

УДК 681.3

Каталог метаданных системы GEO-Ukraine

Куусуль Н.Н., Лавренюк А.Н., Лавренюк С.И., Грипич Ю.А.
Институт космических исследований НАНУ-НКАУ
inform@ikd.kiev.ua

Abstract

Kussul N., Lavrenyuk A., Lavrenyuk S., Grypych Yu. Metadata catalogue of GEO-Ukraine system. This article is devoted to the cataloging of manifold data: in-situ and modeling data. The conceptual approach for catalog system creation according to the international experience is proposed. Architectural features of metadata catalogue, existing solutions, standardization issues, creation metadata profiles and some aspects of metadata catalogue implementation in GEO-Ukraine information system are discussed.

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://deleann.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. Поддерживаются 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 (Y) 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 (Н) 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.

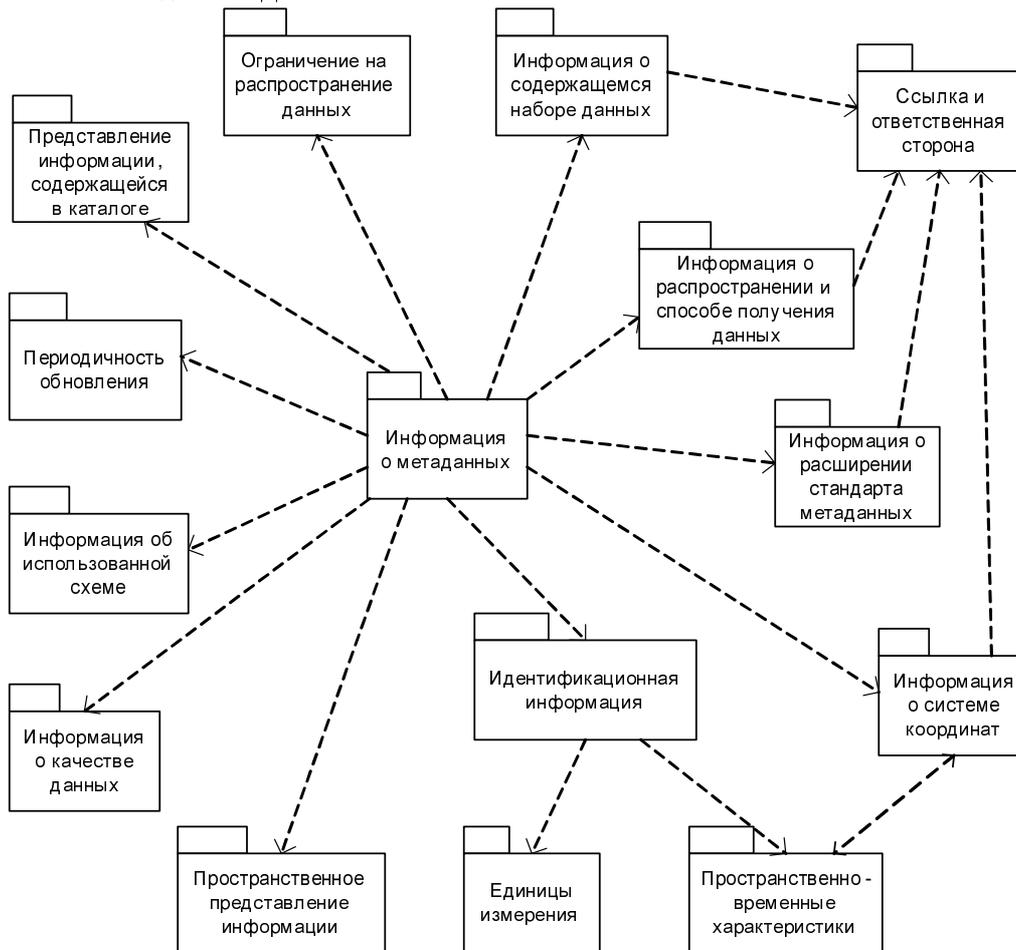


Рисунок 1 – UML-пакеты метаданных

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.

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

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)

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 геопрограммных данных.

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 профілі не змінюються імена, визначення або типи даних елементів метаданих.

4. Профіль повинен містити:

- елементи ядра метаданих;
- всі обов'язкові елементи метаданих у всіх обов'язкових секціях;
- всі умовні елементи метаданих у всіх обов'язкових секціях, якщо набір даних задовольняє умову, необхідну для включення елементів метаданих в відповідну секцію;
- всі обов'язкові елементи метаданих у всіх умовних секціях, якщо набір даних задовольняє задану умову;
- всі умовні елементи метаданих у всіх умовних секціях, якщо набір даних задовольняє умову, необхідну для наявності елементів метаданих і секції.

