

Принципы построения объединенной e-Инфраструктуры геоинформационных ресурсов для оптимизации процедур решения тематических задач дистанционного зондирования Земли*

М.А. Попов¹, Е.Б. Кудашев², С.Ю. Марков¹, М.В. Топольницкий¹, С.А. Станкевич¹

¹ Научный Центр аэрокосмических исследований Земли ИГН НАН Украины

² Институт космических исследований РАН

Аннотация

Статья посвящена обоснованию объединенной российско-украинской e-Инфраструктуры геоинформационных ресурсов, целями которой является обеспечить исследователям простой и быстрый доступ к геопространственным данным и иной научной информации, а также автоматизировать некоторые процедуры обработки и анализа данных. Рассмотрены базовые принципы построения такой объединенной инфраструктуры, а также предложена ее архитектура. Сформулированы направления дальнейших исследований.

1. Введение

Дистанционное зондирование Земли (ДЗЗ) из космоса – научно-технологическое направление, история которого насчитывает около столетия. За этот период появились видовые (иконические) технические средства, способные формировать с космической орбиты высококачественные изображения Земли и расположенных на ней объектов в самых разных диапазонах электромагнитного спектра, накоплены громадные массивы уникальной космической информации, созданы эффективные компьютерные методы обработки и анализа спутниковых данных [1]. Все это, вкупе с современными достижениями в области сетевых информационных технологий, дает многообещающую возможность использовать технологию ДЗЗ как уникальное и мощное средство для решения многочисленных и актуальных задач: народнохозяйственных, устойчивого развития территорий, научных, безопасности и обороны и др.

Необходимым условием реализации такой возможности является наличие соответствующей электронной (e) инфраструктуры в виде простран-

ственно распределенной телекоммуникационной системы, позволяющей организовать и поддерживать единое (интегрированное) информационное поле для сообщества создателей, держателей и пользователей данных. При этом очевидно, что для эффективной навигации в приложениях, т. е. обеспечения возможности находить с необходимыми полнотой и точностью в ресурсной среде такого информационного поля интересующие пользователя данные, информацию, знания и др., в наибольшей степени подходит Интернет или иная подобная сетевая распределенная структура.

Однако при создании единого информационно-коммуникационного поля для облегчения решения различных тематических задач с привлечением данных ДЗЗ, эффективной навигации, осложняющими факторами являются следующие.

- предметная область (область применений) технологии ДЗЗ является весьма обширной и с трудом поддается формализации (строгому описанию);
- данные относятся к различным физическим полям (оптическим, геофизическим, геологическим, ландшафтным и др.), отчего отличаются неоднородностью, имеют самые разные формы и форматы представления;
- точность, достоверность, степень детализации различных исходных данных об изучаемых объектах, явлениях и процессах могут сильно отличаться;
- время, отводимое на поиск и манипуляции с данными, в ряде случаев весьма ограничено, что требует быстрой загрузки потоков больших массивов данных и проведения их максимально оперативного анализа;
- оценивание релевантности данных ДЗЗ, проведение даже относительно несложной компьютерной обработки требуют, как правило, специальных программных систем и наличия у пользователя определенных практических навыков работы с аэрокосмическими изображениями.

Кроме того, следует считаться с тем обстоятельством, что многие ведомства, организации, компании, действующие сегодня на рынке услуг ДЗЗ, пользуются собственными форматами представления данных, протоколами обмена и т. д., созданными достаточно давно и потому не полностью гармонизированными с

современными международными правилами и стандартами. Это в значительной мере относится, в частности, ко многим российским и украинским организациям, работающим с геоинформационными ресурсами в сегменте ДЗЗ.

Цель данной работы – предложить и обосновать принципы и архитектуру объединенной e-Инфраструктуры геоинформационных ресурсов для оптимизации процедур решения тематических задач ДЗЗ, учитывающую наличие выше перечисленных факторов и обстоятельств.

