Ученые записки Таврического национального университета имени В.И.Вернадского
Серия «География». Том 21 (60). 2008 г. № 1. С. 132-141

*УДК 528:004.057.2*

# МЕТОДИЧЕСКИЕ ВОПРОСЫ РЕАЛИЗАЦИИ МЕТАДАННЫХ НА ОСНОВЕ ПРОФИЛЯ СТАНДАРТА ISO 19115

***Салтовец А.А., Николаев В.М.***

В статье рассматривается методика построения системы поддержки метаданных пространственных данных, базирующаяся на группе Международных стандартов серии ISO 19100. Предложены пути и средства реализации системы на государственном и нижележащих уровнях использования пространственных данных.
*Ключевые слова:* Метаданные, стандарты, профили, пространственные данные, XML

Сегодня уже нет необходимости пропагандировать или разъяснять роль и место глобальных и национальных инфраструктур пространственных данных в современной жизни. Они успешно работают и решают свою основную задачу – обеспечивают функциональную совместимость пространственных данных, удобство и простоту их использования. Решение этой задачи базируется на использовании метаданных пространственных данных. Программные системы поддержки метаданных обеспечивают поиск данных, определяют способы и методы доступа к ним, а зачастую и способы их использования.

Рассмотрим методологическую основу создания инфраструктуры пространственных данных (далее по тексту – ИПД). Для создания ИПД необходимо создать сами пространственные данные, описать их – то есть создать их метаданные, а также обеспечить средства и методы поиска и доступа к нужным данным.

Метаданные служат основой и ключевым элементом двух последних процессов. Сами данные, без их описания и возможности опоискования – мертвы и недоступны (или труднодоступны) для пользователей. Но для того, чтобы метаданные могли обеспечивать все, что от них требуется, они должны быть одинаково понимаемы сообществом пользователей и не зависеть от форматов и структур данных или конкретных программных средств по работе с ними. Выполнение этих требований обеспечивается использованием Международного стандарта ISO 19115.

Бурное развитие пространственных данных и расширение спектра их использования в жизни общества в последние годы обуславливает столь же активные усилия по их стандартизации. В Международной Организации по стандартизации (ISO) стандартами в области геоинформатики (геоматики) занимается технический комитет ТС 211, который разрабатывает серию стандартов 19000. Сегодня стандарты этой серии – наиболее наукоемкие и системно построенные из всего множества Международных стандартов. Они полностью ориентированы на информационные технологии. В основу стандартов этой серии

положено понятие эталонной модели стандарта предметной области, которая дает методы описания предмета стандартизации и требований к нему [1].

В соответствии с понятием эталонной модели каждый стандарт этой серии является концептуальной моделью предметной (в смысле стандартизации) области. Для единства описания и простоты применения стандартов в качестве языка для описания концептуальной модели используется универсальный язык моделирования UML [2].

Совокупность стандартов этой серии представляет собой как бы единую виртуальную модель географической (пространственной) информации. Сущности, определенные в одном стандарте, могут быть с легкостью использованы в модели иной области стандартизации. Объектно-ориентированный подход к описанию стандарта позволяет использовать наследование, полиморфизм и инкапсуляцию при создании таких моделей.

Стандарт ISO 19115 занимает в серии одно из центральных мест. И это понятно, так как для описания пространственных данных нужно указать и описать все их свойства и особенности, определяемые в других стандартах серии 19000 [3]. Таким образом ISO 19115 как бы объединяет все остальные стандарты и использует их сущности в своей модели.

Что же представляет собой концептуальная модель метаданных? Корневая UML-диаграмма модели представлена на Рис. 1

