
РАЗРАБОТКА КАТАЛОГА МЕТАДАННЫХ СИСТЕМЫ GEO-UKRAINE

Наталья Куссуль, Алина Рудакова, Алексей Кравченко

Аннотация: *Статья посвящена актуальному вопросу каталогизации разнородных данных – дистанционного зондирования, наземных измерений и данных моделирования. В работе предлагается концептуальный подход к созданию системы каталогизации, с учетом мирового опыта построения таких систем.*

Рассмотрены архитектурные особенности разрабатываемого каталога метаданных, существующие решения для каталогов метаданных, вопросы стандартизации метаданных и создания профилей метаданных, а также некоторые аспекты реализации каталога метаданных в информационной системе GEO-Ukraine.

Ключевые слова: *каталог данных, метаданные, профиль метаданных, стандарт ISO 19115*

Введение

Современные спутниковые данные представляют особый интерес для решения разнообразных задач из областей науки, сельского хозяйства и др., в целях снижения рисков стихийных бедствий, оценки последствий природных и техногенных катастроф. Данные дистанционного зондирования (ДЗЗ) по сравнению с данными наземных наблюдений обладают рядом преимуществ, в т.ч. стоимость, доступность, оперативность. Поэтому круг потребителей аэрокосмической информации в Украине и в мире расширяется.

В Украине достаточно большое количество организаций занимается решением разнообразных тематических задач, связанных с анализом данных ДЗЗ. По мере решения подобных задач в каждой организации накапливается собственный архив геопространственных данных, эффективное использование которого невозможно без его каталогизации.

В Украине разрабатывается информационная система GEO-Ukraine [1], которая должна стать украинским сегментом в международной системе систем наблюдений Земли GEOSS (www.earthobservations.org). В рамках системы GEO-Ukraine (<http://www.geoss-ukraine.org.ua>) одной из приоритетных задач является объединение уже существующих баз геопространственных данных. Однако они имеют сложную структуру, разные форматы, часто большой объем, не позволяющий активно использовать данные в сети Internet. Залогом эффективного использования данных ДЗЗ является инфраструктура хранения, поиска и предоставления этих данных на основе их структурированных описаний – метаданных.

Метаданные предоставляют необходимый и достаточный объем информации необходимый для того, чтобы понять природу и суть содержания описываемого набора данных. В данное время метаданные являются неотъемлемой частью при создании хранилищ данных и их каталогизации. Создание такого каталога метаданных упрощает управление, создание запросов, полноценное использование и понимание данных. Для унификации процессов поиска и предоставления информации в рамках единой системы необходима стандартизация наборов и форматов предоставления метаданных, т.е. необходимо разработать профиль метаданных.

Существующие решения

Проблема создания каталогов метаданных стала актуальной в связи с взрывообразным увеличением объемов данных, поступающих со спутников ДЗЗ. На данный момент каталоги метаданных созданы ведущими национальными космическими агентствами, поставщиками и дистрибьюторами данных ДЗЗ, научными организациями, работающими со спутниковыми данными. Среди существующих зарубежных каталогов метаданных следует упомянуть о каталоге данных ДЗЗ Европейского космического агентства (ESA) (<http://earth.esa.int/resources/catalogues/>) и каталоге NASA EOS Data Gateway (<http://delenn.gsfc.nasa.gov/~imswww/pub/imswelcome/>).

Каталог метаданных ESA Multi-mission Catalogue Service позволяет пользователям выполнять поиск данных ДЗЗ, отображать результаты поиска на карте, анализировать изображения для предварительного просмотра. При поиске имеется возможность указывать необходимый набор спутников и сенсоров, а также задавать временные и пространственные характеристики снимков. Поддерживается два вида пользователей – анонимные и авторизованные. Авторизованным пользователям предоставляется дополнительная услуга заказывать данные ДЗЗ в режиме online. Каталог создан для предоставления данных как со спутников ESA (ERS-1/2, Envisat), так и других миссий (Landsat, PROBA, SPOT). Для этого каталога создано несколько интерфейсов пользователя, в частности Web-интерфейс EOLI-Web (<http://eoli.esa.int/servlets/template/welcome/entryPage2.vm>), и настольное приложение EOLI-SA. Оба приложения созданы на основе технологии Java.

