

УДК 528:004.057.2

К ВОПРОСУ О НАЦИОНАЛЬНОМ ПРОФИЛЕ МЕТАДАНЫХ ПРОСТРАНСТВЕННЫХ ДАННЫХ

Салтовец А.А.

АО «СПАЭРО Плюс», г. Харьков, Украина

В статье рассмотрен международный стандарт метаданных пространственных данных ISO 19115. Предложены пути и способы формирования на его основе национального профиля метаданных
Ключевые слова: стандарт, метаданные, пространственные данные, профили

Как показывает сегодняшний мировой опыт, определяющим для успешного и быстрого развития рынка пространственных данных и сопутствующих сервисов является создание и развитие национальных инфраструктур пространственных данных [1].

Одной из ключевых составляющих, без которых невозможно функционирование таких инфраструктур, являются метаданные пространственных данных и соответствующие сервисы их предоставления и использования.

Проблема поиска нужной информации в большом количестве доступных данных возникла еще тогда, когда пространственные данные не имели столь широкого распространения. Актуальность этой проблемы приобретала все большее значения с развитием служб каталогов документов (библиотеки, архивы и пр.).

В 1995 г. сотрудниками OCLC (Online Computer Library Center) в городе Дублин в штате Огайо была разработана модель метаданных для описания широчайшего диапазона сетевых ресурсов. Эта модель получила название Дублинского ядра (Dublin Core). Первоначально она включала в себя 15 элементов описания ресурса, которые обеспечивали поиск данных по таким параметрам, как заголовок, автор, предмет, дата, авторские права и т.п.[2]

Эта модель постоянно развивается и сегодня включает в себя два уровня описания, дополнительные квалификаторы и пр. На основе Дублинского ядра разработан ряд стандартов и спецификаций, в том числе для языка XML. Международной организацией по стандартизации принят стандарт ISO 15836 «Информация и документация – Набор элементов метаданных «Дублинского ядра». В ряде стран (Австралия, Швеция) модель Дублинского ядра принята в качестве национального стандарта.

Появление и все большее распространение такого специфического ресурса как пространственные данные потребовало существенного пересмотра концептуального подхода к модели метаданных. Если Дублинское ядро достаточно полным образом способно описать практически любой документ, то для описания пространственных данных его существенно не хватает.

Это связано с наличием у пространственных данных таких свойств как пространственный экстенд, масштаб, проекция и т.п. Таким образом, прямое

применение Дублинского ядра к описанию пространственных данных невозможно из-за неоднозначности результатов поиска.

Потребность в описании пространственных данных привела к разработке ряда моделей метаданных, которые разрабатывались как организациями, так и группами разработчиков – наиболее известными из них являются модели, разработанные FGDC и ESRI.

Результаты этих работ были использованы при разработке международного стандарта метаданных пространственных данных, выполненной совместно Консорциумом открытых ГИС (OGC) и техническим комитетом 211 «Географическая информация/Геоматика» Международной организации по стандартизации (ISO). Эти работы были проведены под руководством сотрудника ESRI Дэвида Данко (David Danko). Полное название стандарта – «ISO 19115-2003 Geographic information – Metadata».

Как и остальные стандарты серии ISO 19100, этот стандарт по отношению к метаданным представляет собой описание их общей концептуальной модели, выполненное на языке концептуальных схем, в качестве которого принят UML.

Стандарты серии ISO 19100 разрабатываются в соответствии с единым подходом, определенным в стандарте ISO 19101-2002 Эталонная модель (Reference model). Этот подход определяет создание стандарта как абстрактной, в определенном смысле всеобъемлющей модели определенной предметной сферы.

Для реализации национальной инфраструктуры пространственных данных необходима разработка ряда государственных нормативных документов и стандартов. При этом необходима их гармонизация с международными стандартами. Одним из первоочередных для разработки стандартов является стандарт метаданных пространственных данных.

Целью настоящей статьи является рассмотрение возможных путей создания национального стандарта метаданных, вариантов его информационной структуры и содержательного наполнения.