2. Принципы построения объединенной e-Инфраструктуры

Объединенная e-Инфраструктура должна обеспечивать предельно простой и достаточно оперативный доступ научных организаций различных стран к сетевым геоинформационным ресурсам при решении тематических задач, способствовать повышению уровня информационного взаимодействия в области ДЗЗ. Одновременно в настоящее время создается российско-украинский сегмент e-Инфраструктуры непрерывного доступа к научным ресурсам ДЗЗ, главной задачей которого является обеспечить эффективное решение задач устойчивого территориального развития.

Принципы, положенные в основу построения объединенной e-Инфраструктуры геоинформационных и научных ресурсов ДЗЗ, решающей поставленные задачи, состоят в следующем:

- возможность информационной настройки на конкретно поставленную задачу в рамках предметной области ДЗЗ;
- автоматизация составления плана решения заданной тематической задачи ДЗЗ;
- возможность поиска исходных данных, отвечающих требуемым задачей параметрам, и знаний по содержанию (семантике);
- единая политика администрирования;
- масштабируемость;
- целостность;
- интероперабельность.

При построении объединенной e-Инфраструктуры вся область практических приложений ДЗЗ была разделена в соответствии с классификатором, предложенным в [2], на следующие восемь классов:

- геология;
- землепользование;
- сельское хозяйство;
- экосистемы и лесное хозяйство;
- океан и прибрежная зона;
- водные ресурсы;
- метеорология и климатология;
- чрезвычайные ситуации.

Каждый класс покрывается совокупностью корневых тематических задач, на которые замы-

каются практически все известные задачи ДЗЗ в соответствующем секторе деятельности. Например, в классе «Землепользование» в [2] определены такие корневые тематические задачи, как картирование (1), земельный кадастр (2), анализ урбанизированных и промышленных территорий (3).

Возможность информационной настройки на предметную область означает способность к адаптации при использовании системы для решения задач в различных приложениях и областях знаний. Для этого система должна обеспечивать возможность интеллектуальной обработки метаданных, иметь набор библиотек сервисов для решения типовых задач различных классов, а также позволять синтезировать сложные алгоритмы, включающие наборы этих функциональных элементов.

Система должна обладать свойствами гибкого реагирования на изменения, которые характерны для прикладных научных задач, связанных с проблемами устойчивого территориального развития. Гибкость должна проявляться на уровне поддерживаемых данных, уровне процедур и функций их обработки, а также уровне семантической интерпретации решаемых прикладных задач.

Автоматизация составления плана решения заданной тематической задачи ДЗЗ предполагает наличие некоторых изначальных сведений и знаний об изучаемом объекте, в частности, его характеристик и параметров, особенностей функционирования. Это позволяет далее сформулировать перечень данных, необходимых для решения поставленной задачи. Кроме того, должны быть известны номенклатура и характеристики программного или алгоритмического обеспечения решающих процедур.

При автоматизированном составлении плана решения задачи в методологическом плане целесообразно использовать общий подход Д. Пойя [3], при компьютерной реализации весьма полезным может быть CASE-инструментарий [4].

Для успешного поиска исходных данных, отвечающих требуемым задачей параметрам, и другой нужной пользователю информации необходимо выполнение двух условий. Во-первых, сетевые информационные, программные и другие ресурсы должны быть структурированы в группы и инвентаризованы. Во-вторых, должно иметься простое и однозначное описание каждого отдельного информационного элемента (данных, информации, знаний).

В контексте поставленной в работе общей задачи ресурсы могут быть структурированы в следующие четыре группы.

1. Изображения, картографические и графические материалы.
2. Текстовые материалы (стандарты, нормативно-правовые материалы, научная литература, отчеты, диссертации и т. д.).
3. Внешние сервисы и порталы, базы и банки данных.
4. Динамические данные (видео, анимации, мультимедиа).

Сегодня процедуры составления описания данных, или формирования метаданных, в значительной мере регламентированы [5-7]. В частности, описание геопространственных данных определено международными стандартами ISO 19115:2003 Geographic information – Metadata, ISO 19115-2:2005 Geographic information – Metadata. – Part 2: Extensions for imagery and gridded data. Существуют также американский прототип FGDC-STD-001-1998 названных стандартов, а также российский профиль стандарта ISO 19115 в виде документа ГОСТ Р 52573-2006 «Географическая информация. Метаданные».

