HIEv User Manual

For software version 1.8 %%% Confirm SW version number

%%% Date April 2013 | v1.2

%%% TODO:

Replace all screen dumps

MAC or Windows? – Windows is OK.

What example data and login account should be used for screen dumps? Need to set up Facilities, Experiments, Experiment parameters, have at least 30 files so page buttons are shown, file names should be meaningful and not embarrassing, at least a few files with proper metadata, some TOA5 files with meaningful data, a few packages with meaningful metadata.

Should screen dumps show functions only available to Admins?

There has been a significant rise in the number of sensors and sensor networks used in environmental research in recent years. This growth has brought with it the challenge of managing sensor infrastructure and the data produced by the increasing numbers of deployed sensors.

The HIEv system was developed to address these challenges and specifically aims to:

* Ensure raw data is never lost
* Ensure that data can be used and interpreted in the future
* Allow researchers to make linkages between different types of data
* Make it easier for researchers to get access to the data they need
* Make it easier for technical officers to distribute the data
* Create entries in Research Data Australia (required by ANDS)

As a web application, HIEv is easily accessed from a variety of locations and platforms – all you need is a modern web browser and the URL of the server hosted by the research centre with which you are affiliated /collaborating.

Contents

[Contents 2](#_Toc351989254)

[1 Overview 4](#_Toc351989255)

[1.1 Installing HIEv 4](#_Toc351989256)

[2 Logging in to the system 5](#_Toc351989257)

[2.1 Classes of Users 6](#_Toc351989258)

[3 The HIEv Main Screen 7](#_Toc351989259)

[3.1 Signing Out 8](#_Toc351989260)

[3.2 Changing Your User Settings 8](#_Toc351989261)

[3.2.1 Overview Tab 8](#_Toc351989262)

[3.2.2 Edit Details Tab 8](#_Toc351989263)

[3.2.3 Change Password Tab 9](#_Toc351989264)

[4 HIEv Data File Storage and Metadata 10](#_Toc351989265)

[4.1 Basic Information 10](#_Toc351989266)

[4.2 Information Extracted from TOA5 Files 11](#_Toc351989267)

[4.3 Column Information for TOA5 Files 12](#_Toc351989268)

[5 Facilities and Experiments 13](#_Toc351989269)

[5.1 Creating a Facility Entry 13](#_Toc351989270)

[5.2 Editing a Facility Entry 15](#_Toc351989271)

[5.3 Creating an Experiment Entry 15](#_Toc351989272)

[5.4 Editing an Experiment Entry 18](#_Toc351989273)

[5.5 Setting Up Experiment Parameters 18](#_Toc351989274)

[6 Uploading Data files 21](#_Toc351989275)

[6.1 Uploading RAW TOA5 data files 22](#_Toc351989276)

[6.2 Manual Data Upload Action Summary 23](#_Toc351989277)

[6.3 Automating the upload of data to HIEv 24](#_Toc351989278)

[7 Managing Data Files 25](#_Toc351989279)

[7.1 The Dashboard Tab 25](#_Toc351989280)

[7.2 The Explore Data Tab and File Searching 25](#_Toc351989281)

[7.2.1 Sorting 26](#_Toc351989282)

[7.2.2 Searching 27](#_Toc351989283)

[7.3 The Cart 35](#_Toc351989284)

[7.3.1 Editing the Cart Contents 36](#_Toc351989285)

[7.4 Viewing and Editing a File's Metadata 36](#_Toc351989286)

[7.5 Deleting a Data File 37](#_Toc351989287)

[8 Publishing Your Data 39](#_Toc351989288)

[8.1 Creating a Package 39](#_Toc351989289)

[8.2 Publishing a Package 40](#_Toc351989290)

[8.3 Adding a Package’s External ID 40](#_Toc351989291)

[8.4 Managing Published Packages 41](#_Toc351989292)

[8.4.1 Publishing a second time 41](#_Toc351989293)

[8.4.2 Deleting Published Packages 41](#_Toc351989294)

[8.4.3 Editing Published Packages 41](#_Toc351989295)

[8.4.4 Correcting Published Packages 42](#_Toc351989296)

[8.5 Viewing Published data 42](#_Toc351989297)

