

Модель данных Всемирной организации интеллектуальной собственности стандарта ST.96

А.Б. Корчагин, И.Г. Лисьих, Д.А. Никифоров, Р.Л. Сиваков

Аннотация – Данная публикация продолжает цикл статей, посвященных анализу подходов к моделированию данных и выявлению их технологических особенностей. Публикация посвящена модели, рекомендованной стандартом Всемирной организации интеллектуальной собственности ST.96. Рассмотрена эволюция относящихся к обработке информации стандартов Всемирной организации интеллектуальной собственности по мере развития технологий электронного обмена данными. Представлена логика построения модели ST.96 в целом и ее компонентов: элементов, типов и атрибутов. Проведено сравнение входящих в состав модели структур электронных документов, содержащих сведения о различных объектах промышленной собственности: патентах, товарных знаках и промышленных образцах. Рассмотрены технологические особенности модели ST.96: наличие элементов, используемых для разметки текста; применение одного из шаблонов (паттернов) проектирования XML-схем.

Ключевые слова – информационное взаимодействие; модель данных; совместимость моделей данных; отображение моделей данных; онтология, Всемирная организация интеллектуальной собственности, XML.

I. ВВЕДЕНИЕ

Большинство стандартов Всемирной организации интеллектуальной собственности (ВОИС, World Intellectual Property Organization, WIPO) содержат рекомендации по тем или иным аспектам обработки информации в сфере защиты интеллектуальной собственности. В частности, стандарты 90-х годов посвящены микрофильмам, микрофишам, магнитным лентам и т.п. Например [1, 2]:

- ST.7/B Recommended standard for 16mm roll microfilm for exchange between patent offices / Рекомендуемый стандарт на 16 мм рольный микрофильм для обмена информацией между патентными ведомствами;
- ST.30: Magnetic tape format for the exchange of data

Статья получена 14 июня 2017.

А.Б. Корчагин – кандидат технических наук, доцент Уральского федерального университета, ведущий консультант компании Центр ИТ (e-mail: Alexander.Korchagin@centre-it.com).

И.Г. Лисьих – ведущий консультант компании Центр ИТ (e-mail: Igor.Lisyih@centre-it.com).

Д.А. Никифоров – бизнес-аналитик компании Центр ИТ (e-mail: Denis.Nikiforov@centre-it.com).

Р.Л. Сиваков – президент группы компаний «ЦИТ» (e-mail: Ruslan.Sivakov@centre-it.com).

and documents / Рекомендации по стандартному формату магнитной ленты для обмена библиографическими данными, рефератами и полными текстами патентных документов в машиночитаемой форме.

По мере развития электронного обмена данными, с появлением XML-технологий возникла необходимость в разработке соответствующих времени стандартов, поддерживающих электронное информационное взаимодействие [3].

Сегодня ВОИС формирует так называемую глобальную инфраструктуру управления интеллектуальной собственностью Global IP Infrastructure [4], поддерживая создание и распространение единых программных инструментов, сервисов, стандартов, баз данных.

В рамках Global IP Infrastructure реализуется инициатива распространения данных об интеллектуальной собственности Global Dissemination of IP Data Initiative [5], направленная на повышение эффективности международного информационного взаимодействия. Инициатива уделяет значительное внимание использованию согласованных форматов данных.

К стандартам, отражающим современные тенденции в области информационного взаимодействия, можно отнести следующие [1, 2]:

- ST.25 Standard for the presentation of nucleotide and amino acid sequence listings in patent applications (December 2009) / Стандарт по представлению перечней последовательностей нуклеотидов и аминокислот в патентных заявках;
- ST.26 Recommended standard for the presentation of nucleotide and amino acid sequence listings using XML (June 2016) / Рекомендуемый стандарт представления перечней нуклеотидных и аминокислотных последовательностей с использованием языка XML (проект);
- ST.32 Recommendation for the markup of patent documents using SGML (November 1995) / Рекомендации по разметке патентных документов с использованием SGML;
- ST.36 Recommendation for the processing of patent information using XML (November 2007) / Рекомендации по обработке патентной информации с использованием XML;
- ST.66 Recommendation for the processing of trademark information using XML (March 2012) / Рекомендации по обработке информации о товарных знаках с использованием XML;
- ST.86 Recommendation for the processing of industrial design information using XML (February 2008) / Рекомендации по обработке информации о промышленных об-

разцах с использованием XML;

– ST.96 Recommendation for the processing of Industrial Property information using XML (September 2016) / Рекомендации по обработке информации по промышленной собственности с использованием XML.

Авторы осознанно включили в список стандарт из 90-х ST.32, дабы подчеркнуть влияние развития информационных технологий на разработку отраслевых стандартов.

Подобное же влияние прослеживается и в стандартах ST.25 и ST.26, первый из которых содержит лишь список полей (атрибутов), унифицирующий представление информации о нуклеотидных и аминокислотных последовательностях, а второй уже явным образом ориентирован на XML-технологии.