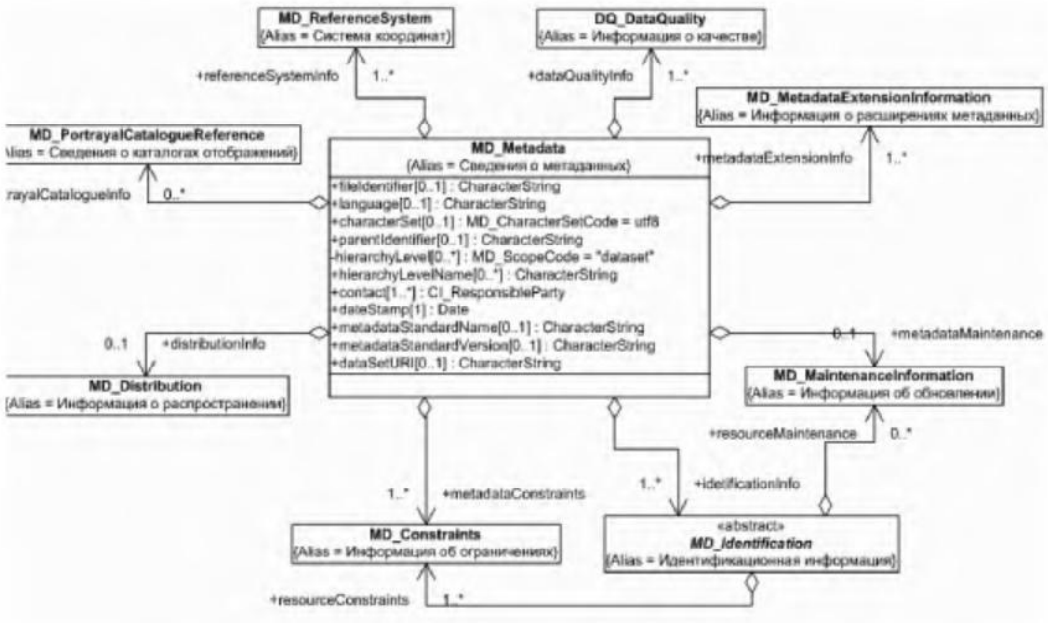


Рис. 1 Сведения о метаданных

Каждый из присутствующих в модели классов представлен своими собственными диаграммами разной степени сложности. Например, информация о распространении выглядит следующим образом (Рис. 2):

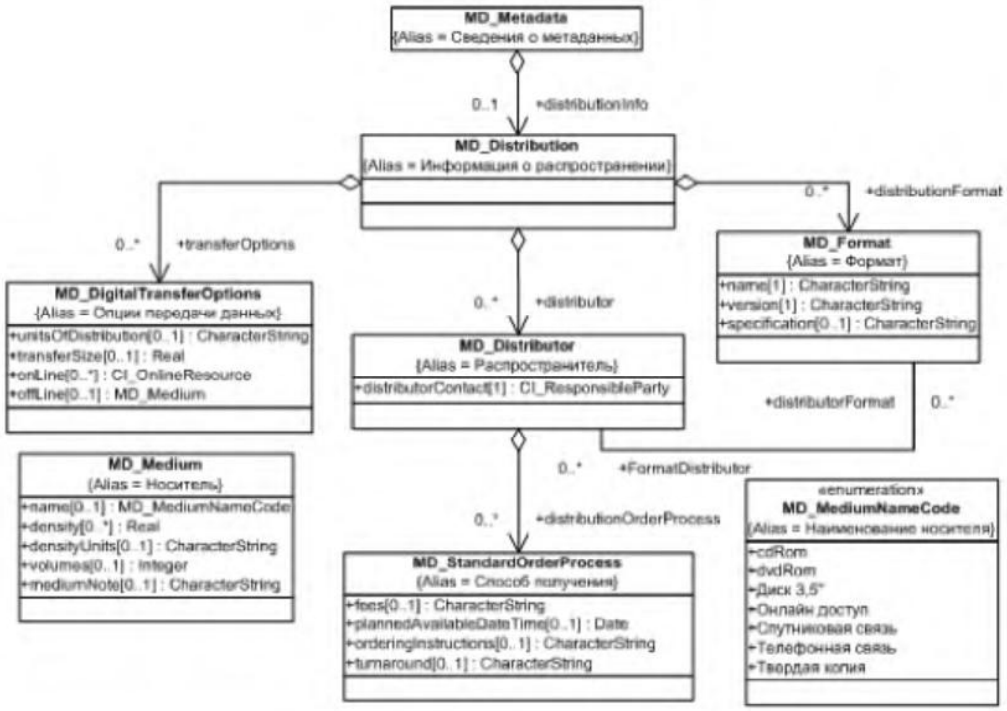


Рис. 2 Информация о распространении

В так называемой всеобъемлющей концептуальной модели метаданных ISO 19115 содержится 16 диаграмм, которые снабжены соответствующим словарем данных.

Изучив все эти диаграммы, автор пространственных данных начинает понимать, что же должно (или может) содержаться в метаданных его данных. Но как это реализовать на практике? Стандарт не дает ответа на этот вопрос – он предназначен для иной цели. А вот другие стандарты из этой же серии как раз и посвящены реализации метаданных на практике.