[9 Downloading files 44](#_Toc351989298)

[10 System Administration 45](#_Toc351989299)

[10.1 Managing Users’ Details 45](#_Toc351989300)

[10.2 Authorising New Users – The Access Requests Tab 46](#_Toc351989301)

[10.3 Managing Column Mappings 47](#_Toc351989302)

[10.3.1 The Column Mappings tab 48](#_Toc351989303)

[10.3.2 Updating from the Explore Data tab 48](#_Toc351989304)

[11 Configuring Tags, Column Mappings and Experiment Parameters 50](#_Toc351989305)

[12 Migrating data to a new system 52](#_Toc351989306)

[13 Revision History 53](#_Toc351989307)

[Appendix A - The Bagit format 54](#_Toc351989308)

[Appendix B - RIF-CS 55](#_Toc351989309)

[Appendix C - Data File Upload Scenarios 56](#_Toc351989310)

1. Overview

HIEv is designed to act as a central repository for environmental research data. Technicians can configure their field PCs to automatically push time-series data from sensors into HIEv, while researchers can use the system to discover and download the latest data available. Rich metadata is stored for physical infrastructure (“Facilities”), the Experiments that run at those facilities, as well as the individual files to support discovery and interpretation.

DC21 Diagram

All files in HIEv are grouped by Experiment, providing a convenient way to organise related files such as cleansed or gap-filled data, and analysis outputs.

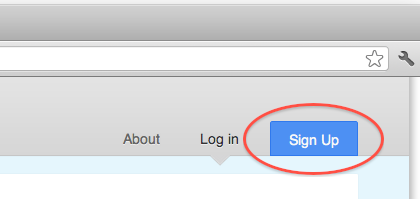
Once finalised, Packages of data can be defined, described and published to ANDS. This enables researchers from outside the organisation that produced the data to discover it, and to request access to download a copy.

* 1. Installing HIEv

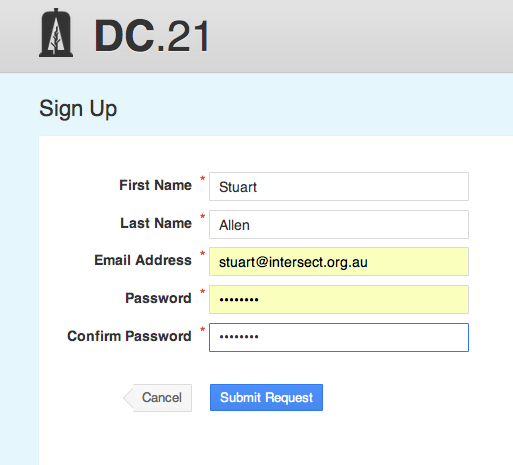
All instructions for installing and upgrading HIEv are held at [Deployment Guide](https://github.com/IntersectAustralia/dc21/wiki/Deployment-Guide) on the project’s [GitHub Wiki](https://github.com/IntersectAustralia/dc21/wiki).

1. Logging in to the system

To begin using HIEv, enter the system URL %%% What is it or how do you find out? into your web browser. Before you can login you are required to have a system account. You can apply for an account by clicking the blue "Sign Up" button on the top left of the screen:

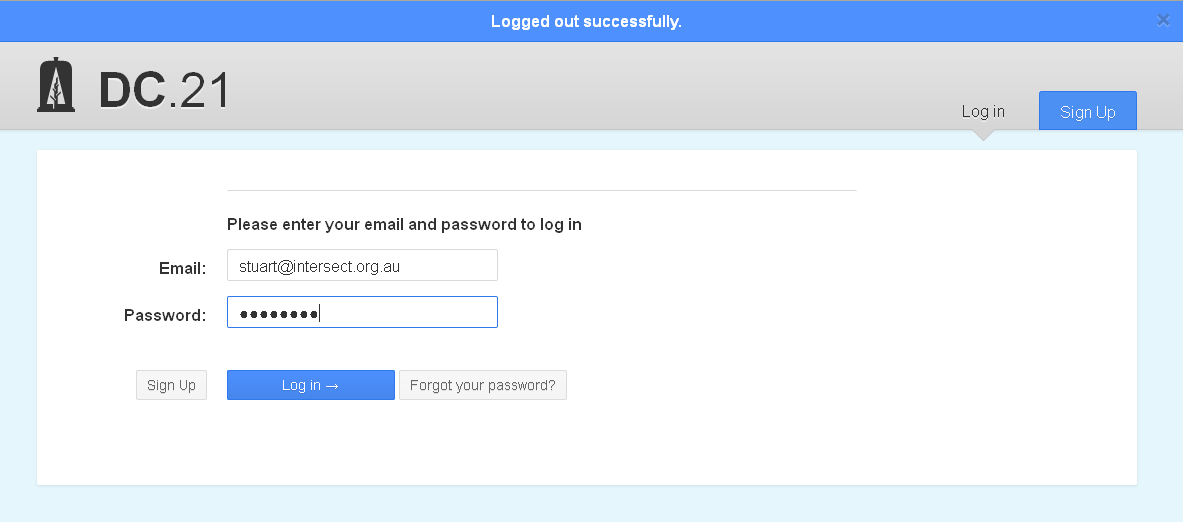


This will take you to a form where you will be requested to enter your first name, last name, email address and chosen password. (Note: Passwords must be between 6-20 characters and contain at least one each of: an uppercase letter; a lowercase letter; a digit and a symbol.)



Once you have filled out the form and clicked "Submit Request" an email will be sent to the system administrator who will either approve or deny your request for access. If your request is approved you will receive an email informing you that you can now login using the password you entered on the original sign up form.

To see the login form, make sure you have the "Log in" tab selected on the top right. Next enter your Email address and password and click the blue "Log in" button below:



Once you have logged in you will be taken to the main screen for the HIEv application.