Метаданные на текстовые материалы в сетях формируются, в основном, с использованием стандарта Dublin Core [5]. Однако следует отметить, что на отдельные группы информационных материалов единые согласованные правила формирования метаописаний в настоящее время отсутствуют. Например, с появлением и введением в практику ДЗЗ многоспектральных и гиперспектральных аэрокосмических изображений стали создаваться библиотеки оптических спектров природных и искусственных объектов [8]. Однако пока отсутствуют стандартизованные на международном уровне правила описания такой информации.

Особенно сложным является формирование метаописаний содержания сетевых знаний (метазнаний). Решение этой задачи невозможно без четкой структуризации знаний, что само по себе является проблемой.

Единая политика администрирования e-Инфраструктуры обеспечивает согласованные между участниками правила доступа и использования ресурсов, унифицированные подходы к обеспечению информационной безопасности. Проведение единой политики администрирования является гарантией доступности и безопасности данных. Система администрирования должна поддерживать территориально распределенных пользователей, узлы и сети. Важная функция администрирования — многопользовательские создание, обработка и обмен геопространственными данными и информационными продуктами на их основе.

Масштабируемость инфраструктуры подразумевает возможность увеличения объема ее информационных ресурсов, расширения функциональности ее сервисов, совершенствования организации данных с целью повышения эффективности их обработки. Принцип масштабируемости гарантирует открытость e-Инфраструктуры к наращиванию, в том числе возможность развития систем инфраструктуры путем введения новых прикладных программ.

Под целостностью информации понимается корректность и непротиворечивость данных про объекты в информационной системе. В базах данных корректность информации часто обеспечивается реализацией доменных ограничений. Существует также понятие ссылочной целостности, по которой понимают корректность всех связей объекта с другими объектами. Инфраструктура как система должна поддерживать целостность во всех указанных смыслах. Реализация принципа согласованности данных позволяет устранять ненужную избыточность в системе, упростить контроль качества и непротиворечивости данных при их обновлении.

Интероперабельность – основополагающий принцип свободного информационного обмена между информационными системами, входящими в состав e-Инфраструктуры, вне зависимости от внутренней логической и физической организации систем. Требование свободного информационного обмена реализуется путем задействования стандартизованных сетевых протоколов обмена данными и языков описания данных и построения запросов (прежде всего, SQL и XML). Интероперабельность обеспечивается глобальной унификацией технических регламентов, протоколов передачи и форматов данных, а также метаданных и поддерживается целой группой специально разрабатываемых стандартов (ISO/IEC 11179 Information technology — Specification and standardization of data elements; ISO 19113:2002 Geographic information — Quality principles; ISO 19114 Geographic information — Quality evaluation procedures и др.).

3. Архитектура объединенной e-Инфраструктуры

Основываясь на рассмотренных выше принципах, может быть предложена общая схема решения тематических задач ДЗЗ, представленная на рис. 1. После постановки задачи программа-планировщик определяет, к какому классу данная тематическая задача относится, анализирует имеющиеся в базе модели тематических задач и выбирает (с участием оператора) наиболее близкую и адекватную модель. Выбор производится с использованием специальной метрики близости.

Далее выбранная модель задачи уточняется и дополняется, исходя из конкретных особенностей поставленной задачи. В частности, могут уточняться требования к номенклатуре и параметрам необходимых исходных данных, к точности конечного информационного продукта, к программным средствам и др.

После этого строится логическая схема методики обработки данных в соответствии с рекомендациями, разработанными в [10]. Основными технологическими операциями методики являются:

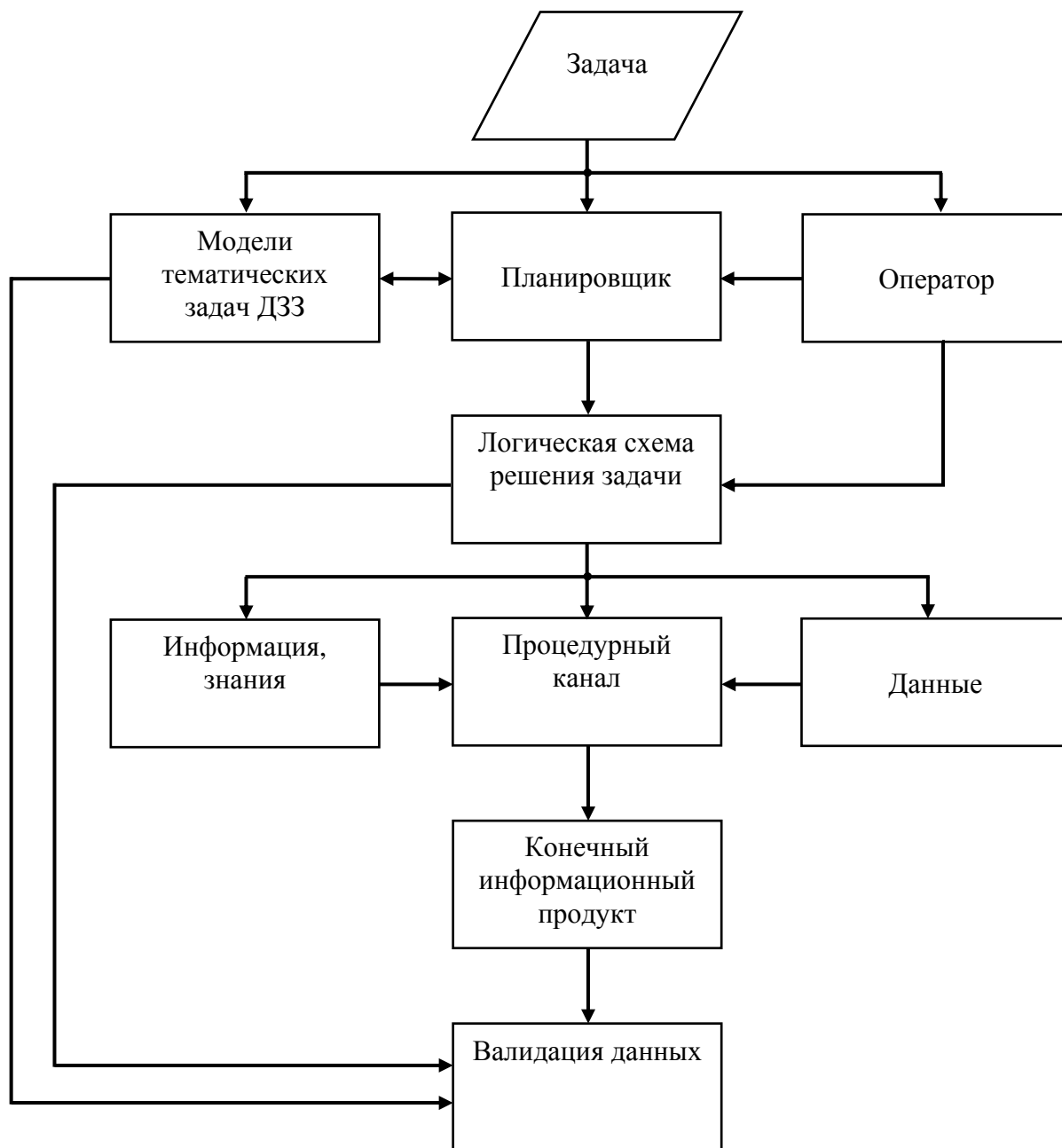


Рис. 1. Общая схема решения тематических задач ДЗЗ

- определение дешифровочных признаков объектов;
- получение репрезентов (обучение);
- анализ вспомогательных данных;
- выбор космических изображений;
- пространственное привязывание;
- классификация;
- интеграция данных;
- публикация конечного информационного продукта;
- оценка точности (валидация).

Эти, а также другие технологические операции, входящие в построенную для конкретной задачи ДЗЗ логическую схему методики обработки данных,

выполняются, если это возможно, с помощью средств программного обеспечения, имеющихся на локальных компьютерах и сервере организации. В случае недостаточности имеющихся в организации средств программного обеспечения, отсутствии в надлежащем объеме исходных данных, осуществляются поиск необходимых информационных материалов и программных средств в глобальной сети и привлечение их в логическую процедуру решения задачи.

