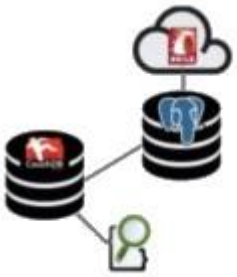


Hatch Tutorial

Last Update 1/12/2015

1 Introduction

This tutorial will give a brief introduction to the primary features of Hatch. Hatch is an online database application currently designed to make the storing and sharing of biological research data easier.

ALERT TO USERS!!!
Hatch is still under development. While the developers try to keep things up to date, features and images given in this tutorial may vary from the actual program.

2 Understanding Documents, Collections, and Metadata

Hatch has been designed to simplify data management by combining standard file organization concepts on top of more advanced database models. A Hatch document can be thought of as just a file containing your data. This could be data from a spreadsheet or it could be a table exported from another database. Documents also may contain metadata (e.g. date, locations, etc.), online and/or downloadable notes.

A collection is like a folder/directory and contains one or more documents. Document actions can be performed on all documents in a collection.

Hatch also has the ability to **validate** data by applying **filters** to documents to insure they are compliant with organizational data standards or to be compatible with database schemas. Depending on the format of your data you may add **metadata** either by embedding it in a document and using an i-filter, by applying a "**meta-form**" to a document/collection, or by using a document's metadata editor. *Metadata is a very important goal of Hatch and is critical for adding context about a user's data*.

3 Becoming a Hatchling

Becoming a Hatch user is simple;

Figure 1. Hatch main menu bar.

1. Go to the Hatch main webpage http://www.datahatch.org/
2. Click the "**Register**" button on the main menu bar.
3. Enter an email to be used as your user/login name.
4. Enter and confirm your password.
5. You should receive a confirmation email shortly; follow the instructions in the email.

Once registered, you can start uploading data.

4 Uploading Data

To start uploading data log into Hatch and click on "**Data**" in the main menu, this brings you to the collections screen. From the collections screen click the "**Upload Files**" button, this brings up the "**Data Import**" screen (see figure below). From the data import screen you can select an existing collection from the drop down menu to add documents to or automatically create a new collection and add documents to it by typing in a collection name in the text field.

Under "what kind of data is this" select a **filter** that fits your data from the filter drop down menu. The selected filter will be applied to each document you add to the collection. Most filters parse and **validate** your data to make sure it is compliant with organizational standards; they also make sure your data is in a nice table format when viewing it within Hatch.
NOTE: depending on the size of your data and how many other documents are being validated it may take some time to finish validating your data.

If you have data that can't or should not be filtered you must still select one of the appropriate items from the list of filters. There are some file types, such as GIS model files, PDF note files, images, etc. which use special encoding that can't be filtered but are still an important part of a user's collection of data. For files like this use one of the "**No-filter**" options from the filter drop down list. Files uploaded with the "no-filter data" option will have a record in Hatch the same as regular filtered documents. Users can add metadata and notes to these non-filterable documents in Hatch in the same way as filtered documents. Since encoded documents must be decoded to be useful, when a user "views" the document they will be presented with a link to download the file and can then open it with the appropriate program. For more information about managing non-filterable files see the "uploads" section.

Warning! Do not try to filter an "encoded" data file (GIS maps, PDF, etc.) this will corrupt the file. When uploading, use one of the "no-filter" options.

Caution! Data validation is not an upload requirement as of the writing of this tutorial however it is *strongly encouraged* that you **ALWAYS** validate/filter your data either on upload or immediately after upload. Data that is not validated may not be compliant with organization/government requirements. It also will be difficult to view in Hatch and may not be searchable.

You can add files one at a time by clicking the "**Add files**" button on the data import menu or you can **drag and drop** files from an open folder. If you wish you can apply a "**meta-form**", which will add a pre prepared set of metadata to each file uploaded. Once you have all the files ready click the "**Start upload**" button. You should see a progress bar next to each file indicating how close it is to being uploaded. Once finished close the data import screen.