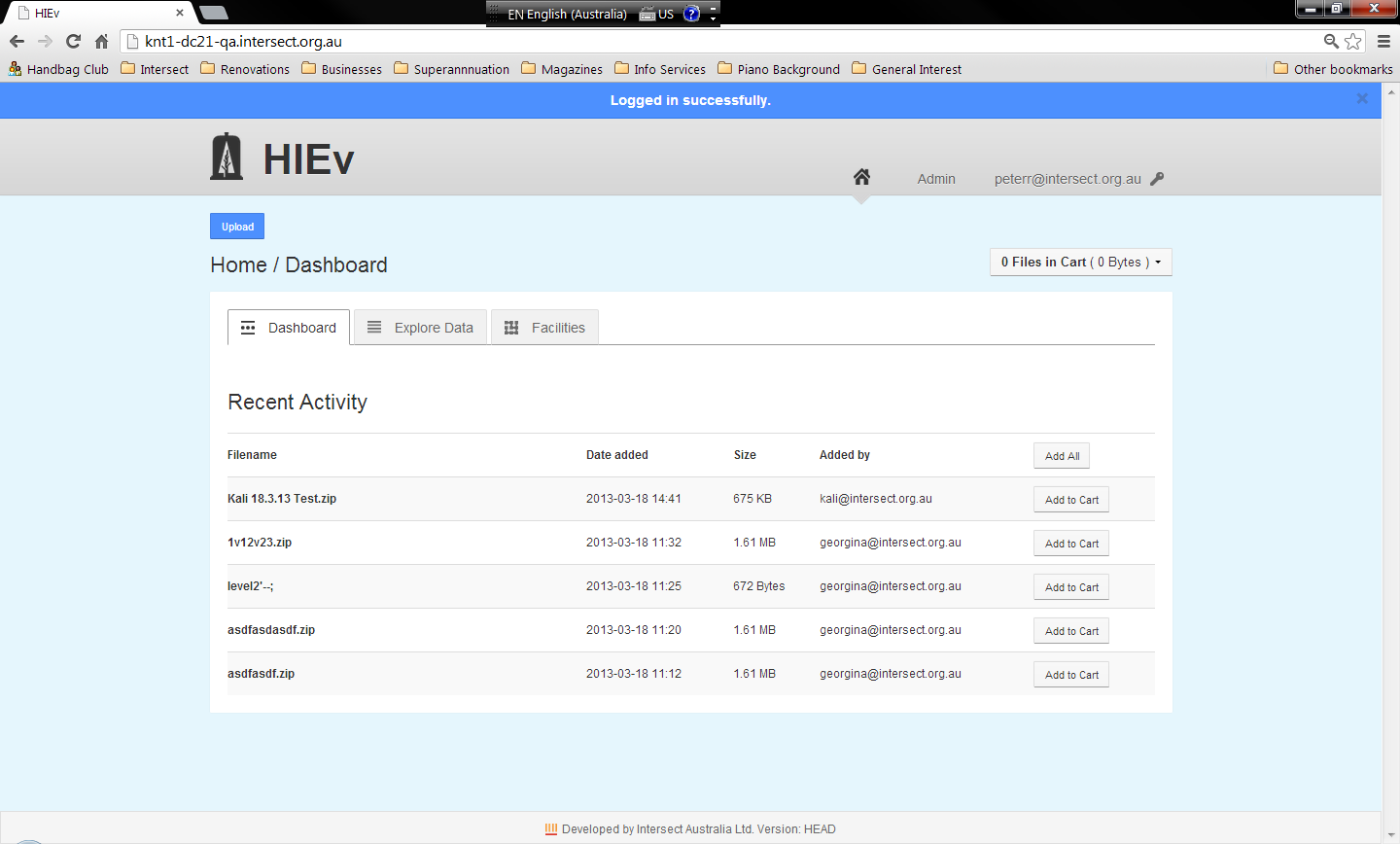
* 1. Classes of Users

HIEv defines three classes of users. You will be assigned to one of these classes by the person who authorises you request for Sign in to the system. These three classes are restricted in the following ways:

|  |  |
| --- | --- |
| Researcher | No access to the Admin tab and its functions.  Cannot edit the metadata of files uploaded or created by other users.  Cannot delete files created by other users.  Cannot use the API functions. The API token can be generated but it cannot be used. |
| API Uploader | No access to the Admin tab and its functions.  Cannot edit the metadata of files uploaded or created by other users.  Cannot delete files created by other users. |
| Administrator | Permission to perform all functions in the HIEv System, including authorising new users’ requests for access to the HIEv System. |

1. The HIEv Main Screen

The Main Screen consists of the following parts:

Figure : HIEv Main Screen

%%% Add reference numbers to this diagram when the final cut & paste is done.

Most of these components are common to many screens in the HIEv system.

|  |  |  |
| --- | --- | --- |
| 1 | Home button | Click to show the Home/Dashboard (as shown in the view above). |
| 2 | Admin button | Click to access Admin functions (see Chapter ). This button is only present if you have Administrator permissions. |
| 3 | Login ID | This is your login name. Click to open a drop down menu of user operations. (See sections 3.1 and 3.2) |
| 4 | Action button | In many screens, there is an action button at the top left corner. It is often an Upload button, which allows you to upload new data files to HIEv. (See Chapter for more information.) However, it may also be a button for another function which is more relevant to the data being displayed in the current view. |
| 5 | Cart status box | The HIEv web interface allows you to add files to a Cart, which operates like an e-Commerce shopping cart. Click in the Cart Status to open a drop down menu of Cart functions. (See section ) |
| 6 | Working area | This contents of this work area changes as you perform HIEv operations. |
| 7 | Version indicator | This shows the version of HIEv which you are accessing. |

* 1. Signing Out

Click on your login ID at the top right of the screen to see a drop down menu. Click on Sign Out to finish your HIEv session.

* 1. Changing Your User Settings

Click on your login ID at the top right of the screen to see a drop down menu. Click on Settings to access the following three tabs.

When finished, click on the Home button to return to the HIEv Main Screen.

* + 1. Overview Tab

This tab displays a summary of your user information.