Как следует из названий стандартов, очевидна их (и заложенных в них моделей) направленность на применение XML-технологий.

Разработку стандартов ВОИС сегодня ведет комитет по стандартам ВОИС, учрежденный в 2009 году. Коми-

тет собирается на ежегодной основе, его членами являются все государства-члены и наблюдатели ВОИС.

На прошедшей в мае-июне 2017 года сессии комитета, в частности, утвержден «радикальный» вариант перехода от ST.25 к ST.26 при котором датой перехода будет служить январь 2022 г., а также принято решение о расширении рассмотренного ниже ST.96 посредством включения в него географических указаний, как было предложено делегацией Российской Федерации.

В целом в ВОИС входят 189 государств, включая Российскую Федерацию. Оперативную работу ведут рабочие (целевые) группы (Task Forces). Участниками целевых групп являются представители национальных и межгосударственных ведомств по интеллектуальной собственности, например, Федеральная служба по интеллектуальной собственности (Роспатент), United States Patent and Trademark Office (USPTO), European Patent Office (EPO), Japan Patent Office (JPO) и др.

Далее рассмотрим модель, заложенную в стандарте ST.96 [2], приходящем на смену ST.36, ST.66, ST86.

II. ПАСПОРТ МОДЕЛИ

Наименование поля	Значение поля
Название модели	Standard WIPO ST.96. Recommendation for the Processing of Industrial Property Information using XML Стандарт ВОИС ST.96. Рекомендации по обработке информации по промышленной собственности с использованием XML
Ответственная организация	The XML4IP Task Force of the Committee of WIPO Standards (CWS) Рабочая (целевая) группа XML4IP Комитета по стандартам ВОИС
Предметная область	ГРНТИ: 10.35 Патентное право. Право промышленной собственности; 10.41 Авторское право и смежные права; 10.89.35 Международное частное право. Патентное право; 10.89.41 Международное частное право. Авторское право OECD: 05.02.DI BUSINESS; 05.05.OM LAW
Описание	Используется для подачи, публикации, обработки и обмена информацией о промышленной собственности
Дата	2012
Тип модели	Модель данных
Формат представления	Таблицы Excel W3C XML-схемы
Источник	http://www.wipo.int/standards/en/st96/
Язык	eng (английский)
Связь с другими моделями	Заменяет: ST.36. Processing patent information using XML; ST.66. Processing trademark information using XML; ST.86. Processing industrial design information using XML.
Охват	Международная модель
Права	Распространяется без ограничений
Номер актуальной версии	2.2
Дата актуальной версии	17.08.2016
Нотация	UML
Стандарт	Standard WIPO ST.96. Recommendation for the Processing of Industrial Property Information using XML Стандарт ВОИС ST.96. Рекомендации по обработке информации по промышленной собственности с использованием XML
Метамодель	Метамодель явно не описана. Методика изложена в Приложении I. XML Design Rules and Conventions

III. ОПИСАНИЕ МОДЕЛИ

Модель, рекомендованная стандартом ST.96, описывает структуру электронных документов, содержащих сведения о различных объектах промышленной собственности: патентах, товарных знаках, промышленных образцах. Данные структуры уже описаны в стандартах ST.36, ST.66 и ST.86. ST.96 направлен на их гармонизацию и объединение. На текущий момент ST.36, ST.66 и ST.86 не отменены, применяются во многих организациях, но осуществляется эволюционный переход к ST.96.

Разработку стандарта ST.96 курирует группа XML4IP. В состав XML4IP входят представители более 20 ведомств по интеллектуальной собственности.

ST.96 включает собственно стандарт и следующие приложения:

I. XML Design Rules and Conventions / Правила и условные обозначения дизайна XML;

II. IP Data Dictionary / Словарь данных по промышленной собственности (далее IP);

III. IP XML Schemas / Схемы XML IP;

IV. XML Schema Technical Specification / Техническое описание схем XML;

V. Implementation rules and guidelines / Правила и инструкции для практической реализации схем;

VI. Transformation rules and guidelines / Правила преобразования и инструкции.

Непосредственно модель представлена в следующих источниках:

Приложение II, включающее ссылку на словарь элементов и атрибутов, а также словарь перечислений (enumeration) в XML-формате;

Приложение III, являющееся набором XML-схем компонентов, определенных в Приложении II;

Приложение IV, содержащее спецификации (технические описания) XML-схем, определенных в Приложении III.