Каталог EOS Data Gateway является распределенной системой поиска и заказа данных ДЗЗ и результатов их обработки. Интерфейс пользователя каталога реализован в виде тонкого Web-клиента. Подсистема поиска позволяет указывать наборы данных, временные и пространственные ограничения для продуктов ДЗЗ, а также некоторые дополнительные параметры. Пользователь может просматривать атрибуты снимков, предварительные изображения данных, а также метаинформацию о наборах данных, спутниках и сенсорах. Каталог предоставляет доступ к данным NASA, в основном к данным приборов на борту спутников Aqua и Terra. Система поддерживает анонимных (guests) и зарегистрированных пользователей. Последние имеют возможность сохранять параметры поисковых запросов и результатов поиска, а также повторно использовать регистрационную информацию при заказе данных.

Среди украинских каталогов метаданных следует упомянуть каталог метаданных, созданный в ГНПЦ «Природа» и в Центре аэрокосмических исследований Земли (ЦАКИЗ). Каталог ГНПЦ «Природы» доступен в Internet по адресу http://www.pryroda.gov.ua/pryroda/search_film.do и позволяет осуществлять поиск спутниковых данных, которыми располагает организация. Каталог создан для облегчения поиска данных клиентами для заказа продукции этой организации. Поисковая система каталога позволяет осуществлять поиск данных, указывая конкретный спутник или сенсор, временные и пространственные ограничения для снимка. Имеется возможность просматривать местоположения снимков на карте и изображения снимков для быстрого просмотра. Каталог ЦАКИЗ на данный момент доступен только в Intranet-сети этой организации и позволяет проводить поиск по атрибутам снимка (спутник, сенсор), временным и пространственным параметрам снимка. Отдельно существует интерфейс внесения новых метаданных. Каталог создан для поддержки работ внутри самой организации. В Украине достаточно большое число организаций занимается тематической обработкой данных ДЗЗ и имеет накопленный архив данных, который зачастую не структурирован. Для эффективного использования существующих данных ДЗЗ необходимо создание каталога метаданных о данных ДЗЗ. Модели метаданных данных каталогов будут рассмотрены далее.

Организация инфраструктуры пространственных данных и их хранения определяются стандартами. Они задают язык и правила взаимодействия участников, без которых это взаимодействие невозможно.

В данное время разработано множество стандартов по представлению пространственных данных. Ряд международных организаций занимается проблемой стандартизации метаданных геопространственных данных.

Среди ряда международных организаций занимающихся стандартизацией геопространственных данных и геоинформационных систем необходимо отметить Федеральный комитет по географическим данным FGDC (<http://www.fgdc.gov/>) разработавший один из первых стандартов для метаданных Content Standard for Digital Geospatial Metadata. Этот стандарт использовался при построении Национальной инфраструктуры геопространственных данных службой США.

Организация Open GIS Consortium (OGC) была основана в 1994 году с целью «обеспечения спецификаций пространственного интерфейса, доступных для всеобщего использования». Основная цель консорциума OGC — создание технологий обеспечивающих прозрачность взаимодействия программных средств разных производителей, возможность конвертирования данных разных форматов и их совместного использования, открытость функциональных интерфейсов и унификация для конечных пользователей (www.opengis.org).

Наиболее общепризнанным и используемым в большинстве международных и национальных проектах является стандарт ISO 19115 Geographic information – Metadata Международной организации по стандартизации ISO/TC 211 (www.isotc211.org). Данный стандарт рекомендован к использованию в рамках международной системы GEOSS и инициативы INSPIRE (Infrastructure for Spatial Information in Europe — Инфраструктура для пространственной информации в Европе), которая является основой для обмена геопространственными данными в рамках европейской программы GMES (www.gmes.info).

Стандарт ISO 19115

Основной задачей стандарта является определение методологии формирования метаданных для географической информации. В стандарте определена терминология, методология и универсальный набор элементов метаданных, посредством которых эти метаданные описываются.

Данный стандарт определяет:

- обязательные (O) и условные (U) пакеты метаданных, их сущности и элементы;
- необходимый и достаточный набор метаданных, для большинства случаев их использования (поиск данных, определение соответствия данных, доступ к данным, передача данных и использование цифровых данных);
- необязательные (H) элементы метаданных, позволяющие расширить стандартное описание геопространственных данных при необходимости.