Сформулируем самые общие требования к национальному стандарту метаданных:

- этот стандарт должен соответствовать международному стандарту для обеспечения использования «наших» метаданных пользователями из других стран;
- этот стандарт должен обеспечивать понимание и использование метаданных всеми пользователями в государстве вне зависимости от предметной области их интересов;
- этот стандарт должен служить инструментом для создания систем метаданных разных предметных областей (так как пространственные данные в геодезии описываются иначе, чем в геологии или трубопроводном транспорте)

На первый взгляд, эти требования в значительной степени взаимно противоречивы. Действительно, если в качестве национального стандарта принять перевод на национальный язык самого ISO 19115, снабдив его соответствующим национальным введением и приведя к требованиям государственной системы

стандартизации, то требования 1 и 3 будут соблюдены. Но при этом, созданные на основе такого стандарта ведомственные нормативы могут существенно отличаться как по структуре, так и по содержанию.

Это приведет к тому, что программные системы, создаваемые в рамках какой-то одной предметной области (а, соответственно, и их пользователи), не смогут оперировать с метаданными другой нормативной сферы.

С другой стороны, если разработать полностью собственный «независимый» стандарт (значительно более простой по сравнению с международным), легко можно удовлетворить требования 2 и 3, но «наши» метаданные не будут понятны миру, а значит и наши пространственные данные останутся невостребованными.

Вернемся к рассмотрению стандарта ISO как абстрактной (концептуальной) модели предмета стандартизации. Этот подход подразумевает создание конкретных реализаций общей модели в виде профилей, которые учитывают специфические требования и нужды отдельных сообществ пользователей (государств, ведомств, корпораций, профессиональных групп и т.п.).

Правила создания профилей стандартов определены стандартом ISO 19106 Profiles. Кроме того, в самом стандарте ISO 19115 кратко описана методика создания профилей, добавления собственных объектов и элементов и тестирования полученного профиля на соответствие стандарту ISO.

Таким образом общая модель взаимодействия стандарта ISO и его профилей выглядит как следующая концептуальная схема (Рис. 1).

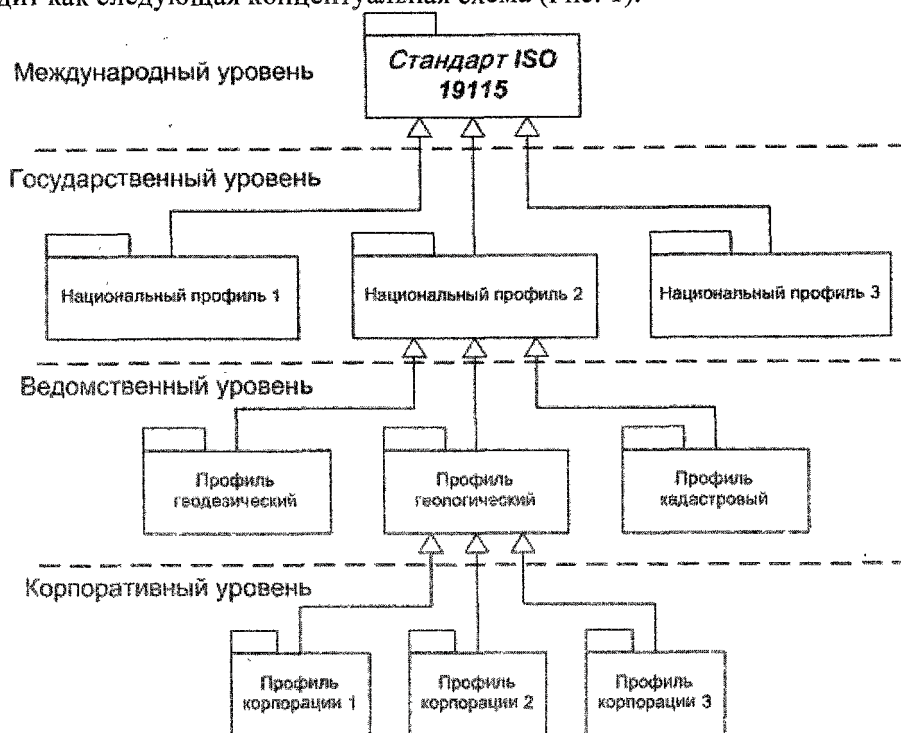


Рис. 1 Профили стандарта ISO

Из этой схемы видно, что профили каждого нижележащего уровня являются потомками вышележащего профиля и связаны с ним отношением генерализации. Иначе говоря, нижележащий профиль наследует определенные свойства и элементы родительского объекта, добавляя к ним свои собственные.

