

Michigan Tech HAZHub Natural Hazards Archive – User Manual

1. Background

The vision for the Michigan Tech Natural Hazards Archive is to provide web access to data files that are of interest to the research community. Natural Hazards Archive (HAZHub) was developed as the Data Archive Plan of the NSF project, "[Remote Sensing for Hazard Mitigation and Resource Protection in Pacific Latin America](#)". This project was sponsored by the National Science Foundation's Office of International Science and Engineering (OISE). The initial archived data was directly related to the research performed by grad students and faculty involved in the project, but the archive is a pro-active archive system that will continue to maintain archives of data sets from current and future geologic hazards and resources research performed at Michigan Tech and other collaborating institutions. Geologic hazards and resources include research in the areas of geohydrology, volcanology, and seismology. The Michigan Tech data collection consists of many types of datasets which cover a wide range of geographic and geologic areas.

The backbone of the database is the open-source [integrated Rule-Oriented Data System \(iRODS\)](#), developed by the Data Intensive Cyber Environments Center (DICE Center) at the University of North Carolina at Chapel Hill. Development of the core iRODS system has been funded by the National Science Foundation and the National Archives and Records Administration.

Current Michigan Tech PIRE datasets include satellite images, ArcGIS maps, seismic data, thermal imaging, technical papers, video, measured and processed data, and other databases related to specific topics and research areas. The types of data to be cataloged are not limited to this list. iRODS will allow new data types to be added to the system as needed. These datasets may be stored anywhere; they need not all reside on the same server.

The data catalog will store descriptive information about each dataset in a MySQL database. This will include searchable fields such as location coordinates, data type, and descriptive keywords. The database will also hold detailed descriptions, usage restrictions, and reference citations.

“iRODS” is the name of the online tool used to catalog the data, but Michigan Tech’s instance of the iRODS database is called “**HAZHub.**” Sometimes these terms are used interchangeably, but they both refer to the underlying database system.

Michigan Tech’s data archive is connected to 700GB of storage space, which has the potential for expansion as needed. Files have been grouped together in tar and gzip formats, and the targeted size for any one file stored in the system is 1-2 GB. To date, 400 GB of data have been uploaded into the repository. The uploaded data is mostly volcano-related datasets from Guatemala and El Salvador, but students are in the process of uploading more volcano, seismic, and geohydrology data.

2. Contents

1. Background.....	1
2. Contents	2
3. User Documentation	3
3.1. Online Access to the HAZHub Natural Hazards Archive.....	3
3.2. Request UserID and Password.....	4
3.3. Access to HAZHub	4
3.3.1. HAZHub iDROP.....	5
3.3.2. HAZHub Online.....	6
3.4. HAZHub Data Content.....	7
3.5. Directory (Collection) Structure and Naming Conventions.....	8
3.6. Metadata.....	10
3.7. HAZHub Search tips	12
3.8. Uploading and Downloading Files.....	16
3.9. Delete Files.....	17
4. Criteria to include data in HAZHub:	17
5. Procedures for Contributing Data to HAZHub.....	18
6. Developments still in progress.....	19

3. User Documentation

3.1. Online Access to the HAZHub Natural Hazards Archive

HAZHub Natural Hazards Archive online documentation

URL: <http://www.mtu.edu/geo/research/hazard-data/>

Access the archive through the [Geological & Mining Engineering & Sciences](http://www.mtu.edu/geo/) (<http://www.mtu.edu/geo/>) web page. Under RESEARCH, select **HAZHub Natural Hazards Archive**.

This website documents the use and expectations of the repository. Userid/password requests are initiated here, and this site provides links to two user interface methods of accessing the data in the archive, HAZHub iDROP and the HAZHub Online web browser.

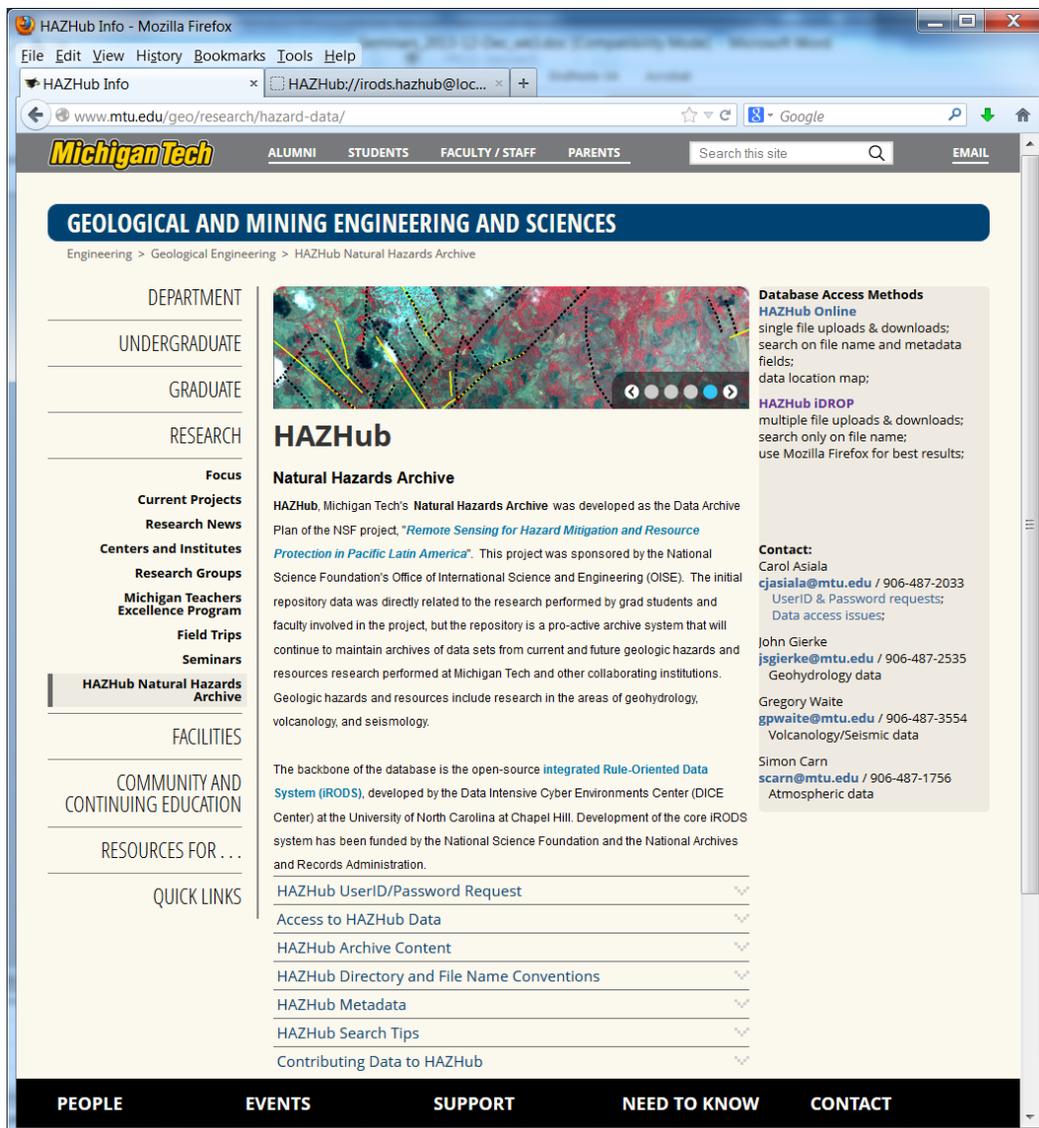


Figure 1. The HAZHub: Natural Hazards Archive web page, is located under the Geological & Mining Engineering & Sciences web site, under RESEARCH.

3.2. Request UserID and Password

Userids and passwords are required to access the data archive, and permissions will be set according to the user's level of use. The root directory of the HAZHub archive is `\hazhub\home\`. Data is then divided into "`\PublicAccess\`", "`\PrivateAccess\`", and "`\Users\`" directories.

`\PublicAccess\` contains datasets that are accessible by the general public. Data should still be cited correctly when referenced in publications and presentations, but it is freely accessible and usable for research and publication. All users will be given access to view and download data in this directory.

`\PrivateAccess\` contains proprietary data that is only accessible by Michigan Tech researchers and students, and is limited in use by copyright and licensing restrictions. General public users will be able to see directory and file names, but will not have permissions to view or download data.

`\Users\` contains a subdirectory for each user with access to the HAZHub archive. This directory is used when Michigan Tech researchers and collaborators need to upload new data into the archive. The archive administrators will review the data and make sure the file & directory names are in compliance with the HAZHub naming conventions, and will also verify that a complete and accurate set of metadata has been provided for each file before transferring data sets to the correct public or private access directory.

To request a userid/password, please email Carol Asiala (cjasiala@mtu.edu) or use the [Online Request form](#). Please include an **Email address** and **Preferred UserID**. If you are requesting access to the `\PrivateAccess\` directory, please provide Michigan Tech advisor/referral name and email address.

Your password can be changed through the iDROP interface. The same UserID and password will work when logging in with either access method, **HAZHub iDROP** or **HAZHub Online**.

3.3. Access to HAZHub

iRODS (www.irods.org) is the open-source software installed on Michigan Tech's server that manages the data archive. **HAZHub iDROP** and the **HAZHub Online** are two web-based user interface programs that allow users to view and access the iRODS-managed database. Both interfaces access the same data, but they each provide unique features that will be helpful to users, depending on their use of the Natural Hazards Archive.

HAZHub iDROP - is a Java Web Start application that can be run directly from a web browser. It is the most current interface being developed by the iRODS organization. It is a working version available for use, but new developments are still in the process of being added.

HAZHub Online - is a tradition web page built on PHP and JavaScript. It is an older interface written in PHP and Javascript. It has search features not yet developed in the iDROP system, but is limited to single file uploads and downloads.

3.3.1. HAZHub iDROP

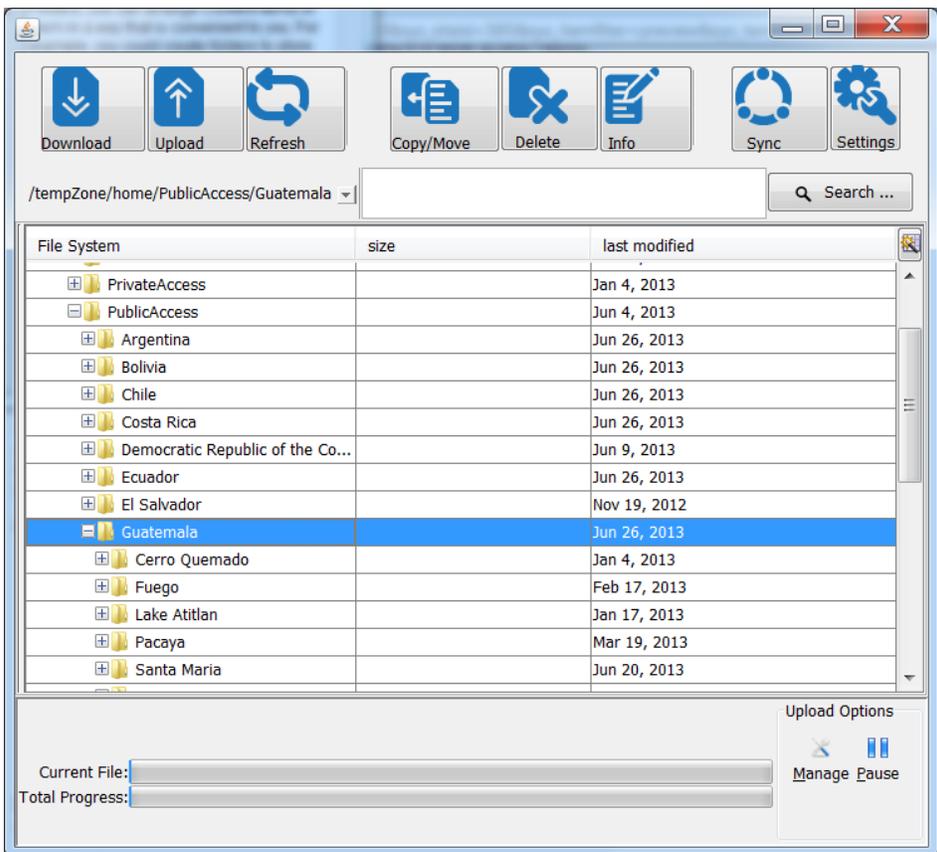


Figure 2. HAZHub iDROP user interface to Michigan Tech's Natural Hazards Archive.

Works best from the Mozilla Firefox web browser.

Link to idrop.jnlp: <http://iren-web.renci.org/idrop-release/idrop.jnlp>

Allow the iDrop Java program to run

Note: (running from other browsers will be checked into further. Currently, Firefox will run idrop.jnlp, but other browsers download the idrop.jnlp file and then are missing support files)

Please log in to your iDrop data grid

Host: hazard-data.geo.mtu.edu

Zone: hazhub

User Name:

Password:

Login Mode: Standard

Check the 'Advanced Login Settings' to change your password or to verify that the Default Resource is set to 'demoResc' (necessary if uploading data files).

HAZHub iDROP Features

- a. The directory structure is similar to the Mac environment
- b. Multiple files may be uploaded or downloaded at once
- c. Searches can be done on file names, but the metadata searches are still under development
- d. Text searches are case sensitive
- e. Clicking on a file in the Search results box, shows the selected file in the directory structure
- f. Files can be renamed if needed
- g. Metadata can be Copied to the clipboard to paste in other applications (Select all metadata values, or use CTL-click to select specific items)
- h. Metadata can be added and deleted, but values cannot be modified.

3.3.2. HAZHub Online

From any web browser, enter the url: hazard-data.geo.mtu.edu

Sign on to iRODS

Host/IP: localhost

Port: 1247

Username:

Password:

Zone: hazhub

HAZHub Online Features:

- a. The directory structure is similar to the Windows environment
- b. Only one file at a time may be uploaded or downloaded
- c. Searches can be done on file names and on metadata values such as Keywords and Latitude/Longitude values.
- d. Text searches are case sensitive
- e. Files cannot be renamed. (use HAZHub iDROP for this)
- f. Metadata can only be copied/pasted one value at a time
- g. Metadata can be added, deleted, and modified
- h. Includes a Map feature for locating datasets and searching by Latitude and Longitude

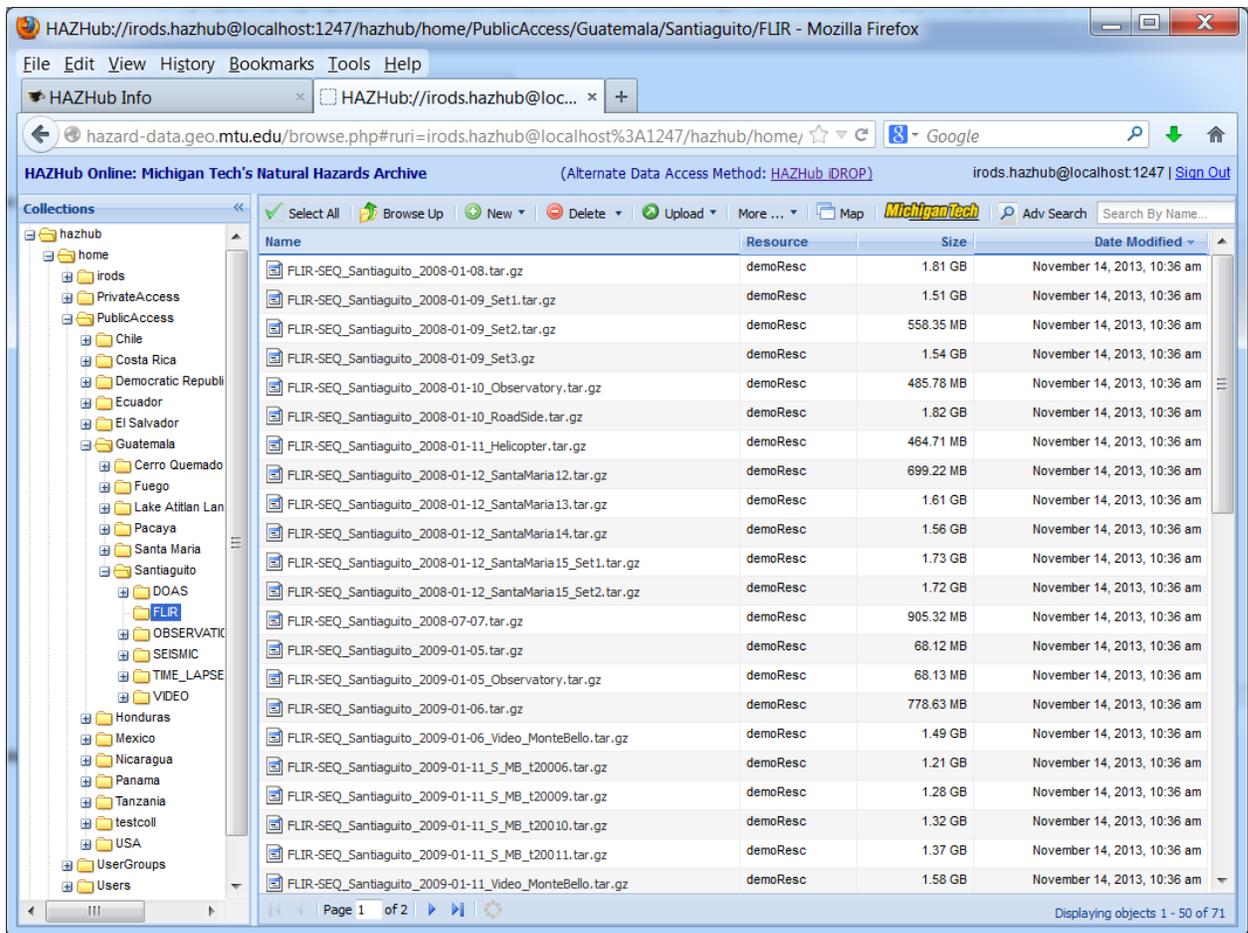


Figure 3. HAZHub Online interface to Michigan Tech's Natural Hazards Archive.

3.4. HAZHub Data Content

The Natural Hazards Archive is a collection of all types of data collected, generated, and processed by faculty and students in support of their research in the fields of volcanology, seismology, and hydrology. The data includes, but is not limited to, remote sensing, GIS, thermal infrared imaging, and seismic data, along with video and photograph recordings of events.

1. Basic data or raw information used for a project, which otherwise is not easily accessible, and which we have permission to put in the repository either for the general public to have access to it, or for only a restricted group of people. Proprietary data will only be accessible by those who have permissions to use it, such as ASTER data downloaded under Michigan Tech's account. Michigan Tech does not have permissions to redistribute ASTER data to the general public, but it is available for use by Michigan Tech researchers. Imagery data such as Landsat and MODIS will not be included in the repository since it is available for free on the Internet. Typical datasets uploaded to the repository include:

- a. Satellite remote sensing datasets in GIS raster format (e. g. geotiff, .img, etc. files).
 - b. GIS datasets acquired for the research purpose (e. g. shapefiles, raster datasets, etc.).
 - c. Instrumental and measurement datasets directly collected during field work, including geophysical (e. g. seismic, acoustic, thermal (FLIR), VLF, 2D_ER, differential GPS, etc), geochemical (volcanic SO₂, water chemistry, etc.), structural rock-mechanics (Schmidt hammer, joint orientation, etc.) and geo-hydrological (e. g. flow rate, well level, etc.) datasets.
 - d. Visual (photos and video), and traditional field (notebook, hand-held GPS, etc.) recorded datasets.
 - e. Analytical and laboratory results from the analysis of samples collected in the field, including geochemical (e. g. rock composition, thin section photographs, etc.).
 - f. Interviews, surveys, and other human subject research datasets, restricted by Institutional Review Board (IRB) constraints).
2. Diverse products generated from the different research projects (e. g. thesis). This should include information in a variety of formats. The most condensed level will be the documents (thesis and papers) describing and presenting the research results, but a variety of other datasets are also expected. Such datasets include:
- a. Datasets and tables in spreadsheet or other general purpose database format (e. g. excel, Dbase, etc.).
 - b. GIS files (raster files, shapefiles, .mxd ArcGis document files) and associated metadata files, including final maps and GIS layers.
 - c. Datasets in specific electronic file formats associated to each discipline (e. g. outputs from seismic source modeling, corrected SO₂ maps/images from the UV camera, etc.).
 - d. Computer code and programming outputs (e. g. Matlab SO₂ software analysis, other code written for specific data processing routines, etc.).
 - e. Documents published in conferences and congresses, including abstracts, posters, presentations, etc.
 - f. Papers published in peer reviewed journals, congress memoirs, etc.
 - g. Other presentations given at workshops, internal (GMES) activities, etc.
 - h. Thesis and report documents.

3.5. Directory (Collection) Structure and Naming Conventions

Directories in the repository are referred to as Collections. The archived data has been divided into 2 main collections, "PublicAccess" and "PrivateAccess". The PrivateAccess collection contains proprietary data with use restricted to Michigan Tech researchers and collaborators, under the guidelines of the data's use policies. The PublicAccess collection will be made available to all archive users, with the request that data be sited correctly in any published documents and presentations.

Directory structure for the PublicAccess and PrivateAccess directories:

/hazhub/home/PublicAccess/
 /Country/
 /Geologic Feature/
 /Data type/
 File names

example: FLIR data from the Sanitaguito volcano in Guatemala will be found under:

/hazhub/home/PublicAccess/Guatemala/Santiaguito/FLIR/

File naming conventions

The file naming conventions were planned to assist users in searches of the data, either by using the search features or just by browsing the collections of data. Many data sets are a combination of multiple files and directories that have been tar'd and zipped. To keep individual file sizes between 1 and 2 GB, some data sets consist of multiple tar.gz files that end with a set number to indicate multiple files for the same data set.

1. *Basic filename format:* TTTTTT_Volcano_yyyy-mm-dd_description_Set#.tar.gz
 - a. **Example:** FLIR-SEQ_Pacaya_2009-01_BuenaVista_Set01.tar.gz
 - b. TTTTTT: general type of data (FLIR, FLIR-SEQ, DOAS, ASTER, TIME-LAPSE, etc.)
 - c. Geologic_Hazard: name of volcano or other specific geologic hazard location if data is not volcano-related
 - d. yyyy-mm-dd: date of data, drop dd and mm if unknown
 - e. description: extra description to help identify data
 - f. Set#: if data is a set of related files, they will be tar'd and zipped in groups of files 1-2 GBs in size. Some data sets will be a set of tar.gz files.
If there are 10 or more sets, they should be numbered with 2 digits for directory sorting. Drop the Set portion of the name if there is only one file.
2. *Filename for thesis:* ThesisMSorPhD_Volcano_PublYear_LastName.pdf
ThesisMS_Santa_Ana_2008_Colvin.pdf
3. *Filename for data directly supporting a thesis:*
ThesisMSorPhD_Volcano_PublYear_LastName_DataType_DateOfData_Set1.tar.gz
ThesisMS_Santa_Ana_2008_Colvin_modis_aqua_2005_Set1.tar.gz
4. *Other Filename conventions:*
Do not leave blank spaces in file names. use "-" inside dates. Use "_" for other items, even volcano names, Santa_Ana

3.6. Metadata

Metadata will be stored for each file in the archive system (Figures 4 and 5). Users will be able to find data by searching on the metadata fields, and will also be able to get a quick review of the data file contents before downloading. Metadata fields will contain:

1. **Contributor Name:** Person contributing the data to the archive. This is usually the faculty member responsible for the data.
2. **Contributor Organization**
3. **Country:** country of data location
4. **Data Use Permissions:** description of how data can be used and distributed
5. **DataID:** Identification number from the database. This will identify individual files in the repository, and will also be included in the Reference Citation field.
6. **Description:** only populated when more detailed information than the keywords has been given
7. **Field of Study:** Volcanology, Hydrology, etc.
8. **Geologic Feature Name:** If the data pertains to a volcano, this will be the volcano name. Otherwise it will be the name of the feature or location the data pertains to, such as the Quito Aquifer System.
9. **Interest Level:** (Technical, General Public). If a data set is labeled “General Public”, it will contain photos or video footage that would be of interest to the general population as well as researchers. If it is labeled “Technical”, then it is a set of data, mostly of interest to researchers.
10. **Keywords:** list of keywords to assist in searching, includes, location, geologic feature, type of data, etc.
11. **Latitude (-n.nnn):** decimal format, positive values for N, negative values for S
12. **Longitude (-n.nnn):** decimal format, positive values for E, negative values for W
13. **Reference Citation:** citation to be used when referencing this data in publications
14. **zLatitude (x1000):** Latitude times 1000
15. **zLongitude (x1000):** Longitude times 1000

These metadata values will be automatically created/updated by a script. They were added to the system because of a database issue discovered in iRODS where numeric comparisons of decimal metadata gave inaccurate results. With this fix, searches will be accurate for Lat/Long values with precision to 3 digits.

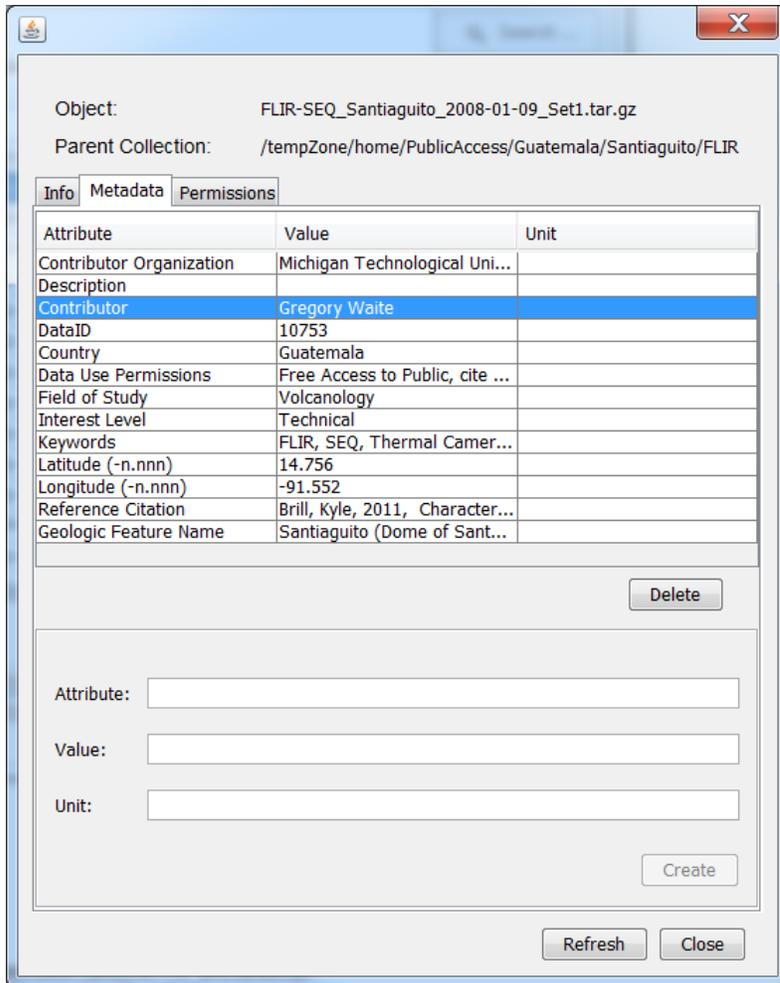


Figure 4. Click the “Info” icon in iDROP to display file information, and then select the Metadata tab. The whole block of metadata can be copied and pasted to other documents. The Reference Citation is especially useful when users need to cite sources in their publications.

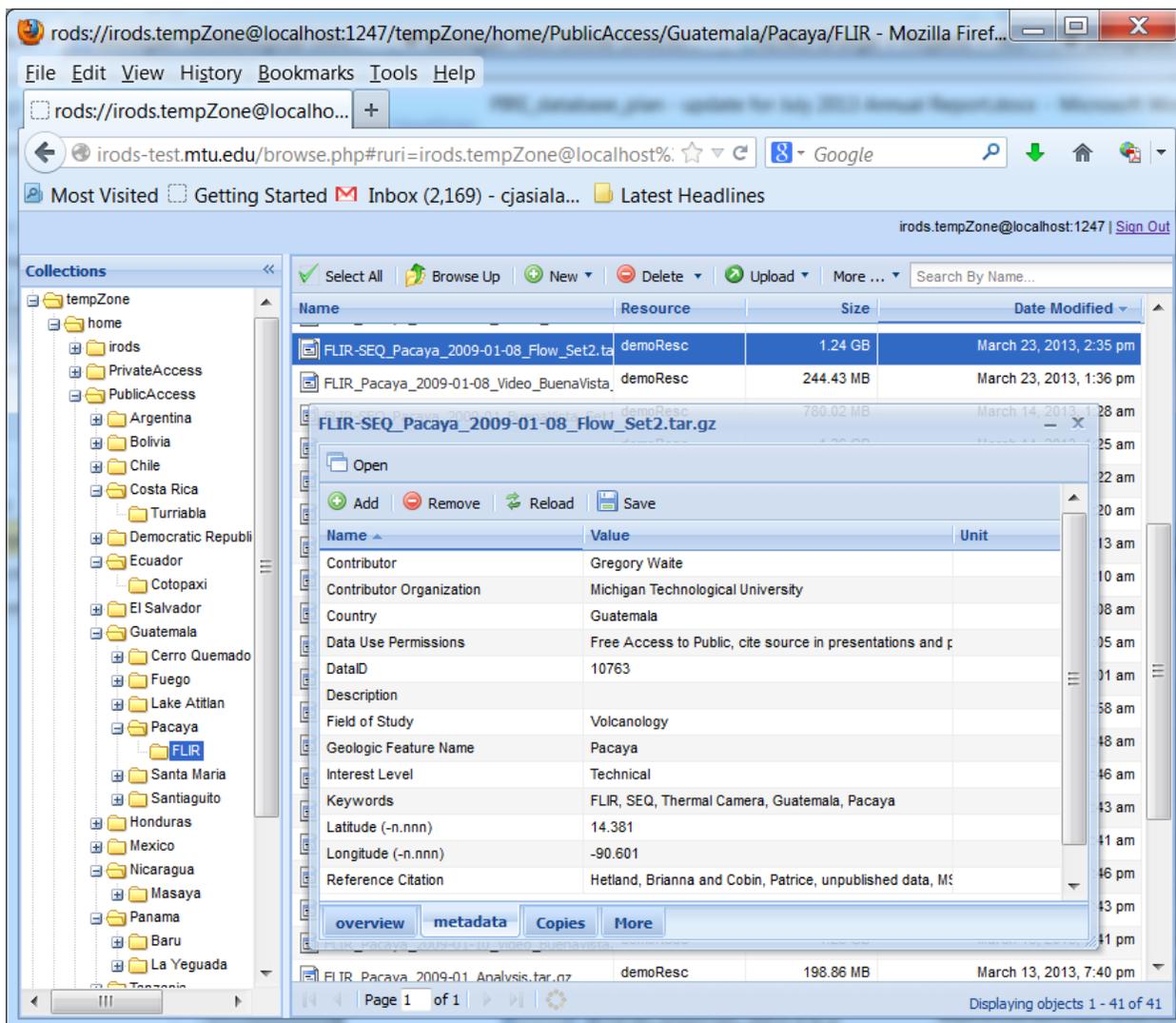


Figure 5. In the HAZHub Online browser, view metadata by double clicking on a file name, and selecting the metadata tab.

3.7. HAZHub Search tips

Searches by File Name and Text Metadata Fields:

- The current text searches are case sensitive.
- Both iDROP (Figure 6) and the iRODS web browser will look for the search text in any part of the File Name field. Searching for 'Pacaya' will find every file with 'Pacaya' in the File Name.
- Since file names are constructed with no spaces, search for 'Santa Ana' as 'Santa_Ana'.
- When searching by metadata values, the user chooses the operator (=, <, <=, >, >=, like). When the 'like' operator is used, the search text is found anywhere within the metadata value field. When the compare operators (=, <, <=, >, >=) are used, the comparison is performed with the first characters of the search text and the metadata value.

Example:

Keywords like FLIR will find all files with 'FLIR' listed anywhere in the Keywords field

Keywords = FLIR will only find the files where 'FLIR' is the only value listed in

Keywords

Keywords > FLIR will only find files where the 1st Keyword follows 'FLIR' in the alphabet

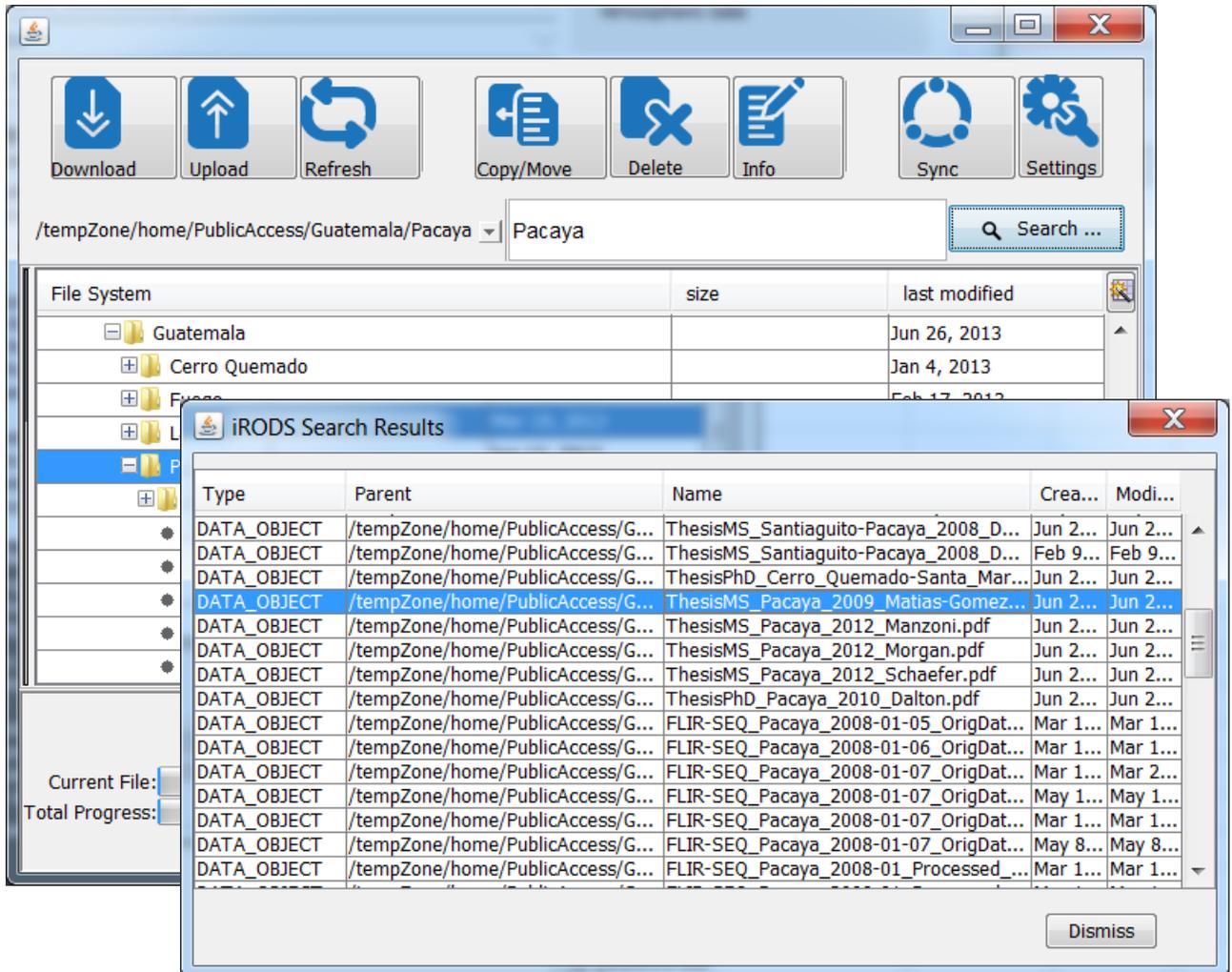


Figure 6. Search results in iDROPS when searching for part of a file name.

Search by Latitude/Longitude – Advanced Search:

- a. Latitude and Longitude are metadata fields, and can only be searched in the **HAZHub Online** interface (hazard-data.geo.mtu.edu, Figures 7 and 8). The metadata search feature in iDROP is currently under development.

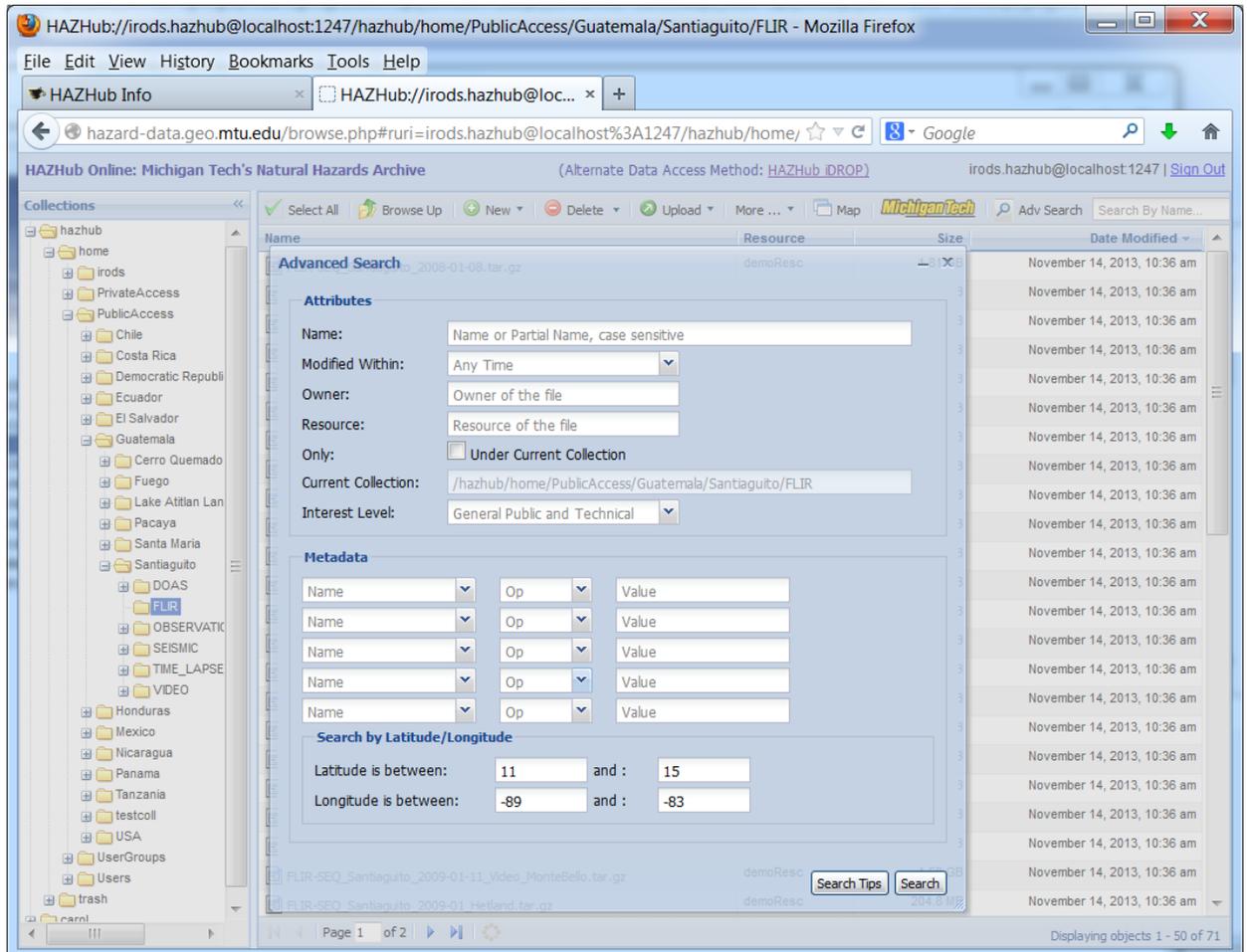


Figure 7. Search dialog for iRODS web browser interface. Open the search box by clicking on the dropdown arrow next to the search icon on the upper right corner of the screen.

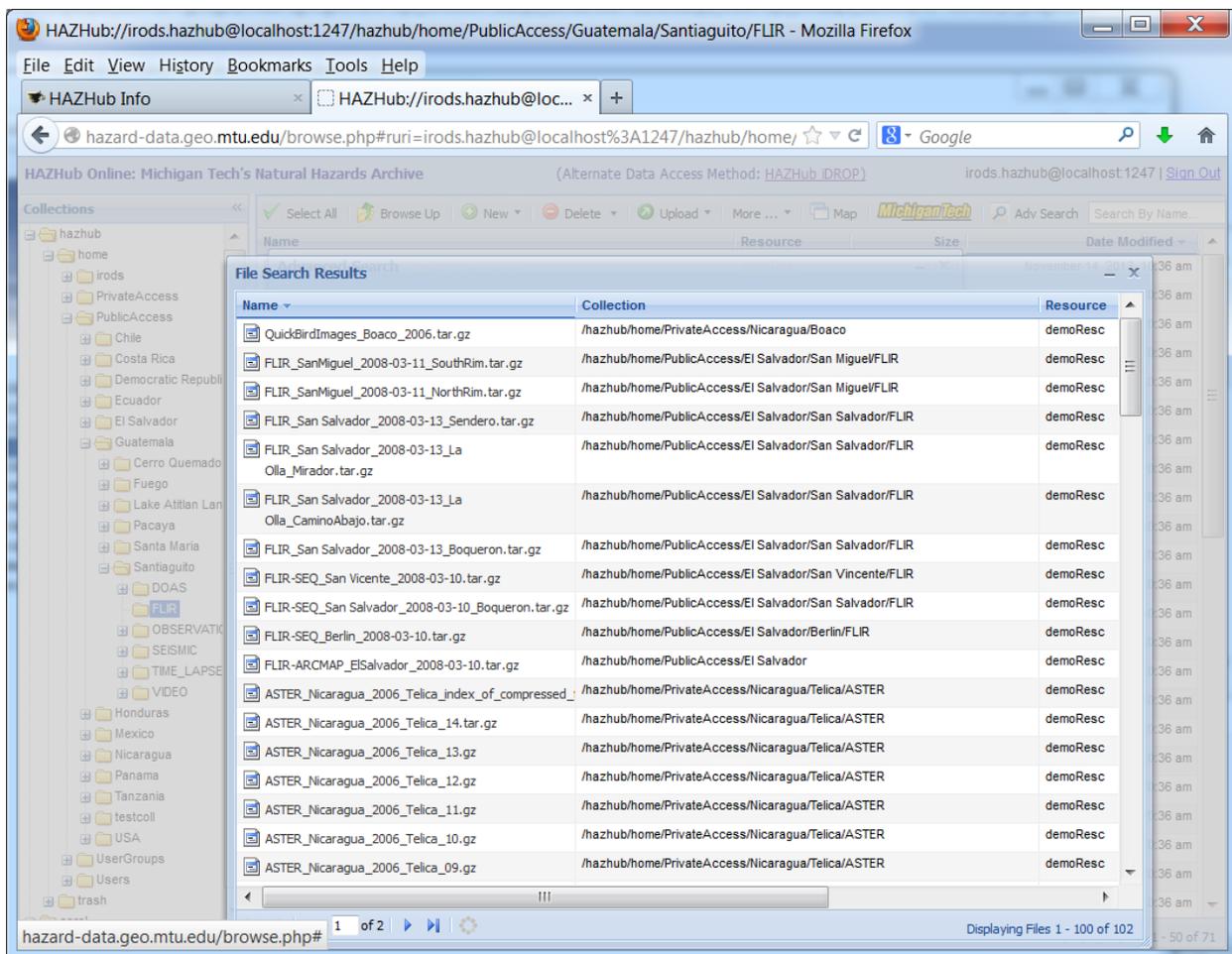


Figure 8. Search results list for the HAZHub Online interface. Each item in the list can be double-clicked to download or view the metadata.

Search by Latitude/Longitude – chosen on location map:

- From HAZHub Online, click on toolbar icon, **Map**
- A new window, HAZHub Online – Location/Search Map opens with all Latitude/Longitude data locations marked on the map.
- A double-click on an icon will show the feature name and country of the data.
- Select the "Draw a Rectangle" box to draw a search area.
- Click on "Search". This opens the "Advanced Search" dialog in the HAZHub Online web page. The lat/long search range will be populated with the SE and NW corners of the user-drawn rectangle.
- Click Search to view the list of data files within the lat/long range.
- To limit the search to just the /PublicAccess/ Files, select the /hazhub/home/PublicAccess/ directory, and then do the search. Check the "Under Current Collection" checkbox in the Advance Search dialog box before clicking on Search.

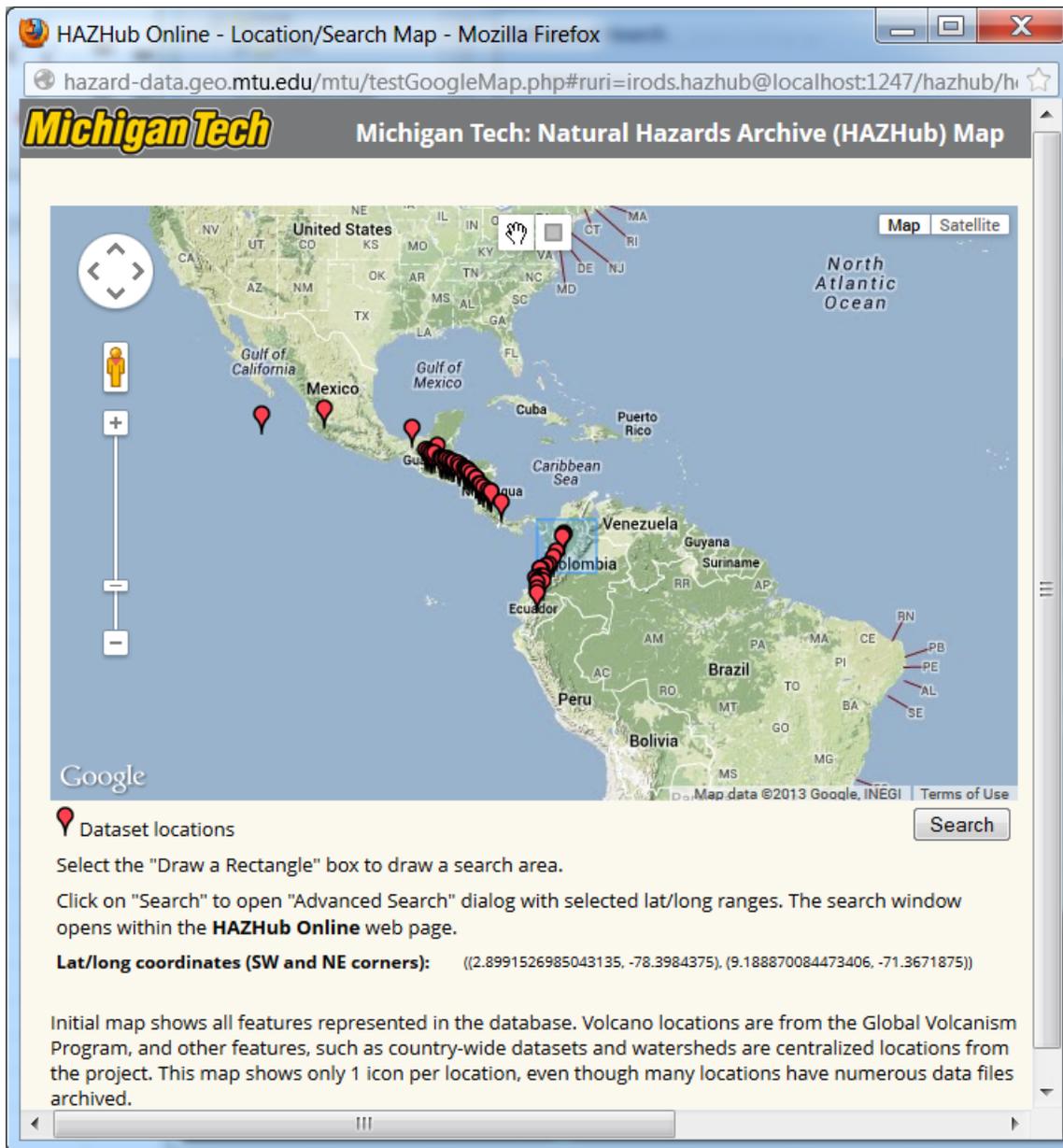


Figure 9. HAZHub Online location map. Users can select an area to search for files within the rectangle boundaries.

3.8. Uploading and Downloading Files

When using iDROP, multiple files can be uploaded or downloaded in one operation. To download, the user selects the files to download, and clicks the “Download” icon. The user selects the destination directory in his local drive from the dialog box (Figure 9).

In the HAZHub Online web browser, download individual files by double-clicking the file, and selecting “Force Download”. Files can be uploaded with the “Upload” tab.

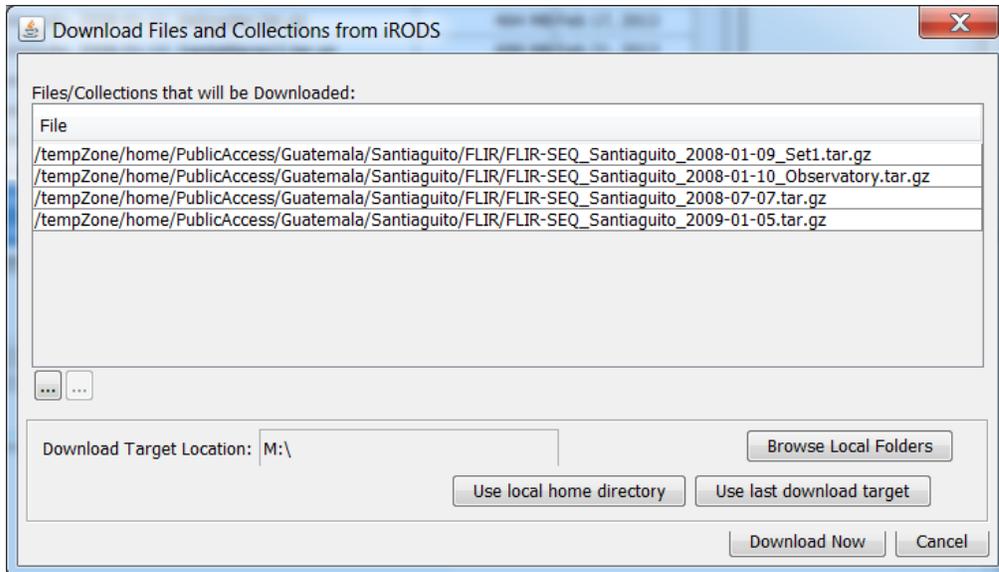


Figure 10. iDROPS interface for downloading files to a local directory. The files listed here are the files that were selected by the user in the data archive.

3.9. Delete Files

Users can only delete if they have read and write permissions.

When deleting files in iDROPS, they automatically get moved to the Trash folder. Even if the file being deleted is in Trash, a new Trash folder is created under the current folder and the file is moved there.

When Deleting files in the HAZHub Online web browser, the user has the option of either permanently deleting the file, or moving it to Trash. Under the “**Delete**” menu, if “**Force/Permenant Delete**” is checked, the deleted file will not be saved under Trash. If this is unchecked, any deleted file will be moved to the Trash folder.

4. Criteria to include data in HAZHub:

The purpose of the data repository is to archive valuable and usable project data created within the PIRE project. For files to be considered valuable and usable the students will apply the following criteria to uploaded data:

1. The files have to be a finished product of some sorts, such that someone interested in the topic covered in the study will find it useful. Intermediate products (e. g. GIS layers used to clip data), unsuccessful trial runs, and "dead end street" tests (e. g. calculations that were not finished or lead to non-conclusive results) should not be included.
2. The files have to be either self-explanatory (e. g. self-contained power point presentations or text documents for a technical but somewhat general audience), or should have some kind of metadata to provide further explanation. This could include an index of content for a set of

files in specific folders, self-explanatory file names, standard file format and names for known file products (e. g. satellite imagery), further documentation in the Keywords and/or Description metadata fields, etc. Files which have some direct relation to a thesis, paper or other document that we can either include in the database (e. g. thesis) or formally cite (e. g. paper) could be included, as long as it can be clearly linked to such a document and the content of the file is clearly described.

3. Redundant information should be avoided; only the final and (hopefully) definitive version should be included in cases where there are several versions of a document or dataset, corresponding to different stages in the development of the document or dataset.
4. Data will be classified as available for use and distribution by the general public, or available only to Michigan Tech researchers. All project results were intended to be made available to the general public, but some supporting data associated with a project may include purchased and copyrighted material for which Michigan Tech does not have permissions to make available in a public database. Examples of this type of data are:
 - a. ASTER imagery which is free, but can only be distributed within the organization that has an authorized download account
 - b. Purchased data sets that cannot be shared
 - c. Photographs and videos taken by organizations that have shared these images solely for Michigan Tech research.

5. Procedures for Contributing Data to HAZHub

Contributions to the Natural Hazards Archive are encouraged, especially from current and former grad students involved in the PIRE project, "Remote Sensing for Hazard Mitigation and Resource Protection in Pacific Latin America".

1. **Data Review**

Check with Carol Asiala (cjasiala@mtu.edu) or John Gierke (jsgierke@mtu.edu) before uploading data. We would like to know what type of data is being added to the archive and make sure contributors follow the archive guidelines we have set up.

2. **Upload to your user directory**

Once you have received a userid and password, a directory under /hazhub/home/Users/<your userid>/ will be set up for you. This is the only place you will have permissions to upload data to the archive.

3. **Combining datasets**

Please combine files that belong together using tar and zip. (ex. files & directories needed for an ArcInfo Map project)

4. **Keep file sizes under 2 GB**

Files over 2 GB may be difficult for some users to download. If a combined dataset will be over 2 GB, then create several files with the same naming convention, but add "Set01", "Set02", etc. to the end of the file names. Users will know the data files belong together, and will download them together.

5. Project Files, layers, and links

If you are uploading a project that contains links to layer files, please include all of the files needed to open and view your project.

6. Metadata

When you upload a file in your Users directory, a set of blank metadata will be generated. Please fill in this data (access through [HAZHub Online](#)), including Latitude and Longitude values. If your data is focused on a particular volcano, use the lat/long coordinates from the [Smithsonian Institution Global Volcanism Program database](#).

7. Private or Public Access

Please indicate the type of data you are submitting. If it can be freely accessible by the general public, it will be moved to /hazhub/home/PublicAccess/. If the data is proprietary and should only be used by Michigan Tech Researchers, then we will move it to /hazhub/home/PrivateAccess/.

8. Notify **Carol Asiala** (cjasiala@mtu.edu) when your upload is complete, and after reviewing, we will move it to the correct directory.

6. Developments still in progress

1. Currently in the process of connecting the iRODS database at the University at Buffalo with Michigan Tech's iRODS database. The iRODS Federation feature will allow Michigan Tech and the University at Buffalo to administrate their own iRODS archive systems, but will allow users of both systems to access the data on both systems. This feature currently being tested and coordinated with the University at Buffalo.
2. Assess the iDROP Transfer Manager to check its usefulness in monitoring the upload/download activity of the data repository through iDROP.
3. Continue uploading project data, and proofread the uploaded files and directories to make sure they follow the outlined naming conventions and metadata structures.