**RA**	3.4.2
TDG	**TDG**	3.4.4
выборка категорий данных	**Data Category Selection**	3.2.3
глобальная информация	**Global Information**	3.3.2
группа ведения реестра	**stewardship group**	3.3.7
группа обслуживания тематической области	**thematic domain group**	3.4.4
группа подачи на рассмотрение	**submission group**	3.3.5
группа принятия решений	**decision group**	3.3.6
группа регистрации	**registration group**	3.3.4
замкнутая категория данных	**closed data category**	3.1.13
замкнутая концептуальная область	**closed conceptual domain**	3.1.14
имя элемента данных	**data element name**	3.3.9
категория данных	**data category**	3.1.3
концептуальная область	**conceptual domain**	3.1.5
концепция элемента данных	**data element concept**	3.1.4
лингвистический раздел	**linguistic section**	3.3.11
модель данных DCR	**DCR data model**	3.3.1
моментальный снимок	**snapshot**	3.6.2
неизменный идентификатор	**persistent identifier**	3.6.3
область значений	**value domain**	3.1.6
объектный язык	**object language**	3.3.13
ограниченная категория данных	**constrained data category**	3.1.10
ограниченная концептуальная область	**constrained conceptual domain**	3.1.11
Орган регистрации	**Registration Authority**	3.4.2
открытая категория данных	**open data category**	3.1.8
открытая концептуальная область	**open conceptual domain**	3.1.9
тематическая область	**thematic domain**	3.4.3
председатель DCRB	**chair of the DCRB**	3.5.1
председатель Совета DCR	**chair of the DCR Board**	3.5.1
председатель Совета по администрированию реестра категорий данных	**chair of the Data Category Registry Board**	3.5.1
простая категория данных	**simple data category**	3.1.12
профиль	**profile**	3.4.5
профиль тематической области	**thematic domain profile**	3.4.5
рабочий язык	**working language**	3.3.12
раздел административной информации	**administration information section**	3.3.3
раздел имен	**name section**	3.3.14
раздел описания	**description section**	3.3.8
реестр категорий данных	**Data Category Registry**	3.2.1

руководитель TDG	**TDG chair**	3.5.2
руководитель группы обслуживания тематической области	**chair of a thematic domain group**	3.5.2
сложная категория данных	**complex data category**	3.1.7
Совет DCR	**DCR Board**	3.4.1
Совет по администрированию реестра категорий данных	**Data Category Registry Board**	3.4.1
спецификация категории данных	**data category specification**	3.2.2
схема аннотации	**annotation scheme**	3.1.15
формат обмена категориями данных	**Data Category Interchange Format**	3.6.1
эксперт	**expert**	3.5.3
эксперт, выносящий решение	**judge**	3.5.4
элемент данных <стандарты по метаданным >	**data element** <metadata standards>	3.1.2
элемент данных <языковые ресурсы>	**data element** <language resources>	3.1.1
языковой раздел	**language section**	3.3.10

Приложение ДА
(справочное)
Сведения о соответствии ссылочных международных стандартов ссылочным национальным стандартам Российской Федерации

Сведения о соответствии ссылочных международных стандартов ссылочным национальным стандартам Российской Федерации приведены в таблице ДА.1.

Таблица ДА.1

Обозначение ссылочного международного стандарта	Степень соответствия	Обозначение и наименование соответствующего национального стандарта
ИСО 8601:2004		*
ИСО/МЭК 11179-1:2004	–	*
ИСО/МЭК 11179-3:2003	–	*
* Соответствующий национальный стандарт отсутствует. До его утверждения рекомендуется использовать перевод на русский язык данного международного стандарта. Перевод данного международного стандарта находится в Федеральном информационном фонде технических регламентов и стандартов.		

Библиография

[1] ИСО 639-1:2002, *Коды для представления названий языков. Часть 1. Двухбуквенный код*

[2] ИСО 639-2:1998, *Коды для представления названий языков. Часть 2. Трехбуквенный код*

[3] ИСО 639-3:2007, *Коды для представления названий языков. Часть 3. Трехбуквенный код для всестороннего охвата языков*

[4] ИСО 1087-2:2000[1)], *Терминологическая работа. Словарь.Часть 2. Применение вычислительной техники*

[5] ИСО 1951:2007, *Представление/обозначение словарных статей. Требования, рекомендации и информация*

[6] ИСО 16642:2003, *Применения компьютера в терминологических целях. Структура терминологической разметки*

[7] ИСО/МЭК 11179-6:2005, *Информационные технологии. Реестры метаданных (MDR). Часть 6. Регистрация*

[8] ИСО/МЭК 19757-2:2003[2)], *Информационные технологии. Язык определения схемы документа. Часть 2. Валидация на основе стандартной грамматики. RELAX NG*

[9] Директивы ИСО/МЭК, Приложение ST к Дополнению ИСО. Процедуры, используемые только в ИСО. *Процедура разработки и сопровождения стандартов в формате базы данных*

[10] TEI. 2009. Инициатива по кодированию текстов. Введение в универсальные документы P5 http://www.tei-c.org/Support/Learn/odds.xml

[11] W3C. *Схема XML. Средства, применение, ресурсы, спецификации и разработка* http://www.w3.org/XML/Schema

[12] Технический комитет OASIS по Relax NG. *Relax NG* http ://relaxng.org/

[13] OMG. *Унифицированный язык моделирования* http://www.uml.org/

[14] W3C. *Пространства имен в XML 1.0* (второе издание). Рекомендация W3C от 16 августа 2006 г. http ://www.w3.org/TR/REC-xml-names/

[15] EAGLES: *Экспертная консультационная группа по стандартам разработки языков* http://www.ilc.cnr.it/EAGLES/home.html

[16] OMG. *UML 2.0. Спецификация OCL [языка объектных ограничений]* http://www.omg.org/docs/ptc/03-10-14.pdf

[17] IETF BCP 47, RFC 5646. 2009. *Теги для идентификации языков* http://tools.ietf.org/rfc/bcp/bcp47.txt

[18] TEI. 2005. *Руководящие указания TEI. P5* http://www.tei-c.org/release/doc/tei-p5-doc/html/

[1)] Отменен.

[2)] Отменен. Действует ISO/IEC 19757-2:2008

УДК 001.4: 006.354 ОКС 01.020, 35.240.30

Ключевые слова: терминология, языковые ресурсы, категории данных

Подписано в печать 30.04.2014. Формат 60x84 1/8.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

ФГУП «СТАНДАРТИНФОРМ»

123995 Москва, Гранатный пер., 4.
www.gostinfo.ru info@gostinfo.ru