

**RAMADDA: Content Management for the Earth Sciences**  
**Jeff McWhirter**  
**jeff.mcwhirter@gmail.com**  
**UNAVCO**

RAMADDA is an open source data and content repository framework. Developed over the last 3.5 years, RAMADDA provides a rich suite of services for creating, managing and accessing a wide range of digital content with a focus on geoscience data. In its development we have tried to touch on a broad range of functionality and user needs - installation, configuration, data ingest, metadata harvesting, publication, knowledge management, search, browse, federation, monitoring, integration with external services, and rich data services.

In this paper we describe a number of the features that RAMADDA provides in order to highlight the range of functionality and services that we believe a geoscience data/information management system needs to provide.

### **An Exercise**

But first, let's start with a simple exercise:

Find all relevant meteorological data for the southeastern United States for the week of August 9th, 2005. This includes radar data, meteorological station data, satellite imagery, atmospheric soundings, forecasting models and, of course, hurricane trajectory model output.

The event in question is Hurricane Katrina and you could probably find a small portion of this data at various data centers, but it would be a non-trivial effort. Here we have one of the most well-known and destructive storms ever to hit North America and there is no centralized place to find data that describes this event.

Now, you may respond:

*But, nobody gathered that data at the time so of course we cannot find it all in one place, so its an unfair task*

Well, let's rephrase the above task:

You know Katrina is going to strike next week and you are tasked with gathering all of the above data. You need to organize this data and make it available for discovery and effective access. Make it so.

Luckily with real-time data distribution systems like Unidata's LDM/IDD, at least you will be able to access most of the above data. However, what technology would you use (or could you use) to make available this data to your user community? The answer to this is

that you would most likely place it on an FTP server and any descriptive information (i.e., metadata) would probably be in a README file.

It is sad to say, but with all of the vast investments and development efforts that the geoinformatics community has made over the last decade in cyberinfrastructure, the only technology that is readily available to you is probably an FTP server. While useful and surprisingly effective, FTP does not represent much advancement in geoinformatics, as it is a 40-year-old protocol. Even many national data centers rely on FTP and README files as their primary means of data management and dissemination.

Now, lets move beyond data. The EarthCube Guidance document states:

*The goal of EarthCube is to create a knowledge management system and infrastructure that integrates all geosciences data in an open, transparent and inclusive manner.*

Note the emphasis on knowledge. Data is not knowledge. Words and pictures are not knowledge. Knowledge needs to be content coupled with data. Furthermore, knowledge cannot exist in a vacuum - knowledge has to exist among a community of practitioners - scientists, teachers, students and the general public.

Let's expand upon our exercise and include using the above data set to enable knowledge creation.

Have a group of students explore the Katrina data by generating visualizations and animations. Have them generate a set of image and Google Earth products, create web content describing what is shown and make it available on the web. Allow other students to comment on these visualizations.

Most likely this would be a difficult task as well with any current geo-informatics system. Ironically, these are tasks that millions of people do every day in their interactions with sites like YouTube, Facebook, wikis, etc.

## **RAMADDA**

The key driver behind the design of RAMADDA has been this idea of integrating data, content and collaboration. RAMADDA provides a broad set of services for creating, organizing and accessing all of the digital “stuff” that is important to the scientific endeavor. This stuff might be a climate model NetCDF file, a Powerpoint file, a wiki page or a collection of images from a field project.

A critical aspect of RAMADDA is that it is a software *framework*. It provides a set of core facilities and services but can be easily extended through its plugin architecture to provide support for new types of digital stuff (content and data), new ways of describing that stuff (metadata), new ways of looking at that stuff (views) and new ways to access that stuff (services).

## **Core content model**

Underlying RAMADDA is a general content model. All entries in the repository are described with a common set of metadata that captures basic spatial, temporal and descriptive information. RAMADDA makes no distinction between a data file, a collection, a Word document or a collection of images. This allows for new approaches to basic content management, e.g., georeferencing images, attaching DIF metadata to Word documents, doing temporal searches for collections.

## **Metadata**

There is an alphabet soup of metadata standards - ISO, DIF, FGDC, Dublin Core. The approach we have taken in RAMADDA is to not focus on a single metadata standard but rather develop a declarative extensible framework that provides support for a wide range of metadata standards and services. The core services include:

- Interactive and programmatic creation of metadata elements
- Metadata storage
- Faceted metadata search
- Metadata HTML display
- Metadata crosswalk

There are also metadata harvesting mechanisms that support the automatic extraction of spatial, temporal, attribute and variable metadata from NetCDF data files, JPEG images, LiDAR data, etc.

## **Model/View Architecture**

Presentation and access to digital content in RAMADDA is provided through a model/view framework. We have decoupled the core data model (repository entries and metadata) from any particular representation of the data model. This has enabled the development of a wide range of content encodings – HTML displays, maps, Google Earth plugin, calendars, timelines, RSS, ATOM, KML, ICAL, Thredds Catalog, DIF, etc.

## **Information and Knowledge Management**

The same content and access services that RAMADDA provides for geoscience data also supports traditional informational content. RAMADDA provides facilities for [wikis](#), weblogs, [FAQs](#), [glossaries](#), [workshops](#) and [ontologies](#).

## **Federation**

A RAMADDA repository can support federated [search](#) across collections of external RAMADDA repositories. Currently, it is assumed that each remote repository is a RAMADDA server so the search functionality and data model is consistent. RAMADDA does support other federated search protocols such as [Open Search](#).

## **Rich data services**

A number of rich data access services have been developed as RAMADDA plugins. The [NLAS/RAMADDA](#) server running at UNAVCO provides access and visualization services for LiDAR data. RAMADDA provides services for [geoscience and sociological observation data](#) and [gridded data](#).

## **Packaged Software**

The community needs to move away from building large and complex custom software systems that only run in a single environment. When evaluating data management systems one need only ask:

*Can I download, install and configure this system?*

Too often, the answer to the above question is no.

From its inception RAMADDA has been built as packaged software. You can easily download and install a WAR to run under a servlet container (e.g., Tomcat) or alternatively RAMADDA is provided as a stand-alone package that only requires Java. RAMADDA is also built on a database abstraction layer that enables it to run under MySQL, Postgres, Oracle or the built-in Java Derby database.

## **Conclusion**

The best way to gain a further understanding of the capabilities that RAMADDA can provide is to explore actual installations of the software. A number of RAMADDA sites are listed here:

<http://facdev.unavco.org/repository/alias/ramadda>