Таким образом, применение собственно стандарта ISO 19115 в качестве государственного стандарта не является оптимальным, так как позволяет создать ведомственные или корпоративные профили столь различные по своим свойствам и содержанию, что их совместное использование в рамках одного государства будет или невозможным или крайне затруднительным.

Рекомендуемым путем является разработка национального профиля метаданных и принятие его в качестве государственного стандарта. С одной стороны он будет являться конкретной реализацией международного стандарта, что обеспечит понимание «наших» метаданных мировым сообществом. С другой же стороны – явится прочной основой для разработки структур метаданных ведомственного и корпоративного уровней.

По такому пути пошли авторы российского стандарта метаданных – ГОСТ Р 52573-06 «Географическая информация. Метаданные», введенного в действие в конце 2006 г. [4]

Следующей «развилкой» на пути создания национального стандарта является вопрос о его структуре. Сопоставляя структуру ISO 19115 и ГОСТ Р 52573, можно отметить, что в российском профиле метаданных отсутствует понятие дочерних профилей, правил и методов их создания, тестирования и регистрации.

По мнению автора статьи это существенным образом обедняет возможности применения такого стандарта и может привести впоследствии к значительным различиям в «локальных» наборах метаданных в разных ведомствах. Возможно, что этот недостаток будет ликвидирован принятием стандарта, соответствующего ISO 19106 «Профили». Однако, его разработка и ввод в действие потребует определенного времени, что может привести к необходимости переделывать уже созданные программные приложения и массивы метаданных.

Следует отметить, что несмотря на наличие ISO 19106, в котором методика и технология создания профилей детально определена, в ISO 19115 тем не менее приведены и краткое руководство по созданию профилей и метод их тестирования. Таким образом, стандарт является методологически самодостаточным и требует обращения к связанному стандарту только для более углубленного понимания процесса создания профиля.

Краткое описание языка UML помещено в российском стандарте в отдельное справочное приложение, тогда как в ISO 19115 оно присутствует в более краткой

форме в разделе «Символы и сокращения». С точки зрения улучшения понимания стандарта российский вариант является более эффективным, учитывая общий уровень компетентности широкого круга пользователей. Возможно, что само описание в украинском стандарте стоило бы сделать еще более широким.

Таким образом, структура содержательных разделов национального стандарта метаданных может выглядеть следующим образом (Табл. 1)

Таблица 1

Информационная структура стандарта

№	Раздел	Подраздел	Содержание
1	Область применения		Область применения стандарта в предметном и юридическом отношении
2	Соответствие		Требования соответствия метаданных конкретным разделам стандарта
3	Нормативные ссылки		Ссылки на связанные стандарты
4	Термины и определения		Термины, используемые в стандарте и их определения
5	Символы и сокращения		Пояснение используемых сокращений
6	Требования		
6.1		Общие требования к метаданным	Информация о применении метаданных
6.2		Пакеты и объекты метаданных	Информация о пакетах и объектах метаданных и их соответствии UML-классам и типам данных
6.3		Описание пакетов	Краткое текстовое описание каждого пакета
6.4		Ядро метаданных	Описание минимального рекомендуемого набора элементов метаданных
6.5		Расширения и профили	Краткое изложение методики построения профилей и расширений, их тестирования
A	Приложение A (обязательное) Схемы метаданных		UML-диаграммы каждого пакета метаданных со ссылкой на словарь данных