5. Взаємозв'язки між сутностями і пакетами метаданих повинні бути визначені строго в відповідності зі стандартом.

6. Профіль повинен бути доступним кожному користувачеві, отримувачеві і використовувачеві метаданих створених в відповідності з цим профілем.

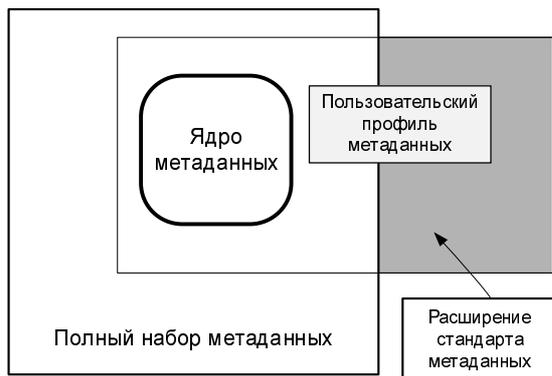


Рисунок 2 - Механізм створення профілю метаданих користувача

Створюваний профіль метаданих для інформаційної системи GEO-Ukraine повинен враховувати існуючі профілі метаданих ГНПЦ «Природа» і ЦАКИЗ. Поля, описуючі дані цих центрів присутні в розроблюваному профілі в відповідності зі стандартом.

В каталозі метаданих ГНПЦ «Природа» містяться наступні атрибути знімків:

спутник, датчик, метод реєстрації зображення, відсоток хмарності (для оптичних знімків), регіон знімка (в текстовому вигляді), дата і час зйомки, інформація про спектральні канали зображення, просторове роздільність, формат і розмір даних, координати кутів знімка, і деякі інші параметри. Додатково

міститься RGB-зображення знімка для попереднього перегляду.

Каталог метаданих ЦАКИЗ містить 4 основних інформаційних групи розробленого профіля: інформація про метаданих, ідентифікаційна інформація, спосіб поширення даних, посилання на власника даних, інформація про інструмент. Інформація про метаданих в даному каталозі організована за ієрархічним принципом, також створені словари, які включають весь перелік допустимих значень відповідних атрибутів.

Необхідним атрибутом метаданих є інформація про рівні обробки космічних знімків. Комітетом CEOS визначені рівні обробки даних L0, L1A, L1B, L2, L3, L4. В межах цієї класифікації рівень обробки знімка варіюється від необроблених даних телеметрії зі супутника (рівень L0) до результатів моделювання з використанням даних ДЗЗ (рівень L4).

Обов'язковим елементом профіля метаданих повинна бути інформація про зображенні попереднього перегляду. При цьому бажано, щоб ці зображення містили дані географічної прив'язки, що дозволить створити повнофункціональний інтерфейс користувача каталогу метаданих.

Каталог метаданих

Каталог метаданих, розроблюваний в Інституті космічних досліджень, є єдиним сховищем метаданих про дані ДЗЗ системи GEO-Ukraine. В майбутньому каталог також буде надавати інформацію про дані вимірювань in-situ, результатах тематичної обробки даних ДЗЗ і результатах моделювання. Створення такого каталогу сприятиме об'єднанню зусиль в межах системи GEO-Ukraine, створюваної під егідою Національного космічного агентства України як національний сегмент міжнародної системи GEOSS [2]. В частині каталог дозволить уникнути дублювання даних в системі GEO-Ukraine, суттєво спростить обробку даних ДЗЗ, розподілену між кількома організаціями (в першу чергу автоматичною), спростить замовлення послуг на обробку даних ДЗЗ.

В результаті аналізу потреб спільноти ДЗЗ в Україні було виділено наступні групи користувачів каталогу метаданих:

- представники спільноти ДЗЗ в Україні, які хочуть отримати доступ до даних ДЗЗ;
- представники постачальників даних ДЗЗ і продуктів їх обробки;

– 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, центром 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.

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.