В простейшем случае метаданные пространственных данных могут быть представлены в виде текстового описания. Но при этом, естественно, возникает целый ряд проблем – в каком формате хранить, как «привязать» к самим данным, как вести поиск, как обеспечить не только «человеческое», но и «машинное» понимание этой информации. Все эти проблемы решаются с помощью языка XML - Extensible Markup Language (Расширяемый Язык разметки). Этот язык и принят в качестве стандартного для представления метаданных.

Так как все стандарты серии 191000 представляют собой концептуальные UML-модели предметной области, то для части из них (в частности для метаданных) нужны правила и методы превращения их из логических абстракций в реальные кодированные описания, которые можно использовать в информационных технологиях.

Общие правила кодирования географической информации в XML-схемы описаны в стандарте ISO 19118 [4]. Детальному же описанию методов кодирования и специфических приемов их применения для UML-моделей метаданных посвящена техническая спецификация ISO 19139 [5].

Географические метаданные представлены в ISO 19115 как ряд пакетов UML, содержащих один или более UML-классов. ISO 19115 обеспечивает универсальное, независимое от кодирования представление метаданных пространственной информации. А ISO 19139, в свою очередь, обеспечивает универсальную реализацию ISO 19115 через кодирование XML-схемы, которое соответствует правилам, описанным в ISO 19118.

Одним из основных понятий XML является понятие пространства имен. Пространство имен – это коллекция имен, идентифицированных ссылкой на URI (Uniformed Resource Identifier – Унифицированный Идентификатор Ресурса), которые используются в документах XML как имена элементов и имена атрибутов.

ISO 19139 определяет следующие пространства имен:

gco Geographic Common extensible markup language (http://www.isotc211.org/2005/gco)
gmd Geographic MetaData extensible markup language (http://www.isotc211.org/2005/gmd)
gmx Geographic Metadata XML Schema (http://www.isotc211.org/2005/gmx)
gss Geographic Spatial Schema extensible markup language (http://www.isotc211.org/2005/gss)
gsr Geographic Spatial Referencing extensible markup language (http://www.isotc211.org/2005/gsr)
gts Geographic Temporal Schema extensible markup language (http://www.isotc211.org/2005/gts)

Кроме них используются следующие внешние пространства имен:

gml Geography Markup Language (http://www.opengis.net/gml)
xlink XML Linking Language (http://www.w3.org/1999/xlink)
xs W3C XML base (http://www.w3.org/2001/XMLSchema)

Основным для метаданных является пространство имен gmd, которое и используется для кодирования классов и атрибутов метаданных.

XML-схема предлагает много альтернатив для того, чтобы структурировать информацию для обмена. ISO 19118 определяет набор правил кодирования для преобразования концептуальных UML-схем, содержащихся в документах серии ISO 19100, в XML-схемы. Но даже с учетом ограничений ISO 19118 существуют вариан-

ты при создании определенных XML-схем. ISO 19139 описывает детали кодирования в XML-схемы концептуальной UML-схемы ISO 19115.

Приведем простейший пример использования правил кодирования:

ISO 19118 определяет, что фундаментальным понятием моделирования в UML является класс. В результате фундаментальные правила кодирования фокусируются на кодировании класса UML и строятся исходя из этого. Важно отметить, что в ISO 19118 свойство представляет собой пару «имя-значение». Оно может представлять атрибут, ассоциацию, агрегацию или композицию. Класс состоит из одного или более свойств. Например, на Рис. 3 класс ***Class1*** имеет три свойства: ***attr1***, ***attr2*** и ***role1***. Для кодирования в XML-схему важно понять, что нет никакого различия между свойствами, которые являются атрибутами UML, ассоциациями, агрегациями или композициями.

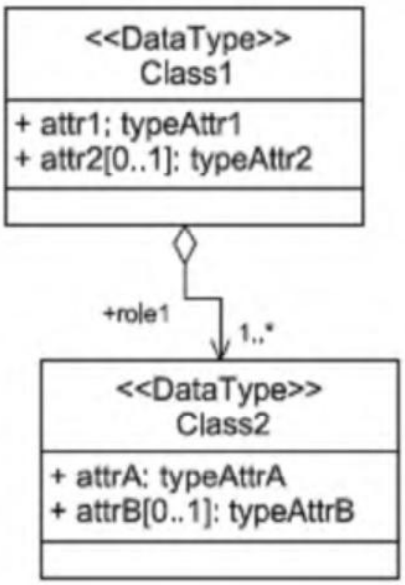


Рис. 3 Пример UML

Класс UML кодируется в XML-схему как сложный тип XML: **xs:complexType**. Этот сложный тип XML в дальнейшем именуется **XML Class Type (XCT)**. В XML-схеме каждый XCT имеет атрибут имени, значение которого – имя класса с суффиксом **_Type:**

```
<xs:complexType name="Class1_Type">
   (...)
</xs:complexType>
```

Все классы UML по умолчанию будут иметь сложное содержание. Для обеспечения этого элемент **xs:complexType** включает в себя элемент **xs:complexContent:**

```
<xs:complexType name="Class1_Type">
```

```
    <xs:complexContent>
      (...)
    </xs:complexContent>
</xs:complexType>
```

ISO 19118 также указывает на необходимость использования в XML-схеме идентификаторов (**id**) и предписывает использовать для этого универсальные уникальные идентификаторы (**uuid**). Для обеспечения необходимых идентификаторов в пространстве имен **gco** существует специальный тип XML-схемы **gco:AbstractObject_Type.**

Все классы UML по умолчанию расширяют **gco:AbstractObject_Type**, добавляя элемент **xs:extension** с атрибутом ***base***, равным **gco:AbstractObject_Type:**

```
<xs:complexType name="Class1_Type">
    <xs:complexContent>
      <xs:extension base="gco:AbstractObject_Type">
        (...)
      </xs:extension>
    </xs:complexContent>
</xs:complexType>
```

Все классы UML по умолчанию будут иметь последовательность, содержащую все свойства класса. Это обеспечивается добавлением элемента **xs:sequence**, содержащего элементы **xs:element** для каждого свойства класса. Атрибутами элемента **xs:element** будут:

- атрибут ***name,*** значением которого является имя свойства;
- атрибут ***type***, значением которого является имя класса с добавлением «**_PropertyType**»;
- атрибуты ***minOccurs*** и ***maxOccurs***

```
<xs:complexType name="Class1_Type">
    <xs:complexContent>
      <xs:extension base="gco:AbstractObject_Type">
        <xs:sequence>
          <xs:element name="attr1" type="ns1:typeAttr1_PropertyType"/>
          <xs:element name="attr2" type="ns1:typeAttr2_PropertyType"
minOccurs="0" />
          <xs:element name="role1" type="ns1:Class2_PropertyType"
minOccurs="1" maxOccurs="unbounded"/>
        </xs:sequence>
      </xs:extension>
    </xs:complexContent>
</xs:complexType>
```

Приведем пример фрагмента XML-схемы, полученной в результате кодирования части UML-модели информации о качестве данных из ISO 19115 (Рис. 4)

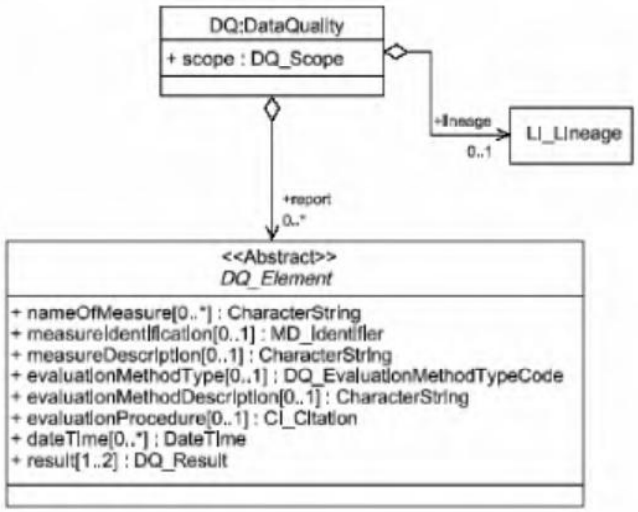


Рис. 4 Часть UML-модели информации о качестве данных из ISO 19115

```
<xs:complexType name="DQ_DataQuality_Type">
   <xs:complexContent>
      <xs:extension base="gco:AbstractObject_Type">
         <xs:sequence>
            <xs:element name="scope" type="gmd:DQ_Scope_PropertyType"/>
            <xs:element name="lineage" type="gmd:LI_Lineage_PropertyType"
minOccurs="0"/>
            <xs:element name="report" type="gmd:_DQ_Element_PropertyType"
minOccurs="0" maxOccurs="unbounded"/>
         </xs:sequence>
      </xs:extension>
   </xs:complexContent>
</xs:complexType
```

Таким образом, используя набор правил из вышеперечисленных стандартов, мы можем преобразовать UML-модель метаданных в XML-схему. Эта большая работа уже проделана специалистами ТС 211 и на сайте технического комитета опубликован набор пространств имен с соответствующими XML-схемами. Следует подчеркнуть, что эти схемы представляют собой результат кодирования **всеобъемлющей концептуальной модели метаданных** и поэтому могут служить только примером методики. Они не предназначены для практического применения в программировании приложений, работающих с метаданными.