Продолжение таблицы 1

№	Раздел	Подраздел	Содержание
Б	Приложение Б (обязательное) Словарь данных		
Б1		Структура словаря	Пояснения элементов структуры словаря данных
Б2		Пакеты метаданных	Табличное представление словаря каждого пакета метаданных
Б3		Типы данных	Табличное представление каждого типа данных
В	Приложение В (обязательное) Расширения и профили		Детальное описание методики создания расширений и профилей
В1		Общие положения	
В2		Типы расширений	
В3		Правила создания расширений	
В4		Общий профиль	
В5		Правила создания профиля	
В6		Тестирование профилей и расширений	
Г	Приложение Г (информационное) Краткое описание языка UML		

Естественно, что предложенная структура учитывает только самые необходимые информационные блоки и в процессе разработки стандарта может изменяться и дополняться.

Рассмотрим теперь собственно содержание стандарта ISO 19115 с точки зрения построения национального профиля.

Полная концептуальная схема модели метаданных выглядит следующим образом (Рис. 2).

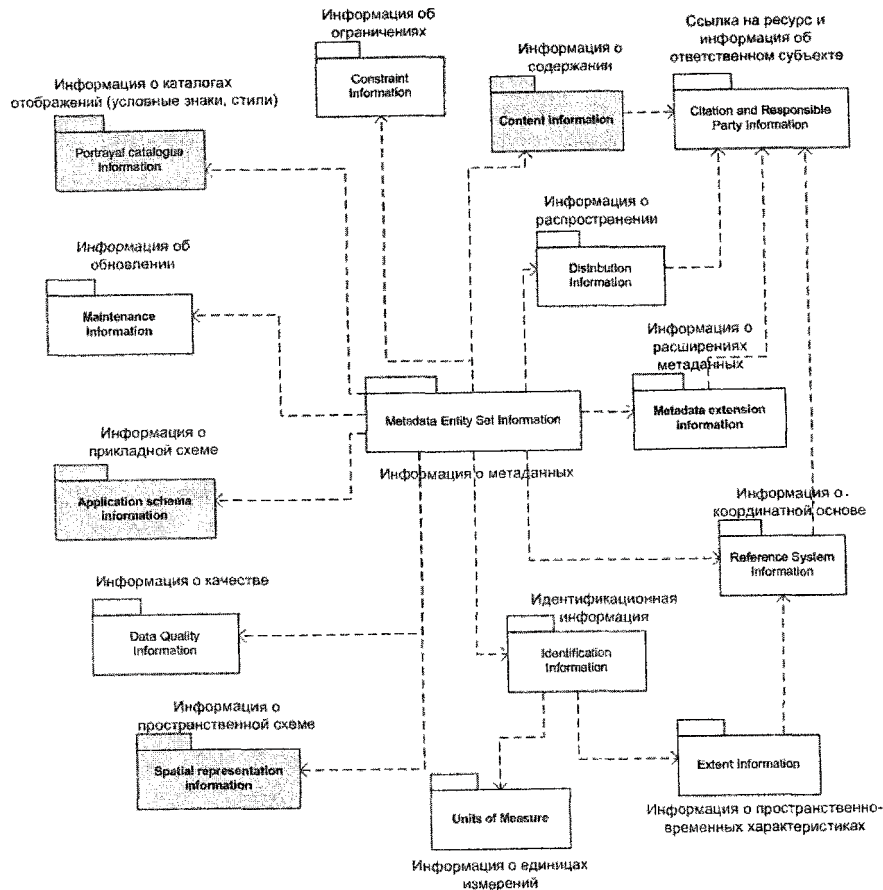


Рис. 2 Полная концептуальная схема метаданных

Рассмотрение содержания пакетов показывает, что ряд пакетов полной концептуальной схемы можно исключить из национального профиля. На Рис. 2 эти пакеты закрашены серым цветом.