Метаданные представляются совокупностью UML-пакетов. Каждый пакет имеет смысловое наполнение и характеризует тот или иной аспект метаданных. Достоинством этого стандарта является то, что он представлен на Универсальном языке моделирования (UML), так как UML-диаграммы могут использоваться для генерации схемы базы данных в полном соответствии с этим стандартом. В стандарте определено 14 UML-пакетов, каждый из которых состоит из одной или нескольких сущностей (рис. 1). Сущность (UML-класс) формально описывает группы объектов, которые обладают одинаковым набором характеристик. Классы содержат элементы (атрибуты), характеризующие конкретный экземпляр метаданных.

Стандартом определен конкретный набор элементов метаданных, однако далеко не все из них используются при описании той или иной географической информации. Поэтому актуальной задачей является создание профиля метаданных для решения конкретной задачи.

Профиль метаданных является ключевым понятием геопространственных стандартов ISO и определяет специфический набор атрибутов метаданных, которые должны быть заполнены в описании для того, чтобы данное описание удовлетворяло данному профилю. Предусматривается, что профили метаданных могут создаваться как отдельными организациями, так и комитетами, работающими в направлении гармонизации. В стандарте полностью прописана процедура создания профиля метаданных.

В качестве основы для формирования профиля определен базовый набор элементов метаданных называемый «ядром метаданных», необходимый для основного документирования географических данных. Элементы «ядра метаданных» предоставляют минимальный объем информации, необходимый для понимания природы и содержания описываемого набора данных:

- информация о метаданных;
- информация о данных;
- информация о способе получения данных;
- информация о системе координат;
- информация о происхождении данных.

Идентификатор файла метаданных (H) (MD_Metadata.fileIdentifier)	Стандарт кодировки данных (O) (MD_Metadata > MD_DataIdentification.characterSet)
Название стандарта метаданных (H) (MD_Metadata.metadataStandardName)	Основная тема набора данных (O) (MD_Metadata > MD_DataIdentification.topicCategory)
Версия стандарта метаданных (H) (MD_Metadata.metadataStandardVersion)	Краткое содержание набора данных (O) (MD_Metadata > MD_DataIdentification.abstract)
Язык создания метаданных (Y) (MD_Metadata.language)	Масштаб или расстояние на местности (H) (MD_Metadata > MD_DataIdentification.spatialResolution > MD_Resolution.equivalentScale or MD_Resolution.distance)
Стандарт кодировки метаданных (Y) (MD_Metadata.characterSet)	Метод пространственного представления (H) (MD_Metadata > MD_DataIdentification.spatialRepresentationType)
Сторона, ответственная за метаданные (O) (MD_Metadata.contact > CI_ResponsibleParty)	Географическое положение набора данных, координаты или географический идентификатор (Y) (MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox or EX_GeographicDescription)
Дата создания метаданных (O) (MD_Metadata.dateStamp)	Пространственно-временные характеристики (H) (MD_Metadata > MD_DataIdentification.extent > EX_Extent)

	> EX_TemporalExtent or EX_VerticalExtent)
Название набора данных (O) (MD_Metadata > MD_DataIdentification.citation > CI_Citation.title)	Информация об Интернет-ресурсах (H) (MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource)
Дата создания набора данных (O) (MD_Metadata > MD_DataIdentification.citation > CI_Citation.date)	Формат данных и версия формата (H) (MD_Metadata > MD_Distribution > MD_Format.name and MD_Format.version)
Сторона, ответственная за набор данных (H) (MD_Metadata > MD_DataIdentification.pointOfContact > CI_ResponsibleParty)	Система координат (H) (MD_Metadata > MD_ReferenceSystem)
Язык создания данных (O) (MD_Metadata > MD_DataIdentification.language)	Информация о происхождении данных (H) (MD_Metadata > DQ_DataQuality.lineage > LI_Lineage)

Дополняя базовый набор метаданных другими элементами, обеспечивается необходимая степень детализации в зависимости от решаемой задачи.

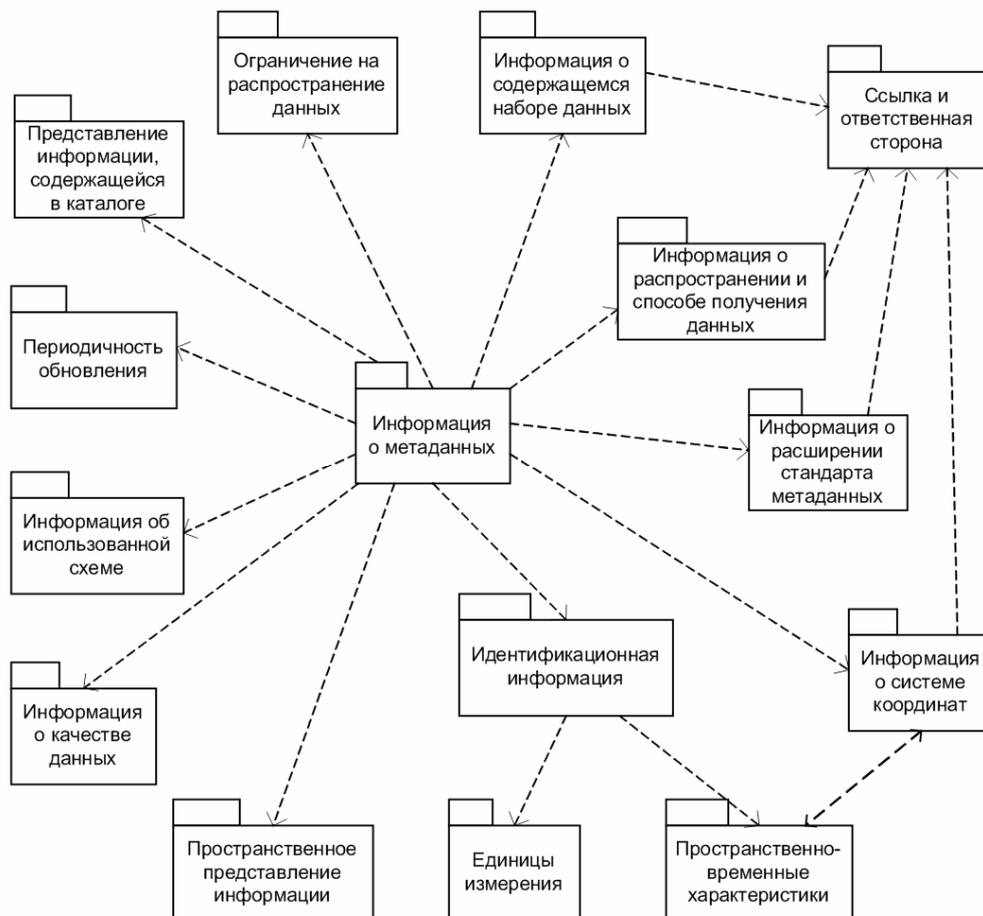


Рис. 1. UML-пакеты метаданных

Создание профиля стандарта метаданных

Профиль метаданных разрабатывается с учетом специфики предметной области в зависимости от решаемых задач и представляет собой некую «проекцию» стандарта на предметную область (рис. 2). Профиль метаданных для украинского сегмента должен быть близким к международным разработкам в этой области и соответствовать интересам Украины в лице организаций, которые являются пользователями геопространственных данных.

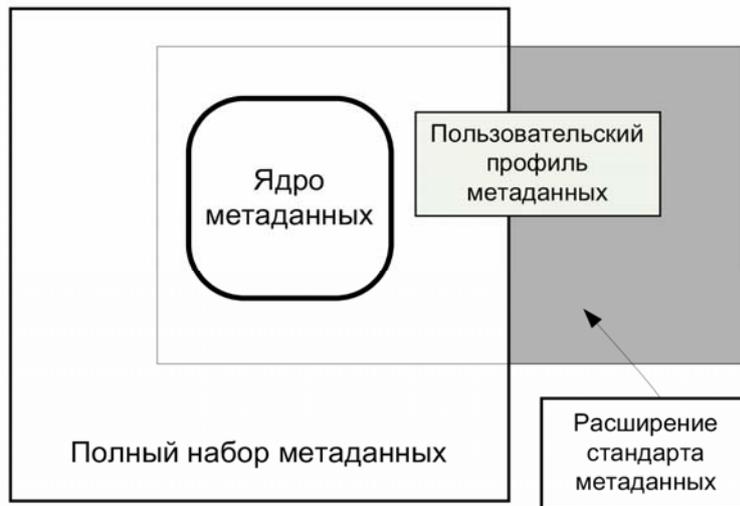


Рис. 2. Механизм создания пользовательского профиля метаданных

В стандарте ISO №19115 прописаны следующие правила создания профиля:

1. Перед созданием профиля необходимо проверить зарегистрированные профили.
2. Профиль создается в соответствии с правилами для определения расширений.
3. В профиле не изменяются имена, определения или типы данных элементов метаданных.
4. Профиль должен содержать:
 - элементы ядра метаданных;
 - все обязательные элементы метаданных во всех обязательных секциях;
 - все условные элементы метаданных во всех обязательных секциях, если набор данных удовлетворяет условию, необходимому для включения элементов метаданных в соответствующую секцию;
 - все обязательные элементы метаданных во всех условных секциях, если набор данных удовлетворяет заданному условию;
 - все условные элементы метаданных во всех условных секциях, если набора данных удовлетворяет условию, необходимому для наличия элементов метаданных и секции.
5. Взаимосвязи между сущностями и пакетами метаданных должны быть определены строго в соответствии со стандартом.
6. Профиль должен быть доступным каждому пользователю, получающему и использующему метаданные созданные в соответствии с этим профилем.

Создаваемый профиль метаданных для информационной системы GEO-Ukraine должен учитывать существующие профили метаданных ГНПЦ «Природа» и ЦАКИЗ. Поля, описывающие данные этих центров присутствуют в разрабатываемом профиле в соответствии со стандартом.

В каталоге метаданных ГНПЦ «Природа» содержатся следующие атрибуты снимков:

спутник, датчик, метод регистрации изображения, процент облачности (для оптических снимков), регион снимка (в текстовом виде), дата и время съемки, информация о спектральных каналах изображения, пространственном разрешении, формате и размере данных, координаты углов снимка, и некоторые другие параметры. Дополнительно содержится RGB-изображение снимка для предварительного просмотра.

Каталог метаданных ЦАКИЗ содержит 4 основных информационных группы разработанного профиля: информация о метаданных, идентификационная информация, способ распространения данных, ссылка на владельца данных, информация об инструменте. Информация о метаданных в данном каталоге организована по иерархическому принципу, также созданы словари, которые включают весь перечень допустимых значений соответствующих атрибутов.

Необходимым атрибутом метаданных является информация об уровне обработки космических снимков. Комитетом CEOS определены уровни обработки данных L0, L1A, L1B, L2, L3, L4. В рамках этой классификации уровень обработки снимка варьируется от необработанных данных телеметрии со спутника (уровень L0) до результатов моделирования с использованием данных ДЗЗ (уровень L4).

Обязательным элементом профиля метаданных должна быть информация об изображении предварительного просмотра. При этом желательно, чтобы эти изображения содержали данные географической привязки, что позволит создать полнофункциональный интерфейс пользователя каталога метаданных.

Каталог метаданных

Каталог метаданных, разрабатываемый в Институте космических исследований, представляет собой единое хранилище метаданных о данных ДЗЗ системы GEO-Ukraine. В будущем каталог также будет предоставлять информацию о данных измерений in-situ, результатах тематической обработки данных ДЗЗ и результатах моделирования. Создание такого каталога будет способствовать объединению усилий в рамках системы GEO-Ukraine, создаваемой под эгидой Национального космического агентства Украины как национальный сегмент международной системы GEOSS [2]. В частности каталог позволит избежать дублирования данных в системе GEO-Ukraine, существенно упростит обработку данных ДЗЗ, распределенную между несколькими организациями (в первую очередь автоматическую), упростит заказ услуг на обработку данных ДЗЗ.

В результате анализа потребностей сообщества ДЗЗ в Украине было выделено следующие группы пользователей каталога метаданных:

- представители сообщества ДЗЗ в Украине, желающие получить доступ к данным ДЗЗ;
- представители поставщиков данных ДЗЗ и продуктов их обработки;
- автоматизированные и полностью автоматические системы обработки данных.

Исходя из выделенных групп пользователей, определены основные прецеденты использования каталога:

- интерактивный поиск данных;
- автоматический поиск данных;

- внесение метаданных в систему.