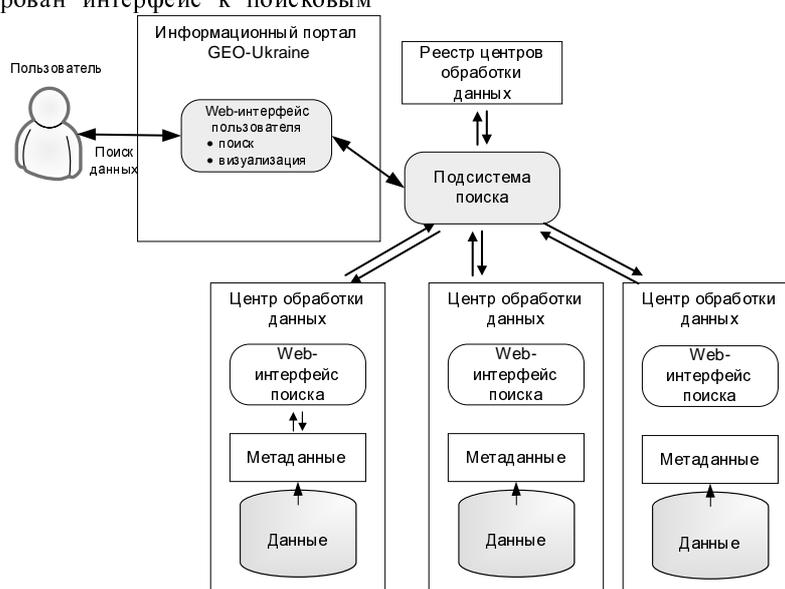


Рисунок 3 - Децентрализованная схема организации метаданных

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).

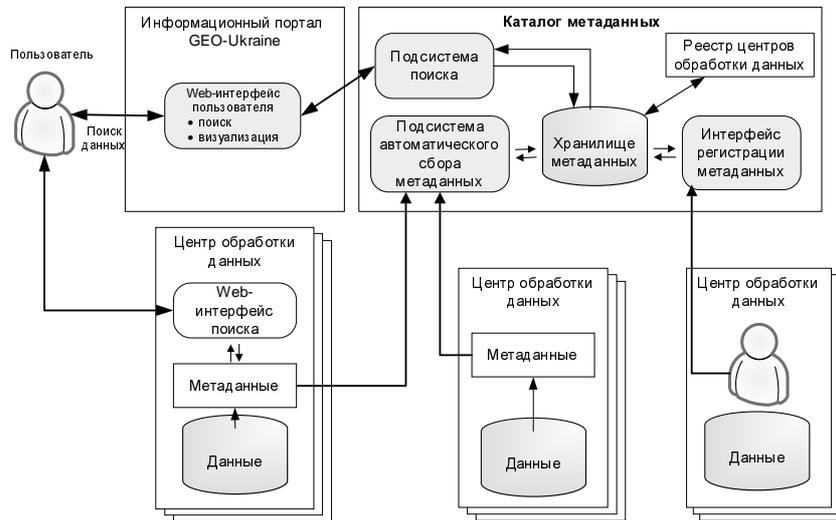


Рисунок 4 - Централизованная схема организации метаданных

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 подхода к организации метаданных, для реализации каталога метаданных системы GEO-Ukraine была выбрана централизованная схема каталога метаданных.

Реализация каталога в системе GeoNetwork

Пилотная версия каталога метаданных реализована с помощью программного обеспечения (ПО) GeoNetwork. GeoNetwork – это свободно-распространяемое программное обеспечение с открытым программным кодом на языке Java и соответствующее условиям

лицензии GNU General Public License v2. Разработчиком системы GeoNetwork является организация FAO (Food and Agriculture Organization of the United Nations).

Среди основных свойств данного ПО необходимо отметить следующие:

- Поддержка разнообразных стандартов метаданных, в том числе ISO 19115 и 19139;
- Возможность задавать собственные профили ISO 19115;
- Создание, редактирование, импорт элементов метаданных;
- Возможность реализовать поиск метаданных по многим критериям, в том числе – геопространственным;
- Возможность поддержки OGC CSW как в роли клиента, который выполняет сбор информации из других каталогов (harvesting), так и в роли сервера, который может быть описан вышеуказанным каталогом;
- Возможность локализации.

GeoNetwork можно интегрировать с многими элементами информационной инфраструктуры, как с свободнораспространяемыми, так и коммерческими. Для хранения метаданных могут быть использованы следующие базы данных:

- McKoi (используется для отладки);

- MySQL;
- PostgreSQL;
- Oracle.

В качестве сервера приложений, в который интегрируется Geonetwork, могут выступать свободно распространяемые продукты Jetty и Tomcat, либо коммерческий

IBM Websphere. Такая гибкость позволяет интегрировать Geonetwork в существующую информационную инфраструктуру.

На рис. 5 и 6 изображены реализованные интерфейсы поиска метаданных и интерфейс редактирования элемента метаданных соответственно.