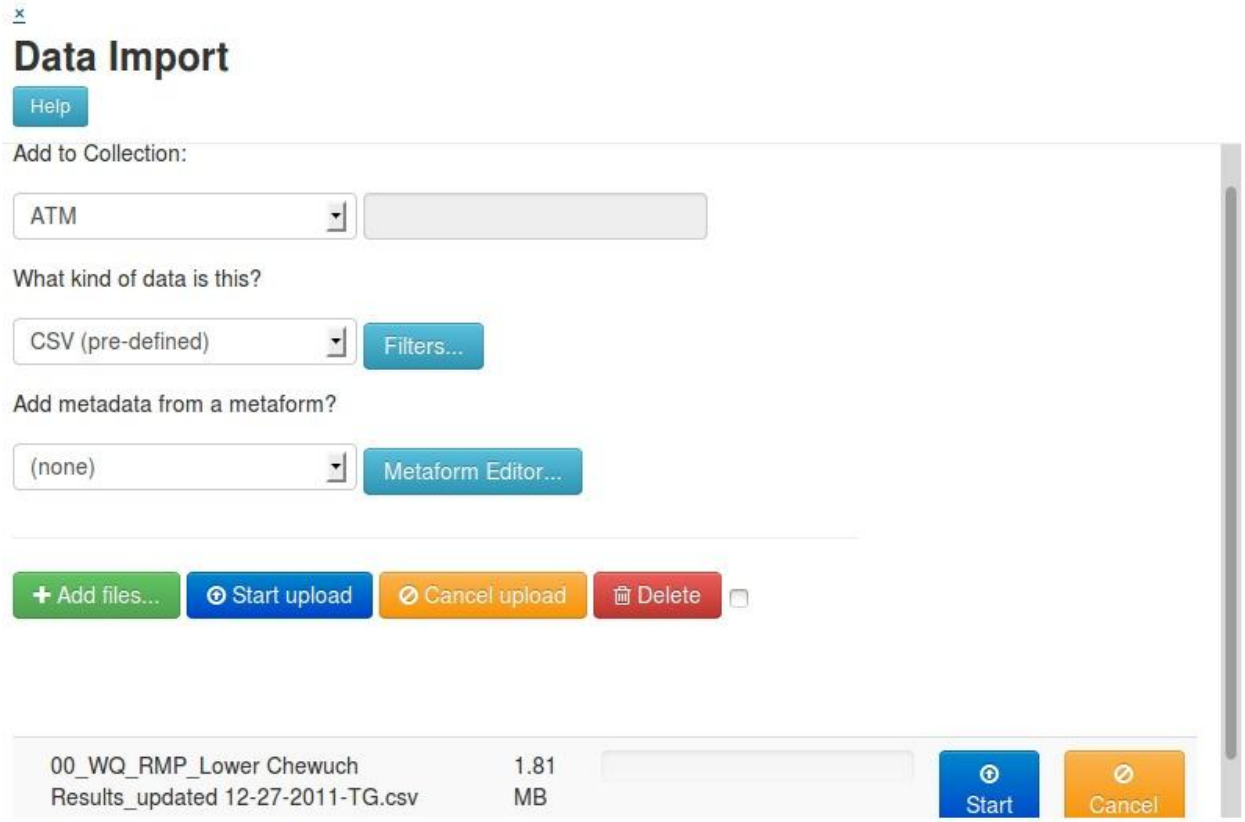


Figure 2. Data import screen.

Now you can see your collections and some of its documents from the collections screen.

5 Collections

Once you have data uploaded you will probably need to begin managing your collections, this can be done by clicking on "**Data**" in the main menu bar. From the collections screen you can create new empty collections or click on a collection to view all its documents. To expand a collection and view any sub-collections click on the "**+**" symbol. Clicking on "+" also shows the "**Action**" button, click the action button to see the collection menu containing commands such

as “**Edit**” and “**Destroy**”. The edit command will take you the collection edit page where you can perform actions such as changing the collection name, selecting a parent collection (making the current collection a sub-collection of the parent), apply a filter to all documents in the collection, adding a meta-form to each document in the collection, or adding the collection to a “**Project**”.

You can link notes to a collection, such as PDF files or images, by going to the collection “Edit” menu and selecting a file from the “Add file as a note” dropdown list. Collection notes can be viewed by clicking a collection name, then clicking the “**Show Notes**” button. See the “Uploads” section for more information.

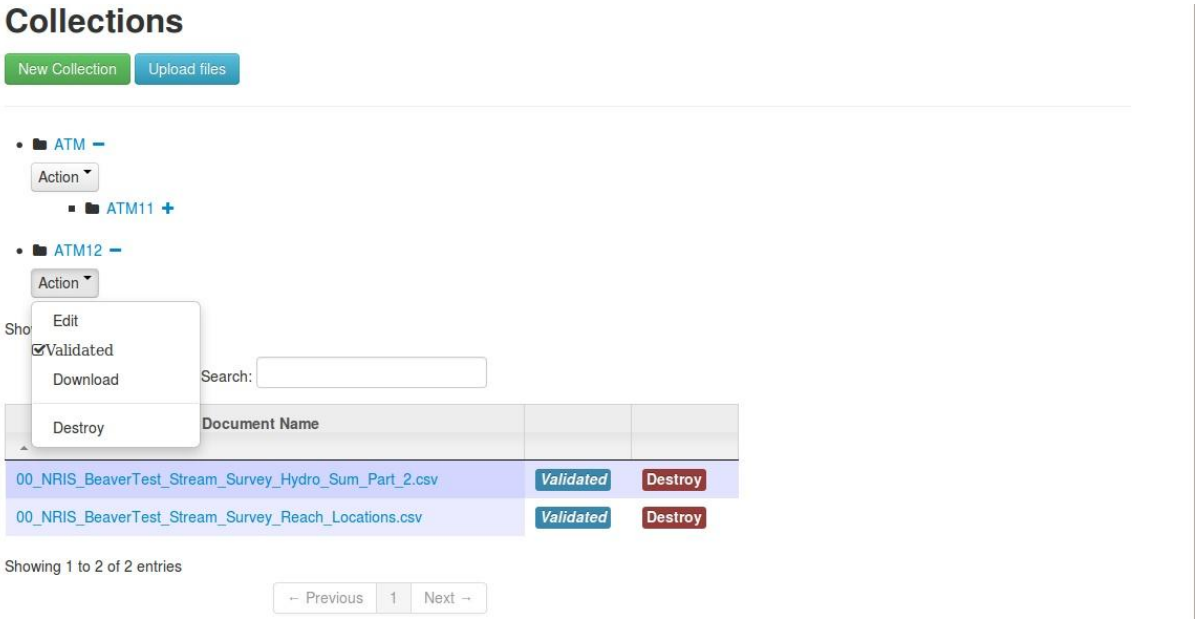


Figure 3. Collections screen.

As with most pages in Hatch there is more than one way to access pages of a collection, such as the edit page.

6 Documents

Once files have been uploaded into a collection you can view them by going to the collection, seeing a list of its documents, and clicking on the document’s name, this will take you document information page. The information page contains metadata about the document, a notes section, and a link to view the actual data. Clicking the “**Download**” button (after the collection name) will download all the documents in the collection. Owners of the document have the option to add/edit metadata, add metadata from a meta-form, and add/edit notes. Owners can access a document’s edit page by clicking the “**edit**” link from the document’s menu. From the edit menu you can do things such as rename the document, move the document to a different collection, or validate the document by using a filter (if it has not already been validated).

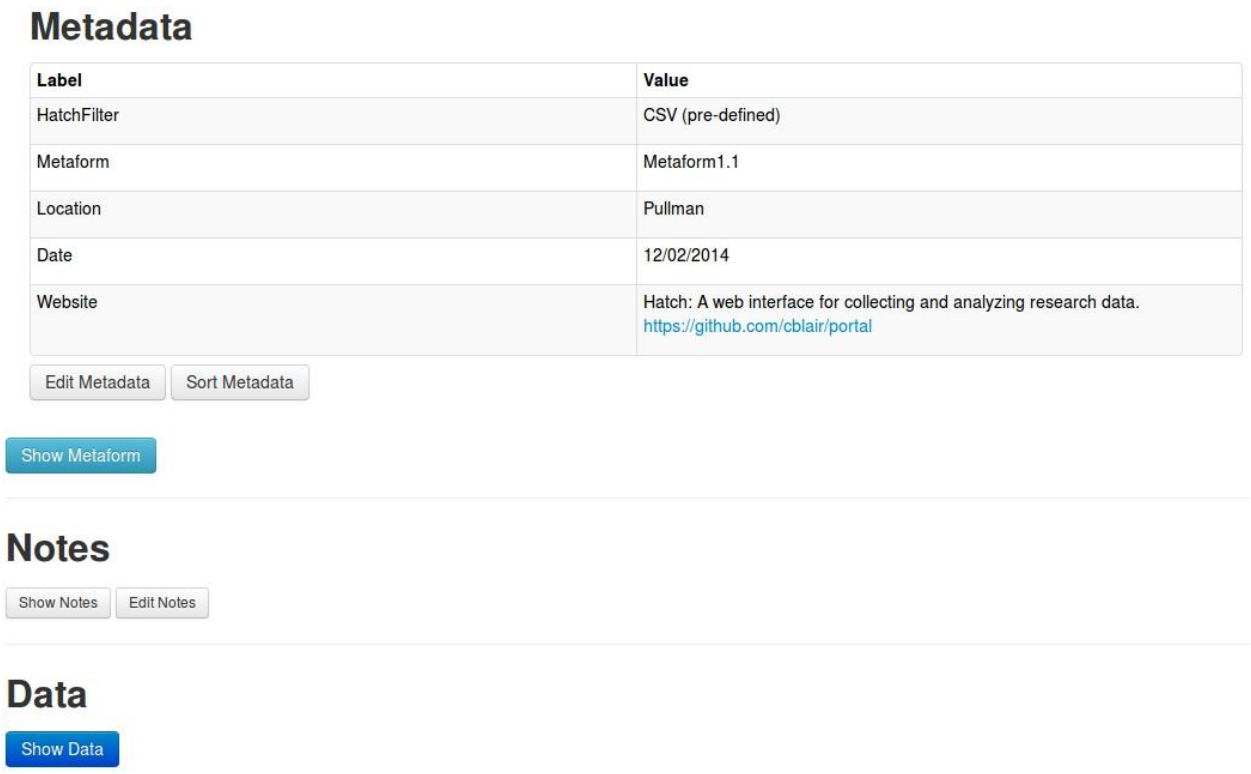


Figure 4. Document info page.

To view the data click the "**Show Data**" button, note that loading the data may take some time depending the document size, your connection speed, etc. If the data has been **validated** by a **filter** it will appear in a table format similar to a spreadsheet or a table in other databases. You can also change how many entries (rows) will appear per page. General information about the data can be seen at the top of the page including the document name, collection, and owner. More information can be seen by clicking the "**Properties**" button. To return to the document information page click the "**Hide Data**" button. To go to the collection page click the "**Back**" button.

Simple notes can be added by clicking the "**Edit Notes**" button. The note editor functions similar to Notepad or other similar text editors. Notes includes support for links to web sites and the link will automatically be generated when changes are saved in the note editor. More complex notes such as PDF files or images can be uploaded and linked to a document by going to the document "**Edit**" menu (not the note editor). These types of notes will appear in the "**Additional Notes**" section as links and can be downloaded. See the "Uploads" section for more information.

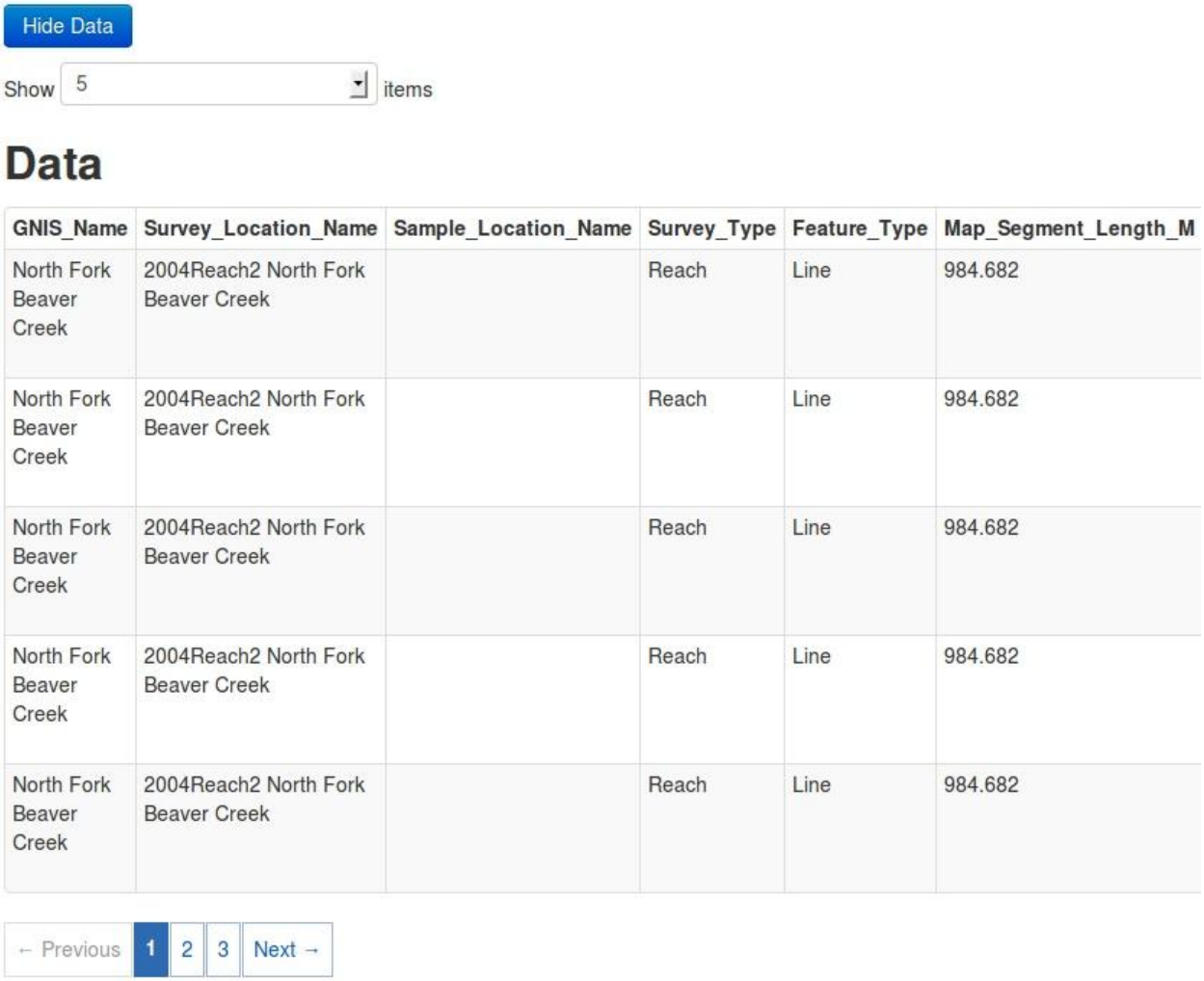


Figure 5. Document data page.

7 Uploads

Uploads provides users a way to manage non-data files, such as PDF and image files. Technically all files uploaded to Hatch are "uploads" regardless if they are plain text data files, encoded data files, note files etc. Hatch will treat uploads differently depending on the filter selected on upload. Generally speaking there are three types of uploads.

- Uploads that contain data and can be filtered.
- Uploads that contain data and cannot be filtered (encoded files).
- Uploads that should not be filtered (notes, PDF files, images, etc.)

Uploads that contain data and can be filtered should be treated as standard documents. Until a document is filtered it's just a raw file and the upload record will be maintained. Once a document has been successfully filtered the upload record will be deleted.

Uploads that contain data and cannot be filtered should also be treated as standard documents. Since the upload cannot be filtered the raw file and the upload record will remain until the upload is deleted.

Uploads that should not be filtered should not be treated as documents since they do not contain data. Uploads of this type are aimed at users who have complex notes, such as PDF files, they wish to link to a document.

Important! If a document has an associated upload, as in the case of a non-filterable data file, it is recommended you delete the document which will automatically delete the upload, metadata, notes, etc. Deleting an upload from the upload menu only deletes the upload; it does not delete the associated document or the document metadata, notes, etc. since this information may be important and should not be accidently deleted.
In the case of an upload with an upload type "note", deleting the upload will automatically remove all links made to the note in any collection or document.

More information about an upload can be seen by clicking the upload file name. From this page you can view/download the file by clicking the "**View file**" button.

8 Meta-forms

Situations may occur were a user needs to repeatedly add metadata to a document. One way to do this is through Meta-forms. Meta-forms provide a graphical interface for adding metadata. Meta-forms are based on the "**i-filter**" concept of **key/value** pairing. The "**key**" acts as a label for the value, for example "Survey Type", "Protocol Name", and "Date" might be keys for a meta-form. Meta-forms can be created by accessing the Meta-form creation page and used in a document from the add metadata menu.

8.1 Using Meta-forms

Adding metadata from a meta-form is easy. Simply open the document you wish to add metadata to and from the metadata page, under the metadata table, click the "**Show Metaform**" button. You will be presented with a dropdown list of available meta-forms, select the meta-form you wish to add and click "**Add Metaform**" button, you will then be directed to the "Metadata Input" page where you can change or fill in any additional information if needed. When you are finished, click the "**Update Metaform**" button, the meta-form metadata will now appear in the document's metadata table.

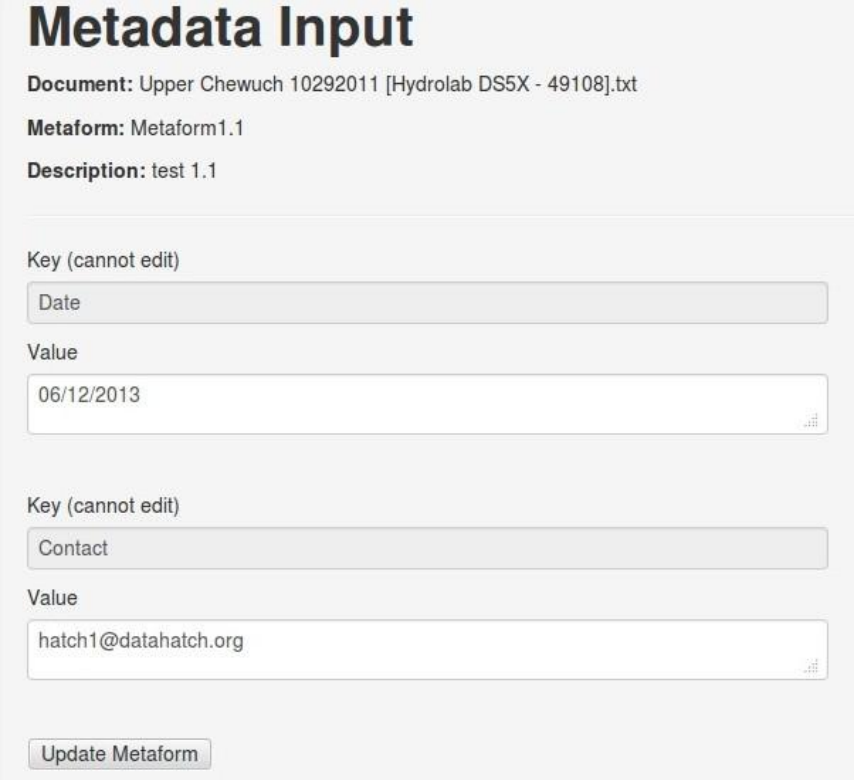


Figure 6. Meta-form metadata input page.

> **Meta-form Tip**
> You can add multiple meta-forms to a document. This modularity helps reduce repetition. For example; you could create a meta-form with just your contact information, preventing the need to repeatedly input your contact information in each meta-form.

8.2 Creating Meta-forms

The meta-form creator can be accessed from the main menu bar or from a document's "**Show Metaform**" link. To start a new meta-form, click the "**New Metaform**" link. You will be presented with a blank meta-form with fields for the meta-form name, description, and a key (label)/ value pairing where the actual metadata will be entered. Each key/value pair will be displayed as a single row in a document's metadata table. You can add more metadata by clicking the "**Add Metadata**" link, or remove metadata by clicking the "**Remove Metadata**" link. When adding metadata the key field MUST be filled in, values can be left blank, but not keys. Any metadata pair with a blank key will be discarded when the meta-form is saved. When you are finished adding or updating your meta-form, click the "**Update Metaform**" button to save changes.

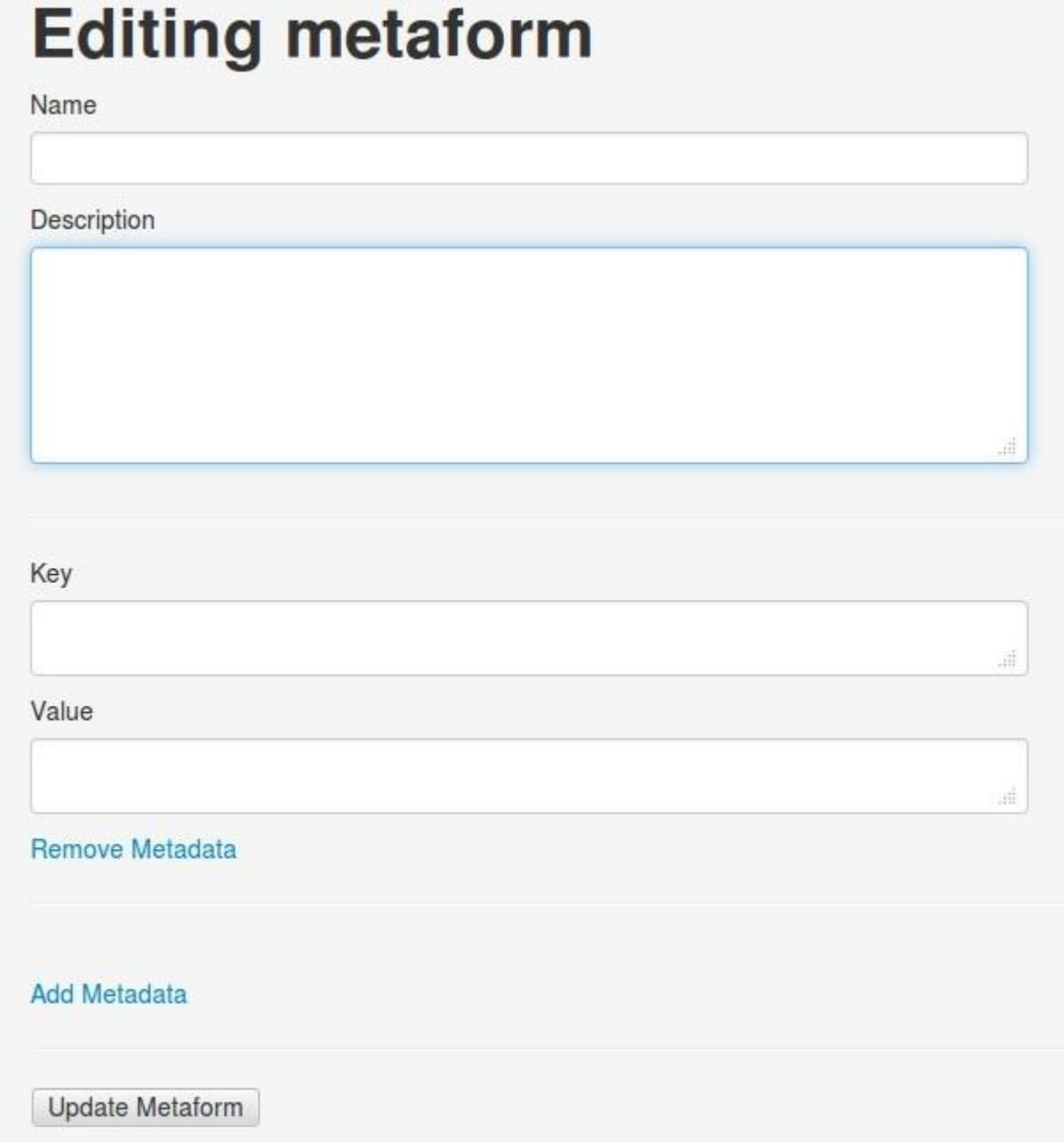


Figure 7. Meta-form editor page.

Once a meta-form has been created, you can view it by clicking the name of the meta-form from the main meta-form menu. When viewing a meta-form (not when editing) owners of a meta-form can sort the rows of metadata by dragging and dropping the rows where desired.

9 Projects

Projects assist with document/collection management and access control. Go to the main project page by clicking on "**Projects**" in the Hatch main menu bar. To create a project click the

"**New**" button to take you to the create project screen enter a name and a description (you can also add a collaborator) then click the "**Create project**" button. To view a project click on the project name from the main project menu, now you can see the collections belonging to the project. You can add collections to a project from the collections edit menu.

All data in Hatch is **private** and is only accessible by its **owner** by default. Data must be added to a project before it can be shared. To make a project public go to the project and click the "**Public**" button on the project menu bar. This will make the project and all the collections and documents in the project viewable by all users of Hatch. Users will have "**read only**" access to a public project, which means they can look at the data but can't access edit menus or any other menu that lets them change things.

9.1 Project User Types

Hatch currently supports three types of users; owners, collaborators, and editors.

- **Owners**: The user that uploads a document is the owner by default. Owners have full access to their documents. Owners can transfer ownership of a project and all of its collections and documents to another user by going to a project's edit menu, scrolling to the bottom of the page, selecting a target user, and clicking the "**Change Owner**" button.

- **Collaborators**: To allow specific users to view, and **only view**, a project go to the project's edit menu where it says "Add a collaborator to this project", select a user's email from the dropdown list, and then click "**Update project**". Collaborators have read only access to a project.

- **Editors**: To give a specific user a **high level** of access including the ability to delete, rename, filter documents, etc. Go to a project's edit menu and under "Add an editor to this project", select a user name from the drop down list and click the "**Update Project**" button. Warning: editors can manage a project in virtually all ways including deleting documents and transferring ownership of the project and all of its contents to another user.

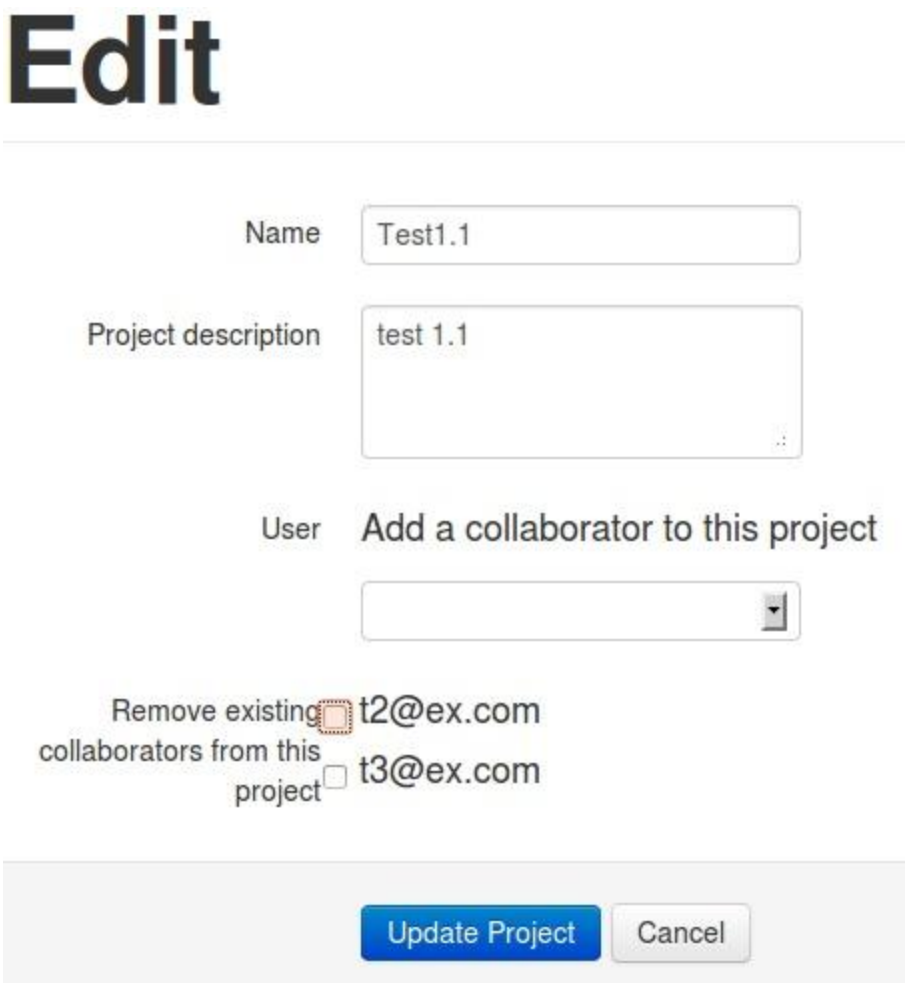


Figure 8. Project edit menu.

10 Search

Search provides users a way to search for documents using a standard web search type of interface. Search will look through metadata and data values to find documents. The main search engine works slightly different than other search fields in Hatch (which are mostly used for filtering) such as collations or jobs.

- Search is NOT case sensitive.
- Search supports the use of "wildcards" (click the "?" to the right of the search field for a detailed description).
- Search look for an *exact* word match unless wildcards are used. For example if you're looking for "Seattle" and typed "sea", search would fail, but "sea*" would succeed.
- Search searches for metadata values but not metadata labels.
- Search searches for data values and data column names.

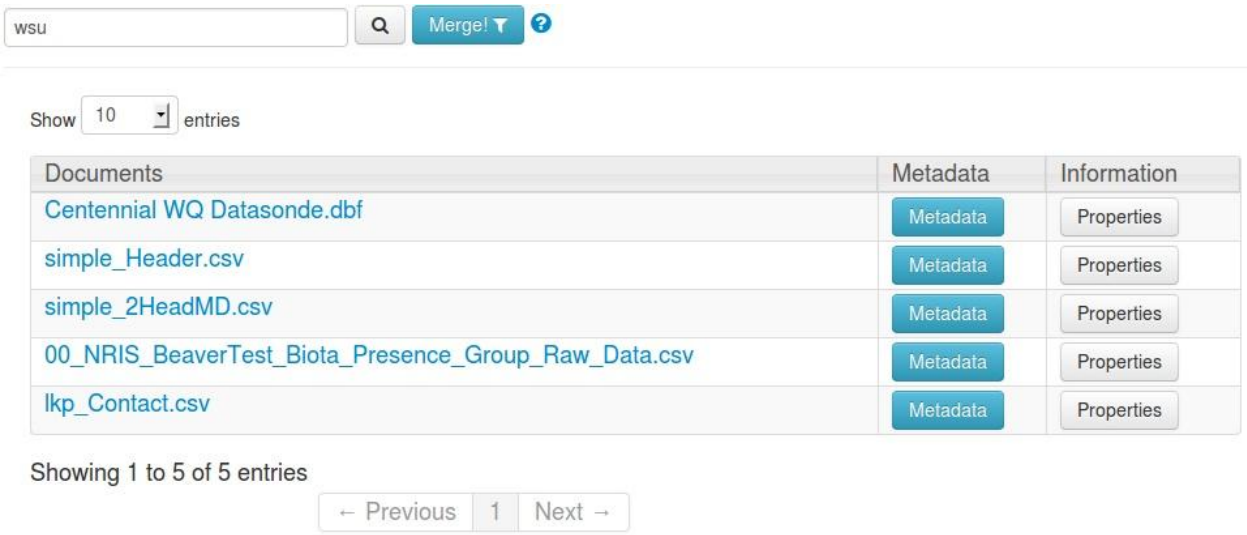


Figure 9. Search example.

10.1 Search Merge

Coming soon…

11 Jobs

Hatch uses "jobs" to manage to manage data validation. When a filter is run on an uploaded file a job is created and added to the job queue. A job status icon will appear in the menu area when you view a document. Validating a file can take time depending on the size of the upload, number of other files being validated, etc.

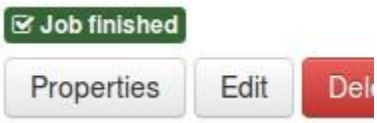


Figure 10. Document jobs icon.

To go to the main jobs page click the "**Jobs**" button from the main data collections page. You can view details of a specific job by clicking the job icon in the document menu or by searching for the document by name from the main jobs page. From the main jobs page you can see a list of all jobs running and finished.

To find specific jobs type part or all of the file name into the "**Search**" field (case sensitive) and hit the "enter" key.

To see (filter) only your jobs type your full user name into the "**Search**" field and hitting the "enter" key, for example "jenson102@wsu.edu" (without quotes).

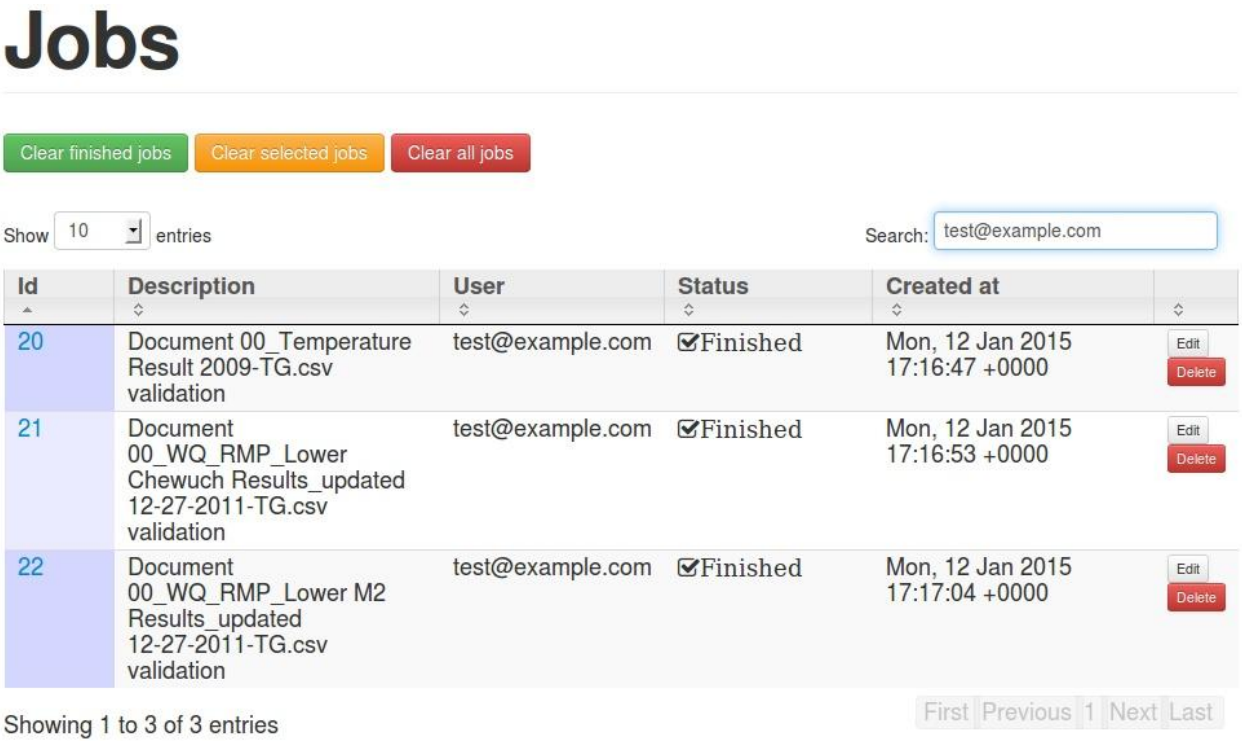


Figure 11. Jobs main menu.

Finished jobs can be removed from the job list by clicking the "**Clear finished jobs**" button. This will clear all your completed jobs both successful and non-successful.

To clear and remove all jobs including jobs that have not finished click the "**Clear all jobs**".

Use the "**Clear selected jobs**" button together with search/filter to clear specific jobs.

12 i-Filters (advanced users)

i-filters provide users the ability to customize the way documents are parsed including the parsing of metadata already in a document. Using i-filters is very simple, simply select a filter from the filter dropdown menu when uploading data or from the "**edit**" menu of a single document or a collection (to apply a filter to all documents in a collection). Several filters are already built into Hatch, such as the CSV filter.

i-filters make use of "**regular expressions**" to parse files. Regular expressions (regex) are beyond the scope of this tutorial however there are many good and free tutorials available on the internet.

Depending on the size of your data and how many other documents are being validated it may take some time to finish validating your data. Small files, less than 5MB, usually take no more than 1-2 minutes. Large files, 15MB-20MB, usually take between 10-20 minutes. Nothing should take more than an hour to filter.

12.1 How i-filters work

When an i-filter is applied Hatch will parse your data following the conditions specified in the filter. Filters have two parts; an optional "**Filter Header**" setting and a "**Data Filter**" setting. Hatch will start by parsing the header, which will go into the metadata section of a document in Hatch. Then Hatch will parse the data part of a document and place it into the data section of a document. Depending on how the i-filter was created irrelevant text can be filtered out and will not appear in Hatch.

The "filter header" setting requires a regular expression to be entered for *each line* of metadata you wish to parse since these lines are likely to be different from each other. If no header filter is given or all are removed then Hatch will assume there is no metadata and move to filtering data. The header filter expects metadata to be in label (key) value pairs. For example "Location: Methow", "Location" would be the label and "Methow" would be the value. Note the use of the colon and parentheses is for the convenience of the reader and is not necessary for the filter to work. Any character (or blank space) can be used to separate labels and values. If a file only contains a value (no label) the filer can generate a temporary label which should be changed using the metadata editor. Having values without labels is not recommended.

The "data filter" setting can use a regular expression or act as a "command line" for some special cases of filtering.

Warning! Do not try to filter an "encoded" data file (GIS maps, PDF, etc.) this will corrupt the file. When uploading, use the "no-filter data" option.

12.2 Special Case "Command Line" Settings

Hatch contains internal i-filter settings which cannot be created practically through the use of regular expressions. These settings are entered in the "data filter" field (not including the ":").

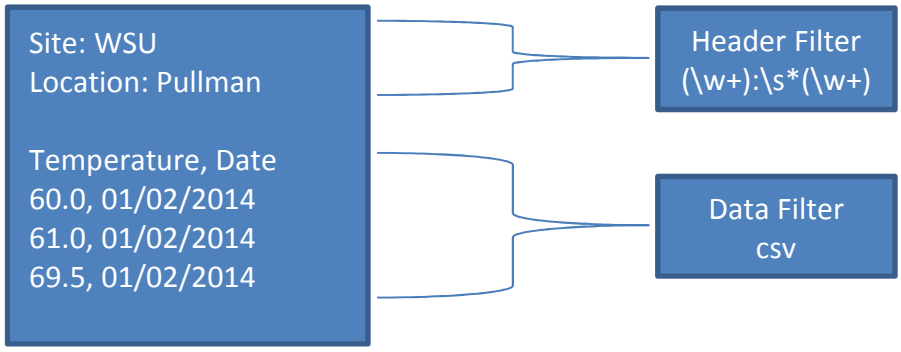
- csv : Parses the data part of a file as a CSV file. The first row is assumed to contain the column names and the following lines are data. NOTE: this option is intended for files with embedded metadata, if the file has no metadata and is already entirely in CSV format it is recommended that you use the pre-defined CSV filter.
- -d : Parses files that contain a "double header". A double header is where the header is split into two lines; the first line containing the column names and the second line contains the unit type. Once parsed the unit will appear in the format.
  column name [unit type]
- -t 1 : Parses a special case double header where the first line is discarded the second line has the column names, the third line has the unit types, and the fourth line is discarded.
- -md : Used in cases were the file contains embedded metadata AND a double header.

These settings can be combined for example; say you have a file with a double header and in CSV format, but no metadata. To create this i-filter you would leave the "filter header" blank and would type "csv -d" in the "data filter" field (without quotes).

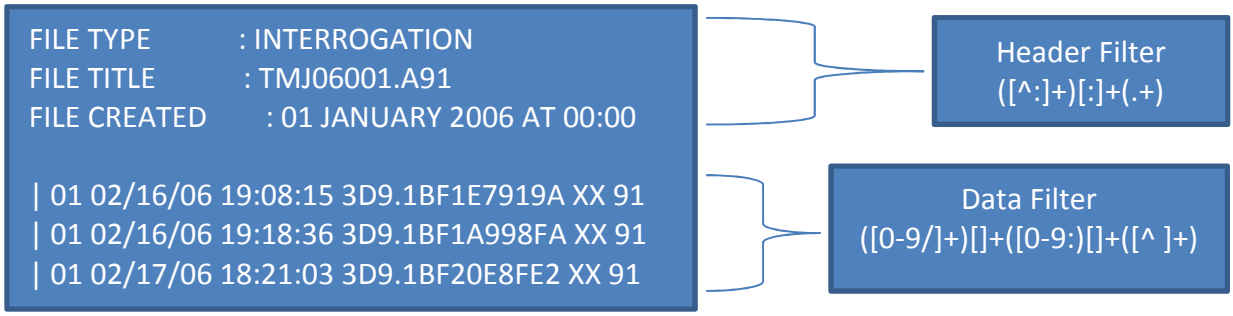
12.3 I-filter examples

To demonstrate the flexibility of i-filters and better show how i-filters work this section will give several examples. Note that to save space duplicate header filters are not repeated for each line.

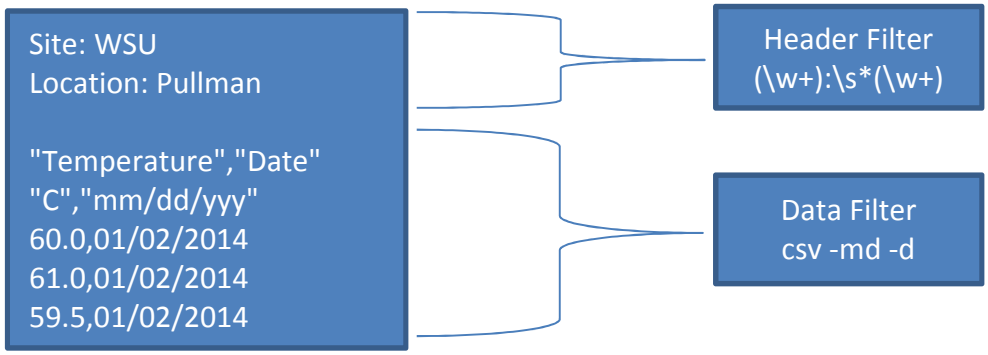
Say you have a file like the one below (left). Hatch will take the 1st and 2nd lines of the file (the header portion) and place them in the metadata table. The "data filter" setting in this case would filter the rest of the document as a CSV file and appear in Hatch as a table.



Another example involving PTAGIS data; the first three lines would be parsed as metadata. However, in this case the data portion is not in standard CSV format so a special regular expression must be created.



Here is an example where the file contains metadata and a double header. Note the need for the "-md" setting.



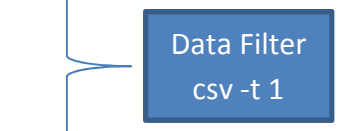
After filtering this file would look like the following document.

Label	Value
Site	WSU
Location	Pullman

Temperature [C]	Date [mm/dd/yyyy]
60.0	01/02/2014
61.0	01/02/2014
59.5	01/02/2014

Here is a special case file containing header lines that need to be discarded.

```
"========","========","=====","======","======","======="
"   Date","   Time","Temp","ODOsat","  ODO","Battery"
"  m/d/y","hh:mm:ss","   C","    %","  mg/L","  volts"
"--------","--------","-----","------","------","-------"
01/06/14,12:40:35,-0.08,91.9,13.47,5.6
01/06/14,12:50:35,0.30,92.1,13.35,5.6
01/06/14,13:00:35,0.34,92.3,13.37,5.6
```



12.4 Limitations

i-filters add great flexibility to Hatch but they can't handle all cases. This section will give examples of files that cannot be filtered. Usually if a file cannot be filtered properly it will fail and an error message can be seen if you view the document and hover the mouse over the "job finished" icon, or you can go to the job page by clicking the icon. Hatch keeps an original copy of file that fails to filter so a failure will not leave your data partly filtered or corrupted.

Sometimes a filter may be misapplied and the filter will not fail but will filter data incorrectly. If this happens the wrong filter was selected, there is a problem with the filter, the file is not in a compatible format, or a random error occurred (during upload, etc.)

Hatch cannot filter more than one metadata pair per line.

```
Site: Pullman  Location: WSU

"Temperature","Date"
60.0,01/02/2014
59.5,01/02/2014
```

Filtering this file would fail because there are two pairs of metadata on the same line.

Hatch cannot filter data that contains duplicate column names (this violates general database rules). Note this only applies to the column names not to the data values.

```
"Temp","Date","pH","pH"
60.0,01/02/2014,81,80
59.5,01/02/2014,80,80
```

This filter will fail because the column name “pH” appears two times on the first row.

Hatch cannot filter a file that contains more columns of data then there are column names. This problem can be hard to detect, note the extra comma after the dates.

```
"Temperature","Date"
60.0,01/02/2014,
59.5,01/02/2014,
```

When Hatch filters this file it will become confused because it expects two columns but finds three columns when it starts filtering the data. Even though the third column is empty this will cause filtering to fail.