ST.96 учитывает наличие национальной, региональной или иной специфики и, как следствие, невозможность создания универсальных XML-схем, необходимых ведомствам по промышленной собственности. Поэтому стандарт лишь рекомендует использование данной модели в ведомствах. Однако при обмене данными стандарт требует ее применения. Цитата из стандарта: «Для обмена ДОЛЖНЫ использоваться лишь XML документы, соответствующие Стандарту ST.96». Таким образом, несмотря на название, стандарт носит рекомендательный характер лишь в границах ведомств, но становится обязательным при информационном взаимодействии между ведомствами. Иначе используемый в обмене XML-документ не будет считаться соответствующим стандарту.

Приложение V содержит рекомендации по адаптации (customization) XML-схем, составляющих Приложение III, к требованиям отдельных ведомств. Несмотря на то, что во многих случаях XML-схемы стандарта ST.96 достаточны для использования в существующем виде,

стандарт допускает их модификацию.

Предполагаются два варианта адаптированных схем: схемы, совместимые (compatible) со стандартом, и схемы, соответствующие (conformant) стандарту. Совместимые схемы могут расширять исходные схемы. XML-документ, прошедший валидацию по совместимой схеме, может не пройти валидацию по схемам стандарта. Соответствующая стандарту схема не имеет расширений, и XML-документ, валидный по соответствующей схеме, валиден и по схемам стандарта.

В рамках Приложения V также предложен построенный на базе языка Schematron [6] программный продукт ST96XSDValidator, позволяющий производить проверку соответствия XML-схем правилам и условным обозначениям, изложенным в Приложении I.

Приложение VI содержит рекомендации по взаимному преобразованию XML-документов, построенных на базе ST.96, и XML-документов, построенных на базе ST.36, ST.66 и ST.86. В частности, Приложение включает таблицы соответствия (mapping tables) элементов и атрибутов различных стандартов ВОИС, а также примеры XSLT преобразований (eXtensible Stylesheet Language Transformations), основанных на данных таблицах.

В части именования элементов, типов и атрибутов модель ST.96 опирается на стандарт ISO 11179, Information technology – Metadata registers (MDR), Part 5: – Naming and identification principals (русский аналог – ГОСТ Р ИСО/МЭК 11179-5-2012).

ST.96 ориентирован на модульный принцип проектирования XML-схем. Разделение схем на отдельные модули позволяет наиболее полно реализовать механизм повторного использования компонентов модели.

Используемый в стандарте термин «компонент» является обобщающим для элементов, типов и атрибутов. Все компоненты разбиты на следующие категории (принадлежат следующим доменам (областям)): Common Components (общие компоненты), Patent Components (компоненты патентов), Trademark Components (компоненты товарных знаков), Design Components (компоненты промышленных образцов), External Standard Components (компоненты внешних стандартов). В соответствии с категориями определены и пространства имен в XML-схемах: com, pat, tmk, dgn, tbl (для схем табличных компонентов OASIS), mathml (для схем компонентов языка математической разметки MathML (Mathematical Markup Language)).

К общим компонентам относятся контекстно-нейтральные, независимые от области охраны промышленной собственности компоненты, либо компоненты, имеющие отношение, по крайней мере, к двум видам объектов промышленной собственности.

Компоненты также классифицируются по уровням: Basic (базовый), Aggregate (агрегированный), Document (документальный). На рисунке 1 представлена логика построения компонентов. В таблице I приведено описание компонентов по уровням.

Базовые – это, в некотором смысле, простые компо-

ненты, а агрегированные состоят из базовых и/или других агрегированных. Документальные компоненты собираются из базовых, агрегированных и (или) других документальных компонентов и предназначены для использования при информационном обмене данными о каком-либо из видов объектов промышленной собственности.

Базовые и агрегированные компоненты могут относиться к общей категории (определяться в общем домене). Базовые, агрегированные и документальные компоненты могут относиться (определяться в соответствующем домене) к компонентам патентов, товарных знаков или промышленных образцов.

К документальным компонентам в домене (области) Patent относятся элементы:

- PatentPublication (зарегистрированный патент);
- ApplicationBody (содержание патентной заявки);
- BibliographicData (библиографические данные патента)

и соответствующие типы:

- PatentPublicationType;
- ApplicationBodyType;
- BibliographicDataType.

Структура элементов домена Patent представлена на рисунке 2.

К документальным компонентам в домене Trademark

относятся элементы:

TrademarkTransaction (полный набор (пакет) передаваемых при информационном взаимодействии сведений о товарных знаках, включая технологические сведения об отправке и получении, о возможных ошибках при взаимодействии, о типе транзакции);

TrademarkApplication (заявка на товарный знак);

TrademarkTransactionBody (набор передаваемых при информационном взаимодействии сведений о товарных знаках, включая сведения о возможных ошибках при взаимодействии, о типе транзакции);

TransactionContentBag (набор передаваемых при информационном взаимодействии сведений о товарных знаках, включая сведения о типе транзакции);

TransactionData (передаваемые при информационном взаимодействии сведения о товарных знаках)

и соответствующие типы:

- TrademarkTransactionType;
- TrademarkApplicationType;
- TrademarkTransactionBodyType;
- TransactionContentBagType;
- TransactionDataType.