После создания конечного информационного продукта в обязательном порядке проводится его валидация. Валидация проводится в соответствии с общими рекомендациями Committee on Earth Observation Satellites (CEOS), представленными на



Рис. 2. Архитектура объединенной e-Инфраструктуры геоинформационных ресурсов

портале <http://calvalportal.ceos.org> этого комитета. Порядок и правила валидации методик обработки данных описаны, например, в [11].

Исходя из рассмотренной схемы решения тематических задач ДЗЗ может быть предложена следующая архитектура объединенной e-Инфраструктуры геоинформационных ресурсов (рис. 2). Она включает следующие компоненты:

1. ГИС-сервер (GIS Server).

ГИС-сервер состоит из одного управляющего компонента Server Object Manager (SOM) и одного или нескольких компонентов-контейнеров Server Object Container (SOC). Он управляет геоинформационными ресурсами (карты, изображения, цифровые модели местности, текстуальная информация и т. д.) и транслирует их веб-приложениям в виде отдельных геосервисов.

2. Веб-сервер (Web Server).

Веб-сервер предназначен для хранения и управления геопорталом. Примером использования веб-сервера может быть Microsoft IIS для операционной системы Windows или Apache Tomcat 5.5.x. Структурно и функционально Веб- и ГИС-сервера могут быть развернуты на отдельных компьютерах. Решение относительно такой конфигурации принимается исходя из задач, которые решаются, количества пользователей, объема данных, финансирования и т.д.

3. Геопортал (Geoportals).

Геопортал представляет из себя совокупность отдельных приложений и сервисов. Используется

для публикации, администрирования и поиска стандартизированных геоинформационных ресурсов. Структурно геопортал состоит из таких элементов:

- веб-приложение (Geoportals Site Starter);
- система управления контентом (Content Management System);
- интерфейс программы-планировщика;
- интерфейс поиска данных (Widget for searching);
- интерфейс предпросмотра картографических данных (Previewer Map Interface);
- сервис загрузки данных (Data Extraction Service).

Геопортал расположен на веб-сервере и является клиентским приложением ГИС-сервера. Пользователи, взаимодействуя с ним, вызывают сервисы, которые являются посредниками между веб-сервером и ГИС-сервером и обращаются к ГИС-серверу через его компонент SOM за необходимым инструментарием (геосервисами). Этот инструментарий реализуется посредством специализированных объектов (так, например, для продуктов компании ESRI – это ArcObjects) на SOC-машинах.

Исходными данными геопортала могут быть локальные и глобальные метаданные, локальные и глобальные веб-сервисы, Вики-проекты, данные сервиса YouTube, документы SharePoint, KML и т. д.

4. Каталогный сервис.

Каталожный сервис предназначен для доступа, поиска, накопления, обработки метаданных геопространственных ресурсов [6, 9]. Для ускорения рабо-

ты каталога, который имеет большие объемы информации, для хранения данных используют распределенную сеть систем управления базами метаданных. Это позволяет проводить индексацию данных для ускорения процессов обработки запросов и повышения уровня их защиты. Каталогный сервис обеспечивает общий механизм для классификации, регистрирования, публикации, поиска, хранения и доступа к метаинформации об имеющихся в сети геоинформационных ресурсах. Виды сервисов каталога выделяются в зависимости от их функций (например, каталогизация данных), типов (например, типы географических объектов, покрытия, символы), он-лайнных данных (например, баз данных, хранилищ и библиотек условных обозначений), типов услуг и он-лайнных сервисов. Сервисы каталога предоставляют метаданные о зарегистрированных геоинформационных ресурсах, поддерживают формирование запросов на информацию по типам ресурсов. Они позволяют публиковать, создавать и редактировать наборы метаданных для геопространственных данных, геопространственных сервисов и геоссылок.

5. Сервис данных.

Сервис данных обеспечивает доступ к содержанию ресурсных информационных материалов, которые находятся в базах и банках геоданных. Процедура организуется таким образом, что каждой единице ресурсов, с которыми работает этот сервис, присваивается уникальное имя. Для повышения скорости поиска осуществляется индексация ресурсов.

6. База моделей тематических задач ДЗЗ.