Рассмотрим практическое применение вышеописанной методики для решения конкретных задач создания системы поддержки метаданных (Рис. 5). Уровень этой системы (национальный, региональный, ведомственный и пр.) не имеет значения, так как для любого из этих уровней необходимо произвести одни и те же действия.

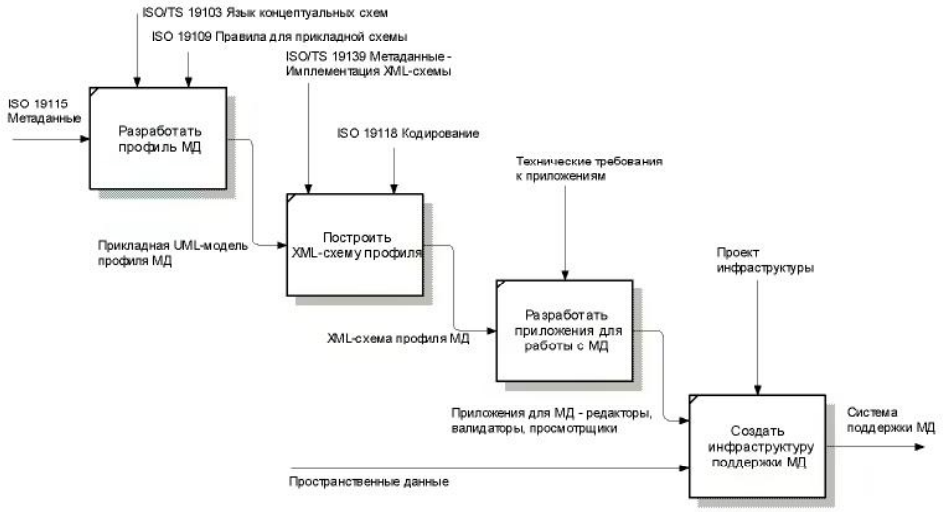


Рис. 5 Функциональная модель методики

Первым шагом создания такой системы является разработка прикладного профиля метаданных для данного уровня системы. Следует кратко остановиться на понятии профиля. Профиль – это концептуальная модель метаданных для данного конкретного сообщества пользователей пространственных данных. Правилам и методике создания профилей посвящен Международный стандарт ISO 19106. В более сжатом и конкретизированном виде раздел о создании профилей включен непосредственно в стандарт ISO 19115.

Модель профиля отличается от всеобъемлющей модели стандарта ISO 19115 тем, что в ней исключены те сущности (классы, атрибуты, отношения), которые не существенны для этого конкретного сообщества пользователей и добавлены так называемые расширения, которые содержат новые сущности, нужные этому сообществу и отсутствующие в стандарте.

Разработка профиля – достаточно трудоемкая и творческая работа, которая требует высокой квалификации и занимает немалое время. Поэтому особое значение приобретает разработка и принятие на государственном уровне национального профиля метаданных, который учитывает обобщенные требования к метаданным для всех пользователей пространственных данных в государстве. Необходимость такой разработки и подходы к ее реализации уже обсуждались авторами в предыдущих публикациях [6]. Национальный профиль метаданных должен являться базовым для всех нижележащих иерархических уровней и, в свою очередь, подвергаться модификации для получения профилей на этих уровнях.

Разработка UML-модели профиля может производиться с помощью таких программных пакетов, как Microsoft Visio или Rational Rose. При этом работа значительно облегчается, если наряду с текстом Государственного стандарта – профиля метаданных доступны также и UML-диаграммы в форматах этих программных средств.

Следующим шагом является разработка XML-схемы профиля метаданных. Общая методика этого процесса была описана выше. Ее конкретная реализация состоит в том, что должно быть произведено кодирование прикладной UML-модели профиля метаданных в конкретную XML-схему. При этом должна учитываться специфика планируемых к разработке программных средств поддержки метаданных. Дело в том, что сама Техническая Спецификация ISO 19139 содержит ряд расширений, обеспечивающих, например, работу в WEB-среде, различные методы передачи метаданных и самих данных и т.п. Все эти специфические моменты должны быть учтены в процессе кодирования XML-схемы.

XML-схема может создаваться в различных приложениях-редакторах – Visual Studio, XMLSpy, Liquid XML Studio и т.п. Следует также отметить, что в Rational Rose имеется возможность полуавтоматической генерации XML-схемы из UML-модели. При создании XML-схемы удобно в качестве заготовок использовать готовые блоки из разработанного ТС211 набора XML-схем стандарта ISO 19115. Результатом проделанной работы является набор XML-схем – файлов формата XSD (XML Schema Definition), сгруппированных по различным пространствам имен.

Полученные XSD используются для разработки программных средств поддержки метаданных – редакторов, валидаторов, просмотрщиков и т.п. Каждое из этих средств работает с метаданными в формате XML, создавая их, проверяя правильность их составления, обеспечивая просмотр или обращаясь с запросами к хранилищу метаданных. Для отображения содержания XML в удобочитаемом для человека виде используется технология XSLT (Extensible Style Language Transformation) [7].

Последним этапом создания системы поддержки метаданных является разворачивание инфраструктуры этой поддержки, состоящей из хранилищ метаданных, средств их публикации, опоискования и связи с самими пространственными данными.

В заключение, с учетом собственного опыта по реализации всех стадий вышеописанной методики, хотелось бы сформулировать некоторые рекомендации.

Для построения инфраструктуры пространственных данных любого (а в особенности, государственного уровня) необходимо:

- разработать и принять в качестве Государственного стандарта национальный профиль метаданных, производный от ISO 19115;
- локализировать и утвердить в качестве Государственного стандарта техническую спецификацию ISO 19139;

- обеспечить публикацию диаграмм UML-модели и XML-схем этих Государственных стандартов в электронном виде;
- проводить популяризацию этих стандартов, средств и методов работы с ними в среде разработчиков геоинформационных систем (публикации, WEB-публикации, курсы, семинары, конференции).

## Список литературы


1. International Standart ISO 19101 Geographic information – Reference model. – ISO, 2002. – 42 p.
2. Technical Specification ISO/TS 19103 Geographic information – Conceptual schema language. – ISO, 2005. – 67 p.
3. International Standart ISO 19115 Geographic information – Metadata. – ISO, 2003. – 140 p.
4. International Standart ISO 19118 Geographic information – Encoding. – ISO, 2005. – 104 p.
5. Technical Specification ISO/TS 19139 Geographic information – Metadata – XML schema implementation. – ISO, 2007. – 111 p.
6. Салтовец А.А. К вопросу о национальном профиле метаданных пространственных данных: - Ученые записки Таврического национального университета им. В.И.Вернадского. Серия «География». Том 20(59). – 2007. – №1 – С. 183 – 190.
7. XSL Transformations (XSLT) Version 2.0: W3C Recommendation - http://www.w3.org/TR/2007/REC-xslt20-20070123/



***Салтовець О. О., Ніколаєв В. М.*** **Методичні питання реалізації метаданих на основі Міжнародного стандарту ISO 19115** // Вчені записки Таврійського національного університету ім. В. І. Вернадського. – 2008. – Серія «Географія». – Т. 21 (60). – № 1. – С. 132-141

У статті розглянуто методіку побудови системи підтримки метаданих просторових даних, що базується на групі Міжнародних стандартів ISO 19100. Запропоновані шляхи та засоби реалізації системи на державному та нижчих рівнях використання просторових даних.



*Ключові слова:* Метадані, стандарти, профілі, просторові дані, XML



***Saltovets A,. Nikolaev V.*** **Methodical aspects of metadata realization based on the International standard ISO 19115** // Uchenye zapiski Tavricheskogo Natsionalnogo Universiteta im. V.I. Vernadskogo. – 2008. – Series «Geography». – V. 21 (60). – № 1. – P. 132-141

The methodology of building spatial data metadata supporting system based on the group of International standards series ISO 19100 is discussed in the article. The approaches and methods of the system realization on the state and underlying levels spatial data using are suggested.



*Keywords:* Metadata, Standards, Profiles, Spatial data, XML

*Поступила в редакцию 22.04.2008 г.*