Структура элемента TrademarkTransaction, включающего остальные документальные элементы домена, представлена на рисунке 3.

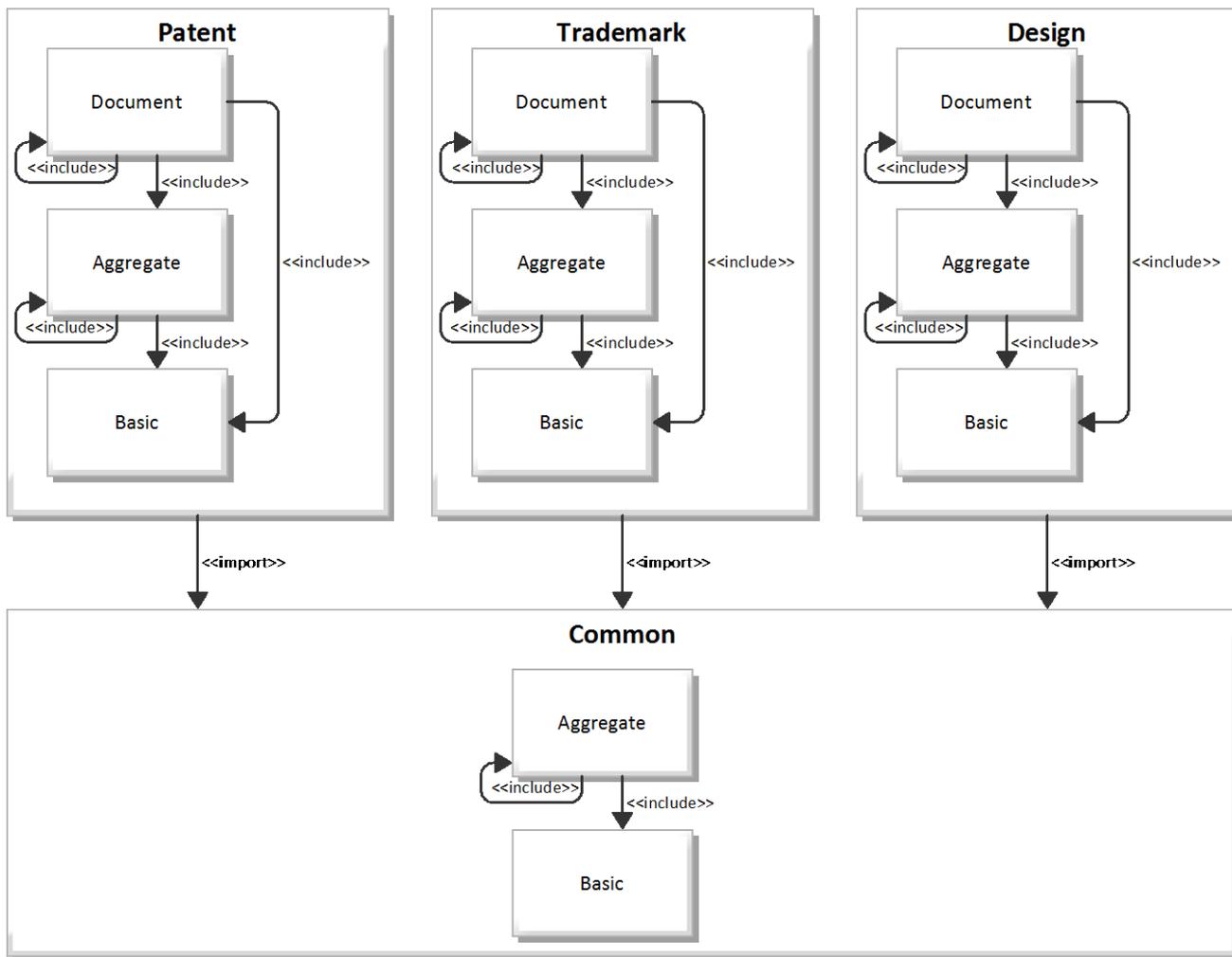


Рис. 1. Домены (области) и виды компонентов ST.96

Таблица I. Уровни компонентов модели в стандарте ST. 96

Уровень компонента	Описание	Примеры
Basic (базовый)	Базовый компонент задается через элемент и соответствующий тип: встроенный тип W3C XMLSchema (Built-in Datatype), простой тип (Simple Type) или сложный тип (Complex Type) с определением <code>xsd:simpleContent</code>	элемент <code>FirstName</code> , имеющий встроенный тип <code>xs:string</code> ; элемент <code>PhoneNumberCategory</code> , имеющий тип <code>PhoneNumberCategoryType</code>
Aggregate (агрегированный)	Агрегированный компонент – комбинация базовых и (или) иных агрегированных компонентов, имеющая вполне определенный предметный (бизнес-) смысл, но независимая от конкретной предметной области	<code>Name</code> (<code>NameType</code>), <code>Contact</code> (<code>ContactType</code>)
Document (документальный)	Документальный компонент включает базовые, агрегированные и / или иные документальные компоненты и отражает конкретную предметную область промышленной собственности (патенты, товарные знаки, промышленные образцы)	элемент <code>ApplicationBody</code> , основанный на типе <code>ApplicationBodyType</code> , представляет патентную заявку.