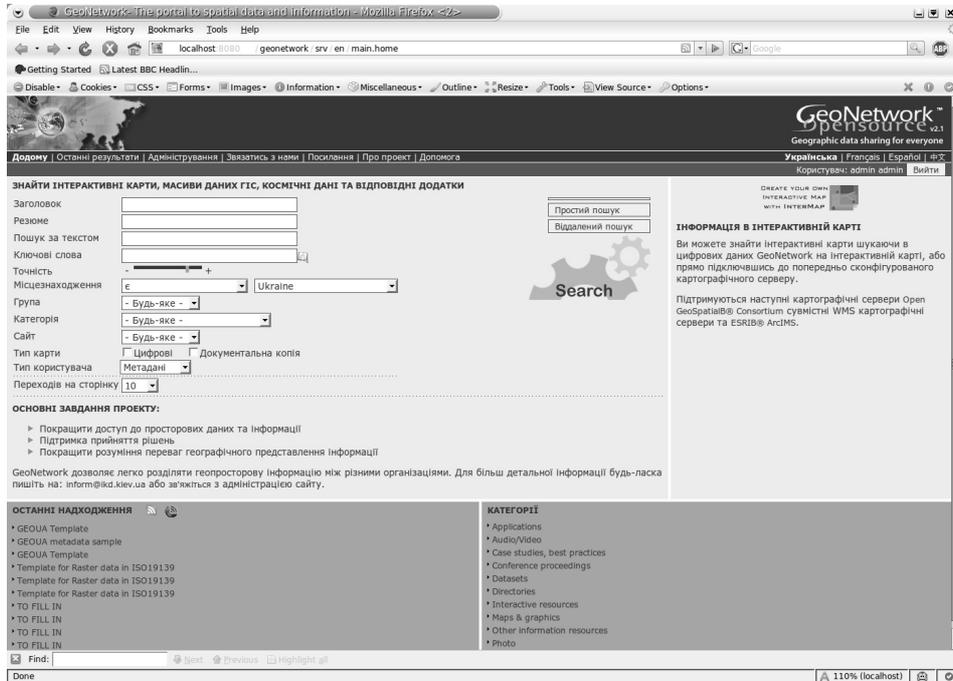


Рисунок 5 - Поисковый интерфейс ПО GeoNetwork

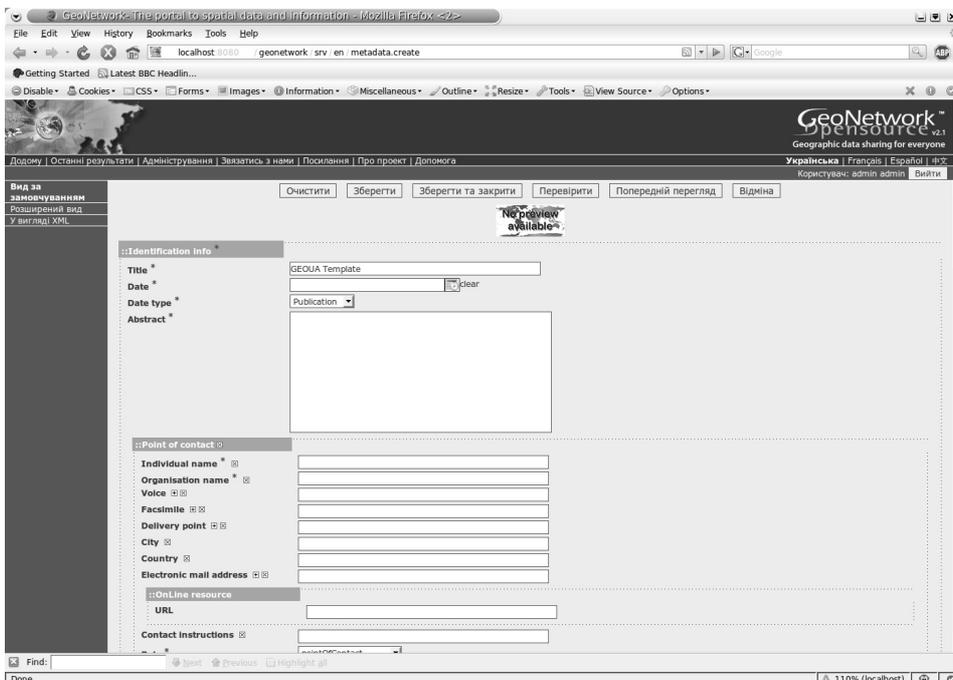


Рисунок 6 – Интерфейс редактирования метаданных GeoNetwork

Выводы

В данной статье предложен концептуальный подход к созданию системы каталогизации, с учетом мирового опыта построения таких систем, а также описана его реализация в системе GeoNetwork.

Создаваемые каталоги данных и метаданных планируется интегрировать с иерархическим каталогом метаданных системы систем GEOSS.

Работа выполнена при поддержке гранта УНТЦ-НАНУ «Разработка Grid-технологий интеграции данных разной природы» (№4928)

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 GEOSS. 10-Year Implementation Plan Reference Document. Noordwijk, Netherlands: ESA Publication Division. — 2005. — 212 p

Поступила в редколлегию 06.03.2009