Данная база предназначена для хранения моделей тематических задач ДЗЗ и позволяет с помощью интерфейса программы-планировщика выбрать наиболее адекватную (релевантную) модель в соответствии с поставленной задачей.

7. Программное обеспечение как сервис (Software as a Service (SaaS)).

SaaS подразумевает предоставление приложений для конечного пользователя в виде сервиса «по требованию» вместо его установки на конкретном рабочем месте или на собственном сервере [12]. Типичным примером услуг SaaS является компания ITT Visual Information Solutions, разработавшая программный продукт ENVI for ArcGIS Server [13]. Данный продукт позволяет публиковать инструменты ENVI на ГИС-сервере, тем самым делая их доступными для всех пользователей внутренней и/или внешней сети. Инструменты ENVI при этом представляются как сервисы при подключении пользователей к серверу, на котором установлены программы ArcGIS Server и ENVI for ArcGIS Server. Функционирование ENVI for ArcGIS Server осуществляется в соответствии с той же идеологией, которая была заложена в программном продукте ArcGIS Server. Согласно этой идеологии работа с инструментами и функциями ENVI на серверном уровне осуществляется в три этапа: создание, публикация и использование.

Кроме того, в рамках «облачных» вычислений [14] существуют и другие направления: платформа как сервис и инфраструктура как сервис.

Платформа как сервис (Platform as a Service (PaaS)) – предоставляется платформа и/или промежуточное (связующее) программное обеспечение в виде сервиса, на которых возможна разработка и развертывание пользовательских приложений. Типичными решениями такого типа являются интерфейсы прикладного программирования (API) и инструментальные средства, а также базы данных и системы управления рабочими процессами, интегрированные средства обеспечения безопасности. Эти решения позволяют разработчикам создавать приложения и запускать их в инфраструктуре, принадлежащей и поддерживаемой поставщиком облачных услуг.

Инфраструктура как сервис (Infrastructure as a Service (IaaS)) – охватывает аппаратные средства и технологию для компьютерных вычислений и хранения данных, операционные системы и другую инфраструктуру, которые предоставляются не как локальные ресурсы, а опосредованно – через обращение к сервисам, размещенным на стороне провайдера.

8. Клиентские ГИС-приложения (Clients).

Клиентские ГИС-приложения устанавливаются на конечных устройствах локальных и удаленных пользователей и подключаются по протоколу HTTP к геопорталу по сетям LAN или WLAN для поиска, публикации и сохранения стандартизированных геоинформационных данных.

4. Заключение

Таким образом, в работе рассмотрены принципы построения объединенной российско-украинской е-Инфраструктуры геоинформационных ресурсов, позволяющей оптимизировать процедуры решения самых разных тематических задач ДЗЗ. Дана общая схема решения тематических задач ДЗЗ, позволяющая formalизовать процессы обработки и анализа данных. Разработана архитектура объединенной е-Инфраструктуры, основанная на геопортальных технологиях и возможностях, предоставляемых глобальными сетями.

Предметом дальнейших исследований является:

- разработка эффективных процедур обмена геопространственными данными и другими научными ресурсами между исследователями, алгоритмическое обеспечение оптимальных условий для устойчивого управления данными и ресурсами ДЗЗ;
- создание профилей стандартов метаданных для описания гетерогенных ресурсов спутниковой информации;
- построение портала геоинформационных ресурсов ДЗЗ, обеспечивающего распределенный стандартизированный доступ исследователей и удаленных пользователей к ресурсам спутниковой и иной научной информации.

Создание реально действующего ГИС-портала позволит интегрировать информационные ресурсы, распределенные на различных сетевых серверах и на компьютерах пользователей и таким образом повысить эффективность решения тематических задач ДЗЗ.