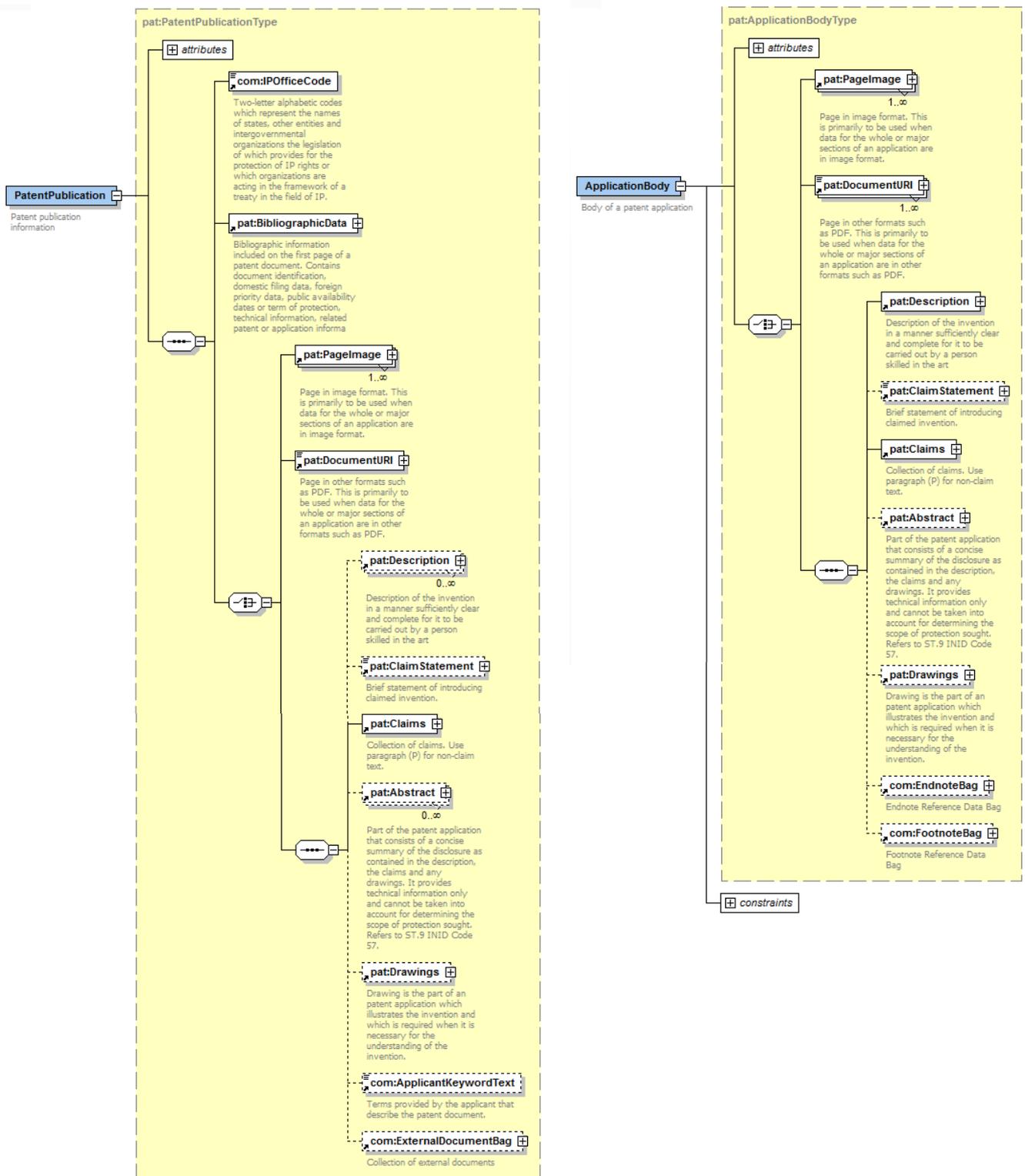


Рис. 2. Структура элементов домена Patent (создано с применением Altova XMLSpy 2017 free trial)