Остановимся на них подробнее:

Пакет «Информация о пространственной схеме» содержит информацию о механизмах и пространственно-временных схемах, которые используются для представления пространственной информации в наборе данных. Поскольку в нашей стране используются только трехмерные координаты, а многомерные сетки и покрытия в ограниченном объеме используются только в области геологических наук, этот пакет является излишним для общенационального профиля. При этом отдельные сущности этого пакета могут быть включены в расширение национального профиля, например, для геологических дисциплин.

Пакет «Информация о содержании» включает в себя информацию о каталогах пространственных объектов, в которых должны быть описаны объекты, содержащиеся в наборе данных. В Украине таких каталогов нет и их создание в ближайшее время маловероятно. Кроме этого в пакете содержатся информационные блоки, описывающие тип покрытия (изображение, физические измерения и пр.) и характеристики изображений (угол освещенности, параметры радиометрической

коррекции и т.п.). Поскольку эта информация либо неприменима к нашим данным, либо является излишне детализированной, этот пакет также является излишним.

Хотелось бы попутно отметить, что сам стандарт метаданных, по мнению разработчиков, не позволяет детально описать такой тип пространственных данных, как снимки. В настоящий момент разрабатывается сопутствующий стандарт ISO 19115-2 полностью посвященный метаданным изображений.

Для нашей страны, которая только начинает использовать метаданные для упорядочения своих пространственные данные, возможностей базового стандарта вполне достаточно.

Пакет «Информация о каталогах отображений» содержит описание ссылки на те наборы стилей оформления (в том числе и условных знаков), которые используются в описываемом наборе данных. Поскольку в нашей практике такие каталоги отсутствуют, то и этот пакет можно исключить из профиля.

Пакет «Информация о прикладной схеме» содержит в себе сущности, описывающие прикладную схему набора данных на каком-либо формальном языке (предпочтительнее – UML). Поскольку использование возможностей UML для разработки схем данных и их документирования в среде производителей пространственных данных у нас крайне мало распространено этот пакет также можно считать излишним.

Отбор на уровне пакетов относительно прост и не требует больших временных затрат. Но для разработки полноценного национального профиля метаданных, который содержал бы необходимые и в тоже время минимально достаточные сущности, нужно провести квалифицированный отбор и на уровнях объектов, элементов и типов данных.

С начала 2006 года возобновил свою работу ТК103 – технический комитет стандартизации «Географическая информация/Геоматика» под руководством проф. Ю.А. Карпинского. В ближайших планах комитета предусмотрена подготовка, в частности, и стандарта метаданных.

Список литературы

1. Карпінський Ю., Лященко А., Стратегія формування національної інфраструктури просторових даних в Україні - Київ: УкрДАГП, 2006. - 107 с.
2. Армс В. Электронные библиотеки: Пер. с англ.-М.:ВИНИТИ, 2001.-274 с.
3. Андрианов В. Стандарты в ИПД: - М.:ArcReview № 2 (37) 2006
4. ГОСТ Р 52573-06 Географическая информация. Метаданные – М.: ИПК Издательство стандартов, 2006. – 48 с.
5. Ромащенко М.І., Драчинська Е.С., Шевченко А.М. Інформаційне забезпечення зрошувального землеробства. – Київ: Аграрна наука, 2005. – 191 с.

Салтовец О. О. До питання про національний профіль метаданих просторових даних // Вчені записки Таврійського національного університету ім. В. І. Вернадського. – 2007. – Серія «Географія». - Т. 20 (59).- № 1. - С. 183-190.

У статті розглянуто міжнародний стандарт метаданих просторових даних ISO 19115. Запропоновані шляхи та способи формування на його основі національного профілю метаданих

Ключові слова: Стандарт, метадані, просторові дані, профілі

Saltovets A. A. To a question on national metadata profile for spatial data // Uchenye zapiski Tavricheskogo Natsionalnogo Universiteta im. V. I. Vernadskogo. – 2007. – Series «Geography». – V. 20 (59). - № 1. – С. 183-190.

In article the international standard ISO 19115 Geographic information - Metadata is considered. Ways and methods of formation on its basis of a national structure of metadata are offered.

Keywords: The standard, the metadata, the spatial data, the profile

Поступила в редакцію 20.04.2007г.