Литература

- [1] Elachi, C. Introduction to the Physics and Techniques of Remote Sensing / C. Elachi, J. van Zyl. John Wiley & Sons, Inc., 2006. – 552 p.
- [2] Багатоспектральні методи дистанційного зондування Землі в задачах природокористування // За ред. В.І. Лялька і М.О. Попова. – К.: Наукова думка, 2006. – 360 с.
- [3] Пойа, Д. Математическое открытие. – М.: Наука, 1970. – 452 с.
- [4] Люгер, Д.Ф. Искусственный интеллект. Стратегии и методы решения сложных проблем. – М.: Вильямс, 2003. – 864 с.
- [5] Kresse, W. ISO Standards for Geographic Information / W. Kresse, K. Fadae. – Berlin: Springer-Verlag, 2004. – 322 p.
- [6] Попов, М.А. Система каталогизации и распределенного доступа к данным дистанционного зондирования Земли: концепция, архитектура, реализация / М.А. Попов, С.Ю. Марков, С.А. Станкевич, А.Л. Бодня // Космічна наука і технологія. – 2007., – Т. 13. – №3. – С. 45-54.
- [7] Куцуль, Н.Н. Каталог метаданных системы GEO-Ukraine / Куцуль Н.Н., Лавренюк А.Н., Лавренюк С.И., Грипич Ю.А. // Наукові праці ДонНТУ. Серія «Інформатика, кібернетика та обчислювальна техніка». – 2009. – № 10 (153). – С. 92-100.
- [8] Baldridge, A.M. The ASTER spectral library. Version 2.0. / A.M. Baldridge, S.J. Hook, C.I. Grove, G. Rivera // Remote Sensing of Environment. 2009. 113. P. 711–715.
- [9] Кудашев Е.Б. Распределенная геоинформационная инфраструктура пространственных данных в задачах устойчивого развития территорий / Е.Б. Кудашев, М.А. Попов, С.Ю. Марков // Труды XVI Всероссийской научно-методической конференции «Телематика'2009». Т. 2. – СПб., 2009. С. 272-274.
- [10] Методика вирішування тематичної задачі «Класифікація покривних елементів ландшафту Землі». – Дніпропетровськ: ДП «Дніпрокосмос», 2004. – 76 с.
- [11] Попов, М.О. Правила та процедура сертифікації методик використання даних дистанційного зондування Землі при вирішенні тематичних задач / М.О. Попов, Т.В. Цимбал // Космічна наука і технологія. – 2004. – Т. 10. – №4. – С. 104-112.
- [12] Кошкарев, А.В. Картографические Web-сервисы геопорталов: технологические решения и опыт реализации / А.В. Кошкарев, В.С. Тикунов,

С.А. Тимонин // Пространственные данные. – 2009. – № 3. – С. 6 – 12.

- [13] Ялдыгина, Н.Б. Новое в линейке программных продуктов ENVI: ENVI 4.8, серверное приложение ENVI for ArcGIS Server / Ялдыгина Н.Б. // Geomatica. – 2011. – № 1. – С. 35-41.
- [14] Гохман, В. ArcGIS в облаке // ArcReview. – 2010. – № 3 (54).

Development of e-Infrastructure of geoinformation resources for optimization of decision procedures of EO Remote Sensing thematic problems

M.A. Popov, E.B. Kudashev, S. Ju. Markov, M.V. Topolnizkii, S.A. Stankevich

The main aim of this paper is to improve the accessibility of Russian and Ukraine Satellite Remote Sensing data, generating satellite information resources, linking the information resources of Russia and Ukraine with the European eo/Porttal (metadata/catalogue) system in order to provide interoperable catalogue access and ensuring that end users are involved, provide a single point of access to data.

Thus, International Scientific community of Remote Sensing has gained the opportunity to access data of Russian and Ukraine EO satellite data providers via a global distributed system. Technologies and principles of building distributed systems considered in this article have great potential for further development, and even today they provide extensive means for arranging distributed heterogeneous data and services in a single global system. The future plan for the next generation of the distributed informational system consists of three main parts: a) correcting shortcomings of the existing system, b) enhancing technical and informational characteristics, and c) providing solutions to new problems which were unknown during the development of the original system.

* Работа выполнена в рамках совместного российско-украинского проекта № Ф40.7/24 «Российско-украинский сегмент глобальной e-Инфраструктуры постоянного доступа к научным ресурсам ДЗЗ для решения задач устойчивого развития территорий» при финансовой поддержке российского и украинского государственных фондов фундаментальных исследований.