Building a Massive Volcano Archive and the Development of a Tool for the Science Community

Final Project Report

Justin Linick
David Pieri, Mentor

Overview

The Jet Propulsion Laboratory has traditionally housed one of the world's largest databases of volcanic satellite imagery, the ASTER Volcano Archive (~10Tb), making these data accessible online for public and scientific use. However, a series of changes in how satellite imagery is housed by the Earth Observing System (EOS) Data Information System has meant that JPL has been unable to systematically maintain its database for the last several years. We have provided a fast, transparent, machine-to-machine client that has updated JPL's database and will keep it current in near real-time. The development of this client has also given us the capability to retrieve any data provided by NASA's Earth Observing System Clearinghouse (ECHO) that covers a volcanic event reported by U.S. Air Force Weather Agency (AFWA). We will also provide a publicly available tool that interfaces with ECHO that can provide functionality not available in any of ECHO's Earth science discovery tools.

Background and Goals

ASTER, short for Advanced Spaceborne Thermal Emission and Reflection radiometer, is currently orbiting the earth on the TERRA platform, as part of the NASA EOS Mission. ASTER has been collecting data since February of 2000, and provides high-resolution images of Earth in 15 different bands of the electromagnetic spectrum, covering a range from the visible to the long wave infrared. These data are used in the creation of maps of land surface temperature, emissivity, reflectance, and elevation. A subset of data has been used in the creation of the Aster Volcano Archive, or the AVA. Currently, this archive is a collection more than 1,500 Very Near Infrared (VNIR) daytime volcano images, and has been made available for scientific use and for the general public. The AVA can be found here: http://ava.jpl.nasa.gov/ and is built and maintained by Howard Tan, Dave Pieri and others at JPL.

For many years the ASTER data was housed within the NASA Land Processes Distributed Active Archive Center (LP DAAC, Sioux Falls, SD), which is a distributed archive center that allows processing and distribution of the ASTER data. At JPL the AVA was kept current using scripts that would login to the LP DAAC, pull out the new data, and update the archive. However in August of 2010 the ASTER meta-data cataloging and maintenance function was moved to a new data system entity, called ECHO. Beyond that point, from the perspective of AVA, systems used for the distribution and download of data were dramatically changed, thus effectively rendering useless AVA's current algorithmic processes. Lacking the resources to invest in building an entirely new system to interact with ECHO, which itself subsequently underwent a series of system architecture changes, the AVA was unable to update its archive. Current methods for connecting to ECHO and retrieving data involve implementation of the REVERB client. However, REVERB, while excellent for small subsets of data, is impractical for AVA's database of more than 1,500 objects. What was needed is an automated process that can interact with ECHO's systems on a machine-to- machine level. Our goal for the project was to complete this process in a manner that allows for a new, complete version of AVA to be kept current while using a minimal investment of human resources for maintenance.

Initial Period

The initial work toward our goal of building a process to update the AVA was very challenging. It took time to research and identify various possible directions that could meet our needs, especially because at the time we were still unsure if any of these methods were even feasible. We initiated several lines of attack, including a detailed Simple Object Access Protocol (SOAP) Java client, which ended up being impractical. It also took some time to identify and contact various persons within ECHO and other communities that could assist us with questions and documentation. At the same time, several of our previous contacts had either transferred or left their positions and were no longer available. After considerable discussion and a prototyping effort, we identified the ECHO REST Application-Program Interface (API) as the best method to start building a process to update the ASTER Volcano Archive.

The ECHO REST API utilizes a RESTful method of communication with ECHO's servers. REST, short for REpresentational State Transfer, essentially is a style of software architecture that takes advantage of existing standards built for the World Wide Web, such as HTTP, URI or XML. It implements these standards in order to facilitate communications and data transfer. This has particular advantages as it is based on proven, existing architectures and is oftentimes faster and less proprietary than, for example, a SOAP implementation. The fact that the architecture appeared to be both functional and well supported, in addition to a recommendation by the ECHO support team, was the reason behind our decision.

Construction of our AVA client

With assistance from the support team at ECHO, we successfully built a process that could utilize the ECHO REST API to login, retrieve a token, retrieve a user id, and use all of these to generate a search query and retrieve the granule results ("granule" is the name given to individual images within the ASTER database). We were also able to retrieve the ECHO metadata for those granules. This allowed us to build a list of granules that covered the 1,500+ objects (i.e., all the volcanoes listed as having been active over the last 10,000yrs, as identified by the Smithsonian Institution Global Volcanism Program) in our database over the entire period of the ASTER mission. Next we built a process to open an empty order, generate options for our orders, add granules to the order, and submit the order through ECHO to whatever data center was housing the appropriate data. This effectively allowed us to leapfrog the ECHO/REVERB client, and instead of attempting to build orders manually, we gained the ability to build and submit orders quickly, efficiently, and on a machine-to-machine level.

However, the ECHO/REVERB system was not designed with the expectation of complete machine-level integration. After a user submits an order, a series of emails are generated informing the user of the processing chain, eventually resulting in an email with a series of file transfer protocol (ftp) links that the user can use to data. This is great for human beings, who can login to their accounts and simply click on a link, but it is a rather inefficient method of machine-to-machine communication. However, since the system was built on the assumption of human-to-machine interactions, it was what we were made to accommodate.

In order to retrieve our order submissions we built additional methods that took advantage of Internet Message Access Protocol (IMAP) and several of Python's standard libraries to login, retrieve, and parse the generated emails for ftp addresses. We then verify the addresses and use them to retrieve our orders via ftp. We also built several processes that verify and track all of the granules, as well as each of the orders, to keep the total number of submitted orders under a user-specified level at all times. This is done to prevent any overloading of ECHO or it's connected servers with order submissions. We also instituted a process whereby anyone can halt the AVA process remotely with a simple email. Finally, we built a processing chain that will generate RGB images and KML super-overlays from our data, and places those on the AVA (ver. 2) database. Soon we will add the final step to that chain which makes them publicly available. This will again allow anyone in the world to see the AVA (ver. 2) data archive, now completely updated and current, and to click on a link to see the imagery itself, or view the ASTER image embedded within Google Earth in .kml format. Each of these files will be annotated with the proper names and positions of the volcanoes in the image, as per the Smithsonian database, with links to their Smithsonian descriptions. Additionally, the entire processing chain can be automated, such that it will keep the AVA maintained and current automatically, with the web server being populated with new data as it is received from the spacecraft. A diagram of steps involved in this process has been appended to this report.

Integration of the AFWA data

The ability to automate searches, submissions and retrieval for all of the data housed by the ECHO database put us in a unique position to take advantage of other datasets as well. The Air Force Weather Agency (AFWA) issues reports of volcano events; these are used worldwide in the direction of air traffic, as the fine silicate dust that comprises volcanic ash is extremely hazardous to aircraft, particularly to turbine engines. The advantage of using the AFWA reports is that we are able to generate a database of volcanic events, for both location and time that stretch back almost a decade. We integrated this into our existing processing chain, such that we now have the unique ability to find and retrieve every frame in the past decade of the ASTER archive that covers every global volcanic event reported by AFWA, both before and after eruption. We do not believe this has ever been previously achieved at our level of coverage for both area and time, or in terms of pure data volume.

Development of a Generic Client

In addition to the development of a process to update the AVA (ver.2) database, we realized that some of the functionality we have developed could be useful to the general scientific community. While most users go through REVERB in order to retrieve their data, it is designed around a manual ordering process. For an individual with a large and complex dataset, a separate tool for accessing data from the ECHO clearinghouse could be useful. This is why, alongside the development of the AVA update process, we have been developing a separate tool that will allow individuals to automate the retrieval of any data available through the ECHO clearinghouse. While the client is not yet complete, as our priority has been updating the AVA, we intend to eventually make the tool freely available for general use, as has been requested of us by the NASA LPDAAC.

Conclusions

As of August 20, 2012, the ASTER data ingest process for the rebuilding and updating of AVA (ver. 2) has still yet to be finalized. For the ASTER daytime imagery, almost 6 TB of an estimated 9-10 TB has been retrieved. We hope that by the end of August a fully updated and now complete archive will be made available to the general public and to scientific users. There is much additional work that can be done using our new tools, including data mining for volcanic event precursors, development of advanced methods of searching and querying AVA (ver. 2) data, and alternative means of building and presenting ASTER data. There are also opportunities involving completion of the generic client and research using the AFWA dataset.

Special thanks are given to those who have provided support and guidance during this project. Specifically I would like to thank the following: Dave Pieri, JPL; Howard Tan, JPL; Dan Pilone, ASTER/ECHO API support; Kathleen Baynes, ASTER/ECHO API support; Dawn and the Engineering staff at LPDAAC user services, USGS Earth Resources Observation and Science (EROS) Center in Sioux Falls, South Dakota. And to the JPL Education Office for making this all possible.

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the 2012 JPL Summer Internship Program and the National Aeronautics and Space Administration.

AVA Update Process