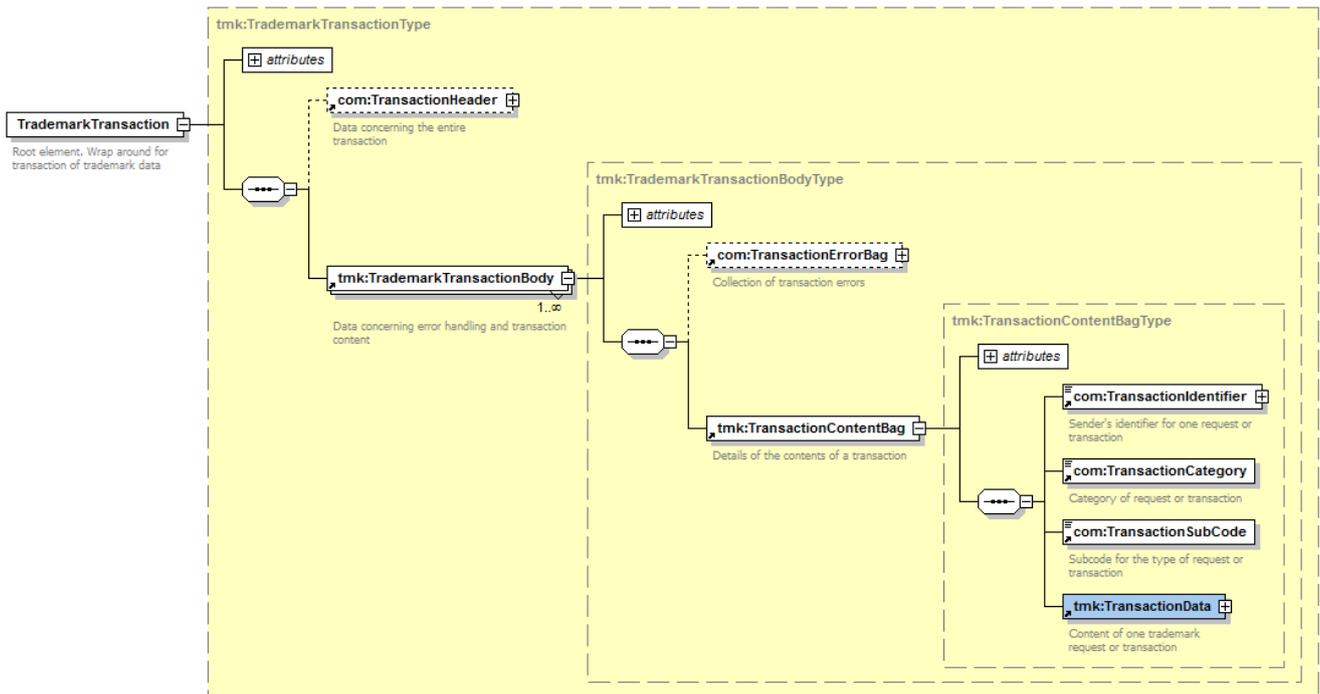


Рис. 3. Структура элемента TrademarkTransaction (создано с применением Altova XMLSpy 2017 free trial)

К документальным компонентам в домене Design относятся элементы:

`DesignTransaction` (полный набор (пакет) передаваемых при информационном взаимодействии сведений о промышленных образцах, включая технологические сведения об отправке и получении, о возможных ошибках при взаимодействии, о типе транзакции);

`DesignApplicationBag` (заявки на промышленные образцы);

`DesignApplication` (заявка на промышленный образец);

`DesignTransactionBody` (набор передаваемых при информационном взаимодействии сведений о промышленных образцах, включая сведения о возможных ошибках при взаимодействии, о типе транзакции);

`TransactionContentBag` (набор передаваемых при информационном взаимодействии сведений о промышленных образцах, включая сведения о типе транзакции);

`TransactionData` (передаваемые при информационном взаимодействии сведения о промышленных образцах)

и соответствующие типы:

`DesignTransactionType`;

`DesignApplicationType`;

`DesignApplicationBagType`;

`DesignTransactionBodyType`;

`TransactionContentBagType`;

`TransactionDataType`.

Структура элемента `DesignTransaction`, включающего остальные документальные элементы домена, представлена на рисунке 4.

Документальные элементы `TrademarkTransaction` и `DesignTransaction` доменов включают остальные документальные элементы домена, но последние, исходя из модульного принципа проектирования, могут применяться как самостоятельные компоненты.

В ST.66 и ST.86 представлены диаграммы классов в нотации языка UML, существенно облегчающие понимание структуры моделей, чего, к сожалению, нет в ST.36 и ST.96. Предполагаем, что в рамках соответствующих целевых (рабочих) групп подобные диаграммы используются, но в открытом доступе они не представлены.

Также, по крайней мере, в открытом доступе отсутствуют концептуальная или онтологическая модели объектов промышленной собственности, хотя на одной из встреч XML4IP (<http://www.wipo.int/export/sites/www/cws/en/taskforce/xml4ip/docs/ohim-20071025.ppt>) были озвучены намерения соответствующей разработки.

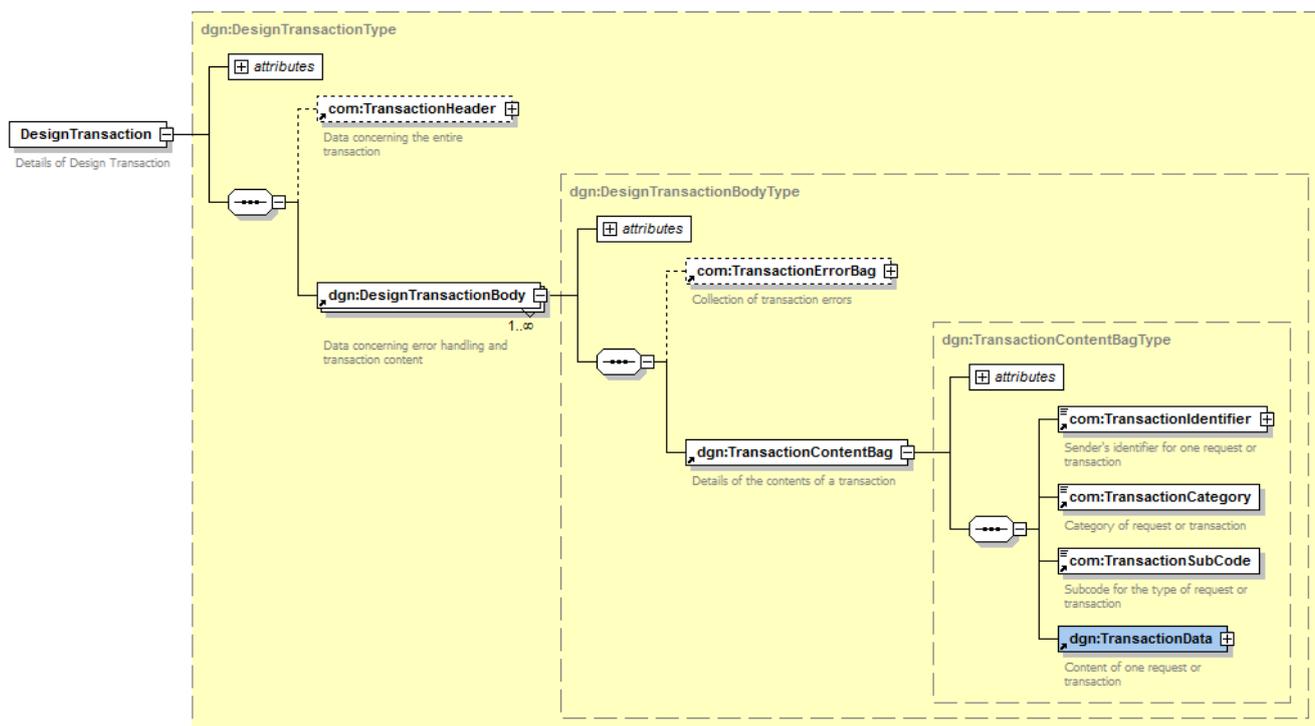


Рис. 4. Структура элемента DesignTransaction (создано с применением Altova XMLSpy 2017 free trial)

IV. ОСОБЕННОСТИ МОДЕЛИ

К особенностям модели авторы относят различные технологические нюансы, возможно различной степени значимости, но которые бросились в глаза при изучении модели.

1. В Common Components есть элементы B (Bold text), Br (Forced line break), LI (List item) и др., используемые для разметки текста, и не несущие никакого прикладного смысла.

Наличие таких элементов в модели, по-видимому,

объясняется необходимостью работы с большими неструктурированными текстовыми документами.

2. В стандарте ST.96 перечислены существующие шаблоны (паттерны) проектирования XML-схем [7], и необходимостью повторного использования типов и элементов обоснован выбор шаблона Garden of Eden (райский сад).

В стандартах ST.66 и ST.86 нет упоминаний о шаблонах проектирования, но судя по XML-схемам (см. листинг 1) в них применялся смешанный шаблон.

```

<xs:complexType name="BasicRecordType">
  <xs:sequence>
    <xs:element name="RecordIdentifier" type="xs:string" minOccurs="0"/>
    <xs:element name="RecordFilingDate" type="xs:date" minOccurs="0"/>
    <xs:element name="RecordNotificationDate" type="xs:date" minOccurs="0"/>
    <xs:element name="RecordEffectiveDate" type="xs:date" minOccurs="0"/>
    <xs:element name="BasicRecordKind" type="BasicRecordKindType" minOccurs="0"/>
    <xs:element name="RecordReference" type="xs:string" minOccurs="0"/>
    <xs:element name="RecordLanguageCode" type="ExtendedISOLanguageCodeType" minOccurs="0"/>
    <xs:element name="TradeMarkKeyDetails" minOccurs="0">
      <xs:complexType>
        <xs:sequence>
          <xs:element name="TradeMarkKey" type="KeyType" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:element name="RecordPublicationDetails" minOccurs="0">
      <xs:complexType>
        <xs:sequence>
          <xs:element name="RecordPublication" type="PublicationType" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
  </xs:sequence>
</xs:complexType>

```

```
</xsd:sequence>
</xsd:complexType>
```

Листинг 1. Фрагмент XML-схемы ST.66

В листинге 2 приведен соответствующий фрагмент XML-схемы стандарта ST.96. См. элемент

RecordPublicationDetails в листинге 1 и его аналог RecordPublicationBag в листинге 2.

```
<xsd:complexType name="BasicRecordType">
  <xsd:sequence>
    <xsd:element ref="com:RecordIdentifier"/>
    <xsd:element ref="com:RecordFilingDate" minOccurs="0"/>
    <xsd:element ref="com:RecordNotificationDate" minOccurs="0"/>
    <xsd:element ref="com:RecordEffectiveDate" minOccurs="0"/>
    <xsd:element ref="tmk:BasicRecordCategory" minOccurs="0"/>
    <xsd:element ref="com:RecordReference" minOccurs="0"/>
    <xsd:element ref="com:RecordLanguageCode" minOccurs="0"/>
    <xsd:element ref="tmk:TrademarkIdentification" minOccurs="0"/>
    <xsd:element ref="tmk:RecordPublicationBag" minOccurs="0"/>
  </xsd:sequence>
  ...
<xsd:complexType name="RecordPublicationBagType">
  <xsd:sequence>
    <xsd:element ref="tmk:RecordPublication" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>
...
<xsd:element name="RecordPublication" type="tmk:PublicationType">
  <xsd:annotation>
    <xsd:documentation>Data regarding a record publication</xsd:documentation>
  </xsd:annotation>
</xsd:element>
...
<xsd:complexType name="PublicationType">
  <xsd:sequence>
    <xsd:element ref="com:PublicationIdentifier"/>
    <xsd:element ref="tmk:PublicationSectionCategory" minOccurs="0"/>
    <xsd:element ref="com:PublicationSubsection" minOccurs="0"/>
    <xsd:element ref="com:PublicationDate" minOccurs="0"/>
    <xsd:element ref="com:PublicationPage" minOccurs="0"/>
    <xsd:element ref="tmk:NationalPublication"/>
  </xsd:sequence>
</xsd:complexType>
```

Листинг 2. Фрагмент XML-схемы ST.96

V. ЗАКЛЮЧЕНИЕ

Стандарты ВОИС способствуют развитию информационного взаимодействия в сфере охраны интеллектуальной собственности. Стандарты направлены на гармонизацию данных о различ-

ных объектах интеллектуальной собственности, а также на применение современных решений в области информационных технологий.

БИБЛИОГРАФИЯ

- [1] List of WIPO Standards. http://www.wipo.int/standards/en/part_03_standards.html (дата обращения: 17.03.2017)
- [2] Федеральная служба по интеллектуальной собственности (Роспатент). Стандарты ВОИС http://www.rupto.ru/docs/interdocs/stand_wipo (дата обращения: 17.03.2017)
- [3] Корчагин А.Б., Лисьих И.Г., Никифоров Д.А., Сиваков Р.Л. Модели данных для информационного взаимодействия. International Journal of Open Information Technolo-

- gies ISSN: 2307-8162 vol. 5, no.3, 2017
- [4] Global IP Infrastructure http://www.wipo.int/global_ip/en/ (дата обращения: 17.03.2017)
- [5] Global Dissemination of IP Data Initiative: Data Specifications http://www.wipo.int/global_ip/en/ip_data_initiative/data_specifications.html (дата обращения: 17.03.2017)
- [6] ISO/IEC 19757-3:2016. Information technology – Document Schema Definition Languages (DSDL) – Part 3: Rule-based validation – Schematron
- [7] Шилова А. Шаблоны проектирования XSD. <https://habrahabr.ru/company/cit/blog/259167/> (дата обращения: 17.03.2017)

Data model of the World Intellectual Property Organization standard ST.96

A.B. Korchagin, I.G. Lisikh, D.A. Nikiforov, R.L. Sivakov

Abstract – The publication continues the series of articles devoted to the analysis of approaches to data modeling and revealing their key technological features. This article describes the model recommended by the World Intellectual Property Organization standard ST.96. The evolution of WIPO standards for data processing with the development of data exchange technics is reviewed. The general logic of ST.96 model, as well as the logic of its components in particular (elements, types and attributes) is presented. A comparison of data structures dealing with different objects of industrial property such as patents, trademarks and industrial designs is performed. Some technical features of ST.96 model, such as elements for text markup and XML-schemes design patterns are reviewed.

Keywords – information exchange; data model; data model compatibility; data model mapping; ontology, the World Intellectual Property Organization, XML.