Сообщество GEO-Ukraine состоит из территориально распределенных организаций, которые занимаются обработкой данных ДЗЗ, центрами приема данных ДЗЗ, обработкой и хранением. В такой распределенной системе существуют два подхода организации метаданных в единое хранилище: централизованный и децентрализованный.

Реализация децентрализованного способа и соответствующей архитектуры единого хранилища метаданных показаны на рис. 3. В рамках децентрализованного подхода метаданные хранятся непосредственно в организациях, которые хранят эти данные. В рамках каталога метаданных ведется реестр таких центров и создана распределенная подсистема поиска данных. Для организации подсистемы поиска данных определена и стандартизирована модель метаданных, в каждом центре обработки данных создана локальная подсистема поиска данных, а также специфицирован интерфейс к поисковым системам для каждого центра. В центральной подсистеме поиска созданы адаптеры подсистем поиска для каждого центра обработки данных. Создание такой системы существенно облегчает стандартизация поисковых подсистем центров обработки данных. Для взаимодействия внешних пользователей с каталогом метаданных может быть создан единый интерфейс, в частности на Web-технологиях. По такой схеме устроен каталог метаданных EOS Data Gateway.

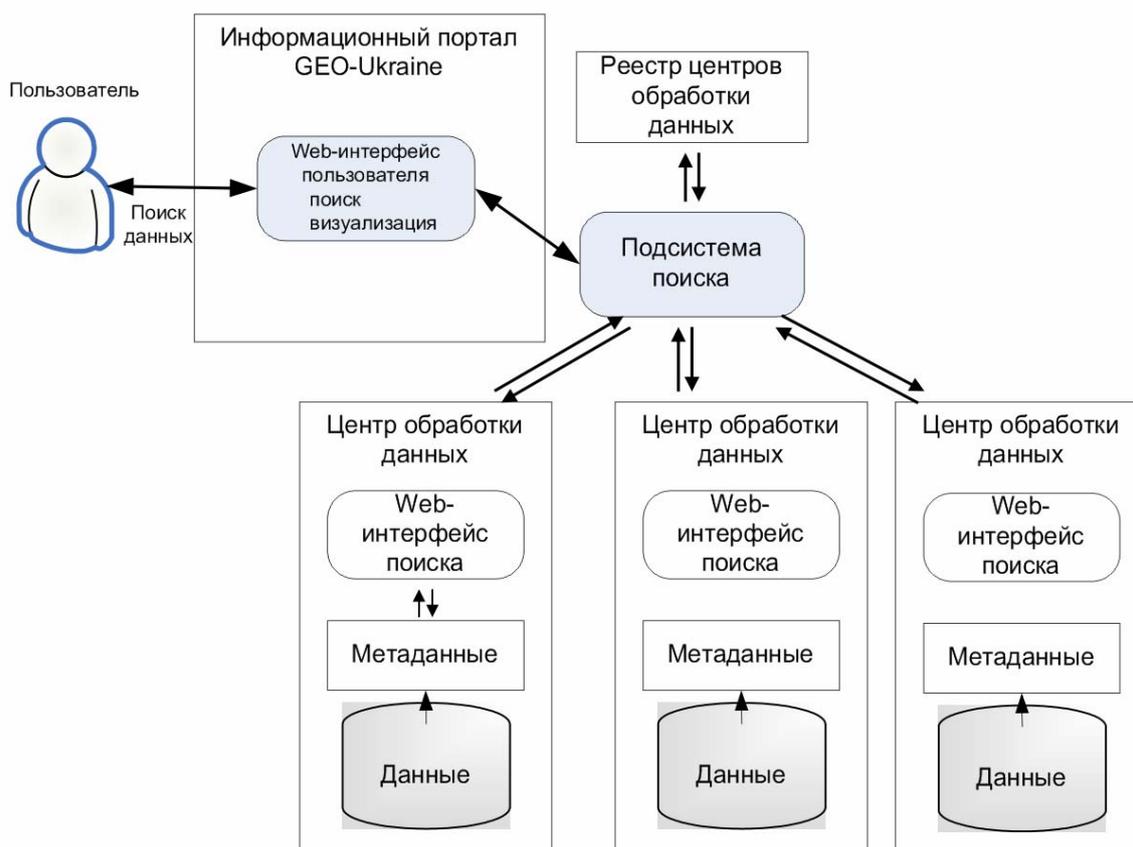


Рис. 3. Децентрализованная схема организации метаданных

Достоинством данной схемы является высокая скорость обновления метаданных, поскольку они вносятся локально каждым центром обработки данных. Однако она обладает существенными недостатками, а именно:

- скорость поиска лимитирована наихудшими показателями параметров связи между центральной подсистемой поиска и центрами обработки данных, а также наихудшей производительностью подсистем поиска метаданных;
- организациям, которые желают быть представленными в каталоге, необходимо создать и поддерживать информационную инфраструктуру, в частности обеспечивать хранение метаданных, поиск по метаданным, обеспечивать надежность предоставления услуг.

В рамках централизованной схемы метаданные обо всех данных в системе хранятся в выделенном хранилище. Данная схема поддерживает несколько типов центров обработки данных (рис. 4):

- центры, поддерживающих собственное хранилище метаданных;
- центры первого типа, которые дополнительно обеспечивают поисковый интерфейс пользователя к собственному хранилищу метаданных;
- центры, не имеющие собственного хранилища метаданных.

Для работы с центрами первого и второго типов должен быть определен протокол сбора метаданных для каждого центра. Для поддержки центров третьего типа должен быть создан специализированный интерфейс, предназначенный для внесения пользовательских метаданных в каталог.

Данная схема не обладает недостатками предыдущей благодаря возможности обновления метаданных в режиме offline и поддержки нескольких типов центров обработки данных. В то же время частота обновления сведений о данных ограничена частотой сбора информации с центров обработки данных.

По данной схеме организованы большинство поисковых систем общего назначения, в частности поисковые системы Google (www.google.com) и Yahoo (www.yahoo.com).



Рис. 4. Централизованная схема организации метаданных

При создании каталога метаданных для системы GEO-Ukraine необходимо учитывать следующие особенности украинских организаций, занимающихся обработкой данных ДЗЗ:

- различное состояние информационной инфраструктуры в подобных организациях;
- существование организаций, для которых создание соответствующей инфраструктуры нецелесообразно (например, в случае небольших объемов данных);
- недостаточная обеспеченность многих организаций необходимым доступом в Internet.

Учитывая эти особенности и преимущества централизованного похода к организации метаданных, для реализации каталога метаданных системы GEO-Ukraine была выбрана централизованная схема каталога метаданных.

В докладе будут рассмотрены рекомендации для представления и хранения пространственных метаданных, разработанные в Институте космических исследований НАНУ-НКАУ на основе стандарта ISO 19115.

Выводы

В данной статье предлагается концептуальный подход к созданию системы каталогизации, с учетом мирового опыта построения таких систем.

Каталоги данных и метаданных системы GEO-Ukraine планируется внедрить в международный каталог метаданных ECHO (Earth Observing System Clearing House, www.echo.eos.nasa.gov).

Работа выполнена при поддержке проектов УНТЦ №3872 «Разработка эффективных GRID-технологий экологического мониторинга на основе спутниковых данных» и INTAS-CNES-NSAU №06-1000024-9154 «Data Fusion Grid Infrastructure».

Литература

- [1] Федоров О.П., Куссуль Н.Н., Шелестов А.Ю. Задачи и перспективы развития в Украине информационной системы наблюдения Земли из космоса// Проблемы управления и информатики. — 2006. — №6. — С. 116–121.
- [2] Global Earth Observation System of Systems GEOS. 10-Year Implementation Plan Reference Document. Noordwijk, Netherlands: ESA Publication Division. 2005. 212 p

Авторы

Куссуль Наталья Николаевна - Институт космических исследований НАН и НКА Украины, профессор, доктор технических наук, заведующий отделом, проспект Академика Глушкова 40, 03680 Киев, Украина; e-mail: inform@ikd.kiev.ua

Рудакова Алина Игоревна - Институт космических исследований НАН и НКА Украины, младший научный сотрудник; проспект Академика Глушкова 40, 03680 Киев, Украина; e-mail: alinarudakova@yahoo.com

Кравченко Алексей Николаевич - Институт космических исследований НАН и НКА Украины, аспирант, младший научный сотрудник; проспект Академика Глушкова 40, 03680 Киев, Украина; e-mail: oleksiy.kravchenko@gmail.com