%%% Screen dump of tab

|  |  |
| --- | --- |
| User Name: | Your valid email address which you use for logging on. |
| First Name Last Name: | Your name. |
| API Token: | A string of characters which you can use as an authorisation token in scripts which you write to make use of the HTML API for HIEv. See the API definition on the GitHUB WIKI for DC21/HIEv for instructions on using the HTML API.  Clicking on Generate Token will cause a token to be displayed in this field. Copy and paste it into the required place in your API scripts.  Clicking on Re-generate Token will cause the current token to be invalidated and a new token to be generated and displayed. You must replace the token value in your API scripts with this new token so that your API scripts continue to work. This button is only displayed if a valid token is available.  Clicking on Delete Token will invalidate the displayed token. Your API scripts will no longer work. This button is only displayed if a valid token is available. |

* + 1. Edit Details Tab

The Edit Details tab allows you to update your First Name and Last Name. Click Cancel to return to the Overview Tab without accepting changes, and click Update to store the changed values you’ve entered.

%%% Screen dump of tab

* + 1. Change Password Tab

Use this tab to change your HIEv logon password. You must correctly enter your current password and the values you enter for New password and Confirm new password must be identical for your password change request to be processed. Click Cancel to return to the Overview Tab without accepting changes, and click Update to store the changed password you’ve entered.

%%% Screen dump of tab

1. HIEv Data File Storage and Metadata

The HIEv System stores uploaded data files using a database structure. In addition to storing the files themselves, HIEv also stores metadata about each file. This metadata falls into three categories.

* 1. Basic Information

The Basic Information is metadata entered by the user when the data is uploaded. It consists of the following fields.

|  |  |
| --- | --- |
| **Name** | The name of the file as it is stored in the HIEv system. |
| **Type** | The Type of the file is a single value that describes the data contained within the file. This value is chosen from a constrained list of possibilities defined by the system administrator. The file's **Type** is generally aimed at tracking data through its various stages of processing.  The Type **Package** indicates a file containing a collection of data files which is intended for Publishing. See Chapter for more information. |
| **File Format** | There are two known file formats. Possible values for this field are:  TOA5 The file was inspected on upload and discovered to be TOA5 format. TOA5 files are processed differently on file upload. See section for more information.  BAGIT The file is a Package which is formatted as a BAGIT ZIP file. See for more information on BAGIT files.  UNKNOWN The file format is not known to the HIEv system. |
| **Description** | The description entered by the user to describe the contents of this file. |
| **Tags** | The file Tags are a set of flags that have been given by the user to the file from a constrained list of possibilities defined by the system administrator. No tags or multiple Tags can be applied to any file. |
| **Experiment** | This field indicates which experiment produced the file. Each file can be associated with only one experiment. Any user with the appropriate permissions can create experiments. See Chapter for more information. |
| **Facility** | The Facility field indicates which facility the above experiment was run at. Each file must be associated with exactly one facility. Any user with the appropriate permissions can create facilities. See Chapter for more information. |
| **Date added** | The date on which this file was added to the HIEv database. For Packages, it’s the date on which it was Packaged. |
| **Size** | The size of this file. |
| **File ID** | File IDs are unique integers which are assigned and used internally by the HIEv system to identify files. File IDs cannot be changed by the user. In general, they will not change, but in the case of TOA5 files, may sometimes change after uploading further TOA5 data. |
| **ID** | This field provides the opportunity for users to enter an additional ID which has been used outside of the HIEv system to identify this data. It is a character string and is typically used to enter the ID which has been assigned to a Published Package on an external data store for Published experimental data. No two files in the HIEv system can have the same non-null ID. See section Adding a Package’s External ID  Once a Package has been Published and harvested, an ID is generated for it by the external Published data store.  To edit a Package’s metadata:   * Navigate to the Package file using the Explore Data tab. * Click on the Package file’s filename to open the metadata view screen. * Click on the Edit Metadata button to open the metadata edit screen.   %%% Edit Package Metadata screen dump   * Enter the new ID into the metadata ID field. * Click Update to save the change.   See section 7.4 Viewing and Editing a File's Metadata for more information on using this edit function.  Note Editing most fields of a Package’s metadata is not recommended. It does not modify the data in the matching RIF-CS file. Please see section 8.3 **Error! Not a valid bookmark self-reference.** for more information about restrictions on modifying Packages’ metadata.  Managing Published Packages for more information about the use of this ID. |
| **Added by** | This field indicates the user who uploaded the file to the HIEv System. For Packages, it’s the user who Packaged it. |
| **Published** | This applies to Packages only and indicates whether the Package has been Published or not. |
| **Published date** | This applies to Packages which have been Published only and indicates whether the date on which the Package was Published. |
| **Start time End time** | These fields apply to non-TOA5 data files and Packages only. They hold the dates which were manually entered when the file was uploaded or created. These dates indicate the start and end times of the data in the file. (For TOA5 files, start and end times are automatically extracted and stored as part of the Summary Information.) |

* 1. Information Extracted from TOA5 Files

Summary Information is stored for TOA5 files only. It is collected automatically from the TOA5 data file and is not editable by HIEv users.

%%% Is it worth listing all of these? The creators of the data should know better than me what it is!

|  |  |
| --- | --- |
| **Start time End time** | The first and last times of the observations found in the TOA5 data file. |
| **Sample interval field** | The frequency of samples in the data file, if relevant. |
| **Datalogger model** | The model of data logger used to generate the TOA5 file. |
| **Station name** | %%% Are these fields manually entered at some point, or extracted automatically, or does it vary? Is there anything useful we can say, or should we just list them? |
| **Serial number** |  |
| **Os version** |  |
| **Dld name** |  |
| **Dld signature** |  |
| **Table name** |  |

* 1. Column Information for TOA5 Files

Column Information is collected automatically from TOA5 files when they are uploaded and cannot be changed by HIEv users. This automatically collected Column Information for a TOA5 data file is displayed whenever the TOA5 data file’s metadata is displayed.

%%% Screen dump of example with Name column highlighted.

In addition to the Column Information display there is one more column, which is found by searching for the TOA5 column heading in the HIEv System’s Column Mapping table. This extra descriptive value is displayed as the Name column in the metadata display. See section for more information about setting up the Column Mapping table.

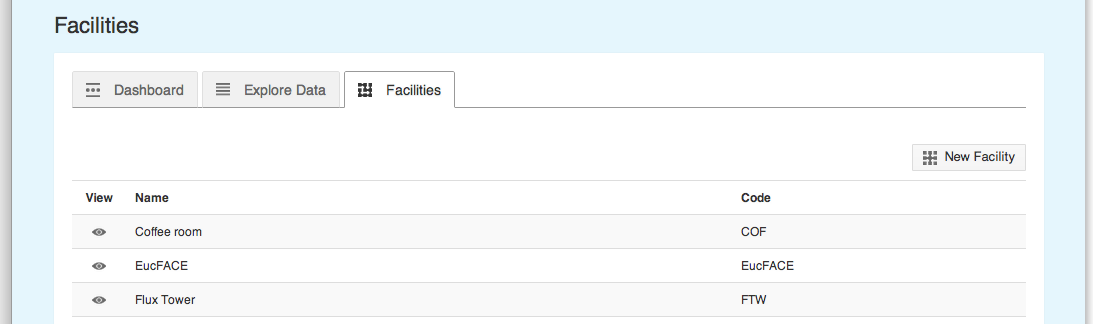
1. Facilities and Experiments

All data files uploaded to the HIEv system must be associated with an Experiment. In turn, each Experiment is associated with a Facility.

Therefore entries for Facilities and Experiments must be created before the associated data files are uploaded.

|  |  |
| --- | --- |
| Facility | A Facility represents any instrument, or discrete set of instruments that are used in concert. This could be a multi-million dollar fixed Facility with hundreds of instruments and sensors or a single piece of portable equipment that is taken out into the field. |
| Experiment | An Experiment is a set of related tests performed at a single Facility for which the data will be uploaded into the HIEv system. |

Facilities and Experiments entries are entered and updated using the Facilities tab on the main HIEv screen.

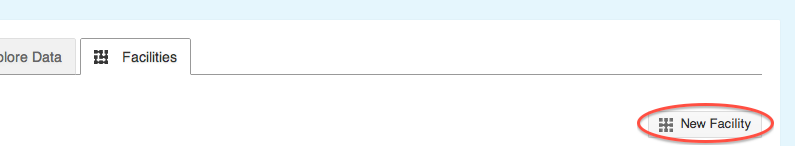


The Facilities tab lists all of the Facilities currently defined in the HIEv system.

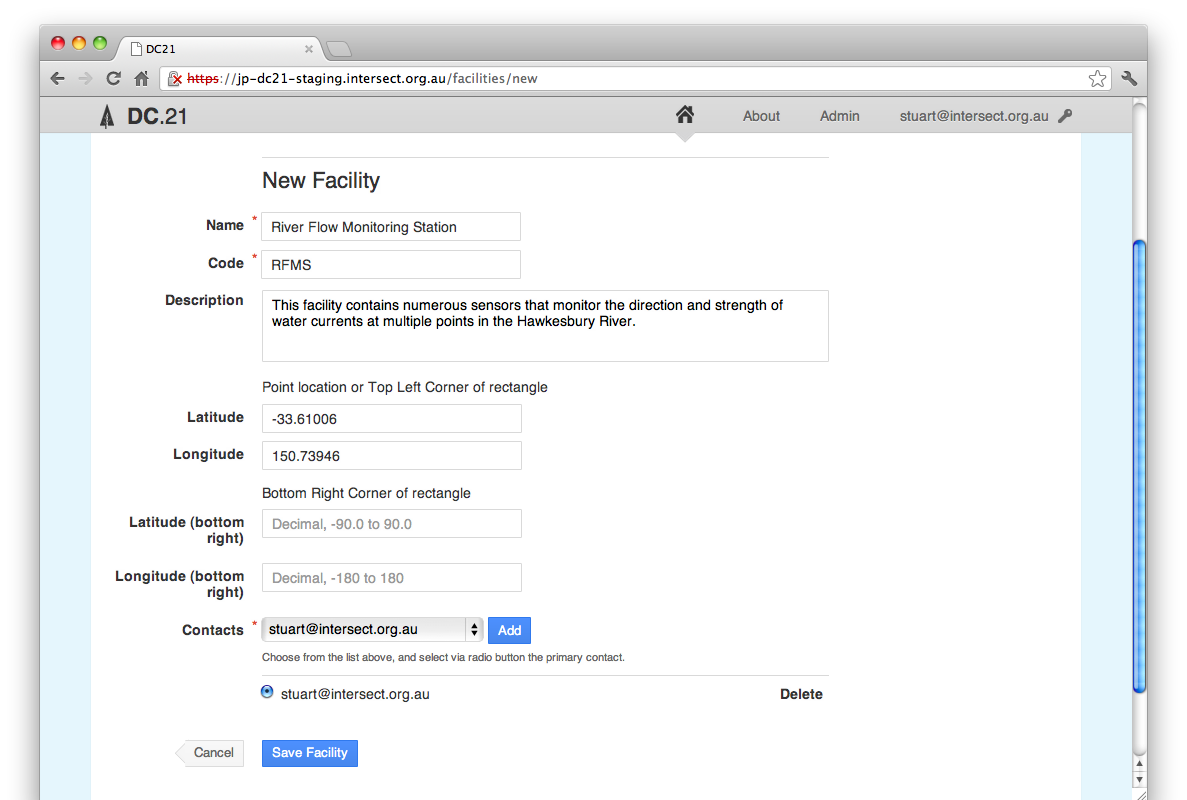
* 1. Creating a Facility Entry

Note Take care. Once created, a Facility entry cannot be deleted. This restriction prevents Experiemnts which reference a Facility from becoming invalid if the Facility were to be deleted.

Facilities are added by clicking the New Facility button at the top-left of the Facilities tab:



This will display a form that allows you to add details about the facility you would like to create:



Enter the details for the Facility.

|  |  |
| --- | --- |
| Name | The Name for the facility is a short, plain-English title that will be used in the application interface to refer to the facility. |
| Code | The Code for the facility is a short unique string. %%% When is this used? The Name is used in dropdown lists! |
| Description | The Description of the facility should be as comprehensive as possible, describing details that would help a researcher both discover the facility when searching and assist the researcher in being able to interpret the data that is produced by the facility. These details would include things such as:   * The purpose of the facility * Types of sensors installed at the facility * Location of the sensors within the facility |
| Latitude and Longitude fields | The Latitude and Longitude for the facility are expressed in Decimal Degrees (<http://en.wikipedia.org/wiki/Decimal_degrees>) and can be taken directly from Google Maps.   * If a single set of co-ordinates is given it is considered to be the central point for the Facility. Enter them only in the Latitude and Longitude fields, leaving the fields for the bottom-right corner empty. * If two sets of co-ordinates are given they are considered to be a rectangle that bounds the facility. Enter the northwest corner in the first two fields and the southeast corner in the fields labelled “bottom right”. |
| Contacts | The Contacts for the Facility are selected from the users registered within the HIEv system. There must be at least one Contact for each Facility.  %%% Contacts dialog screen dump with a couple of contacts already added. Or ensure this info is clear in the screen dump above, perhaps with a highlighting circle.  Select a Contact from the email addresses shown in the dropdown list and press the Add button to add it to the Contact list shown below this question.  To remove an incorrectly added Contact, press the Delete button corresponding to that Contact’s email address in the Contact list.  Ensure that the primary Contact is highlighted correctly with the radio buttons to the left of the Contacts in the Contact list. |

Click on Save Facility to return to the Facilities list.

Once Facilities have been created they will appear in the list on the Facilities tab:

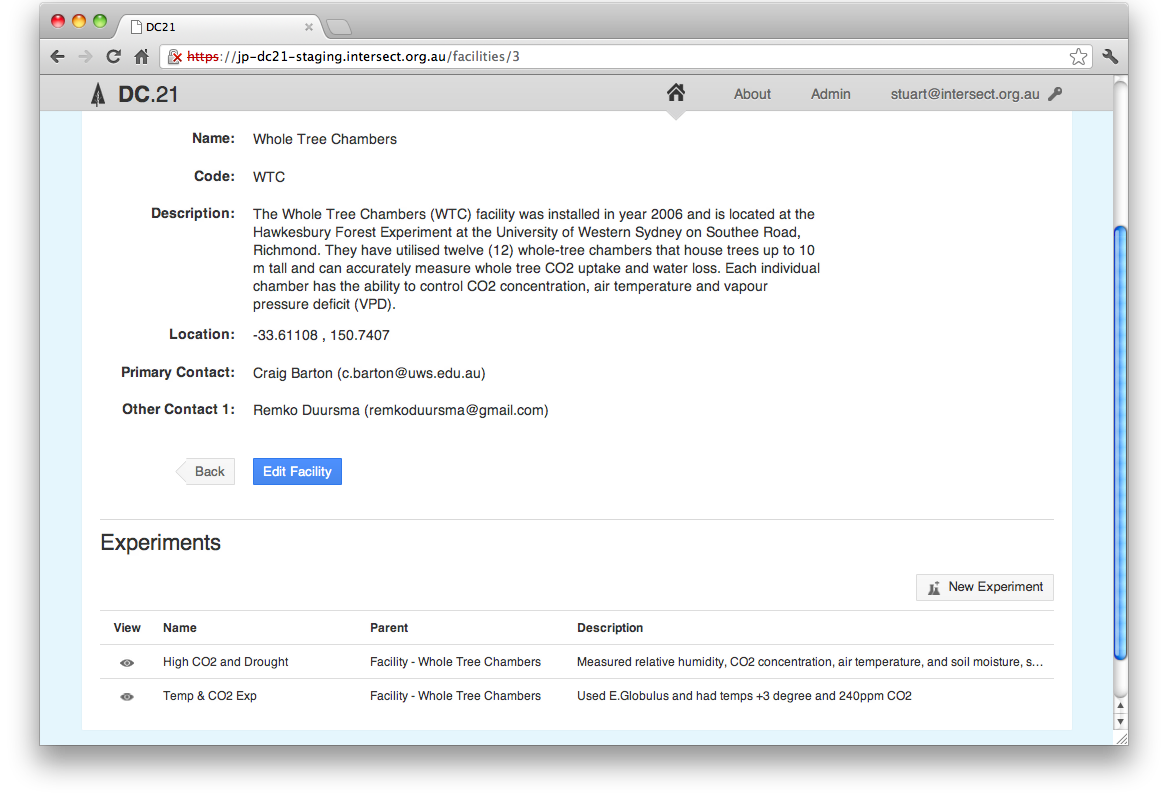
* 1. Editing a Facility Entry

To modify any of the parameters for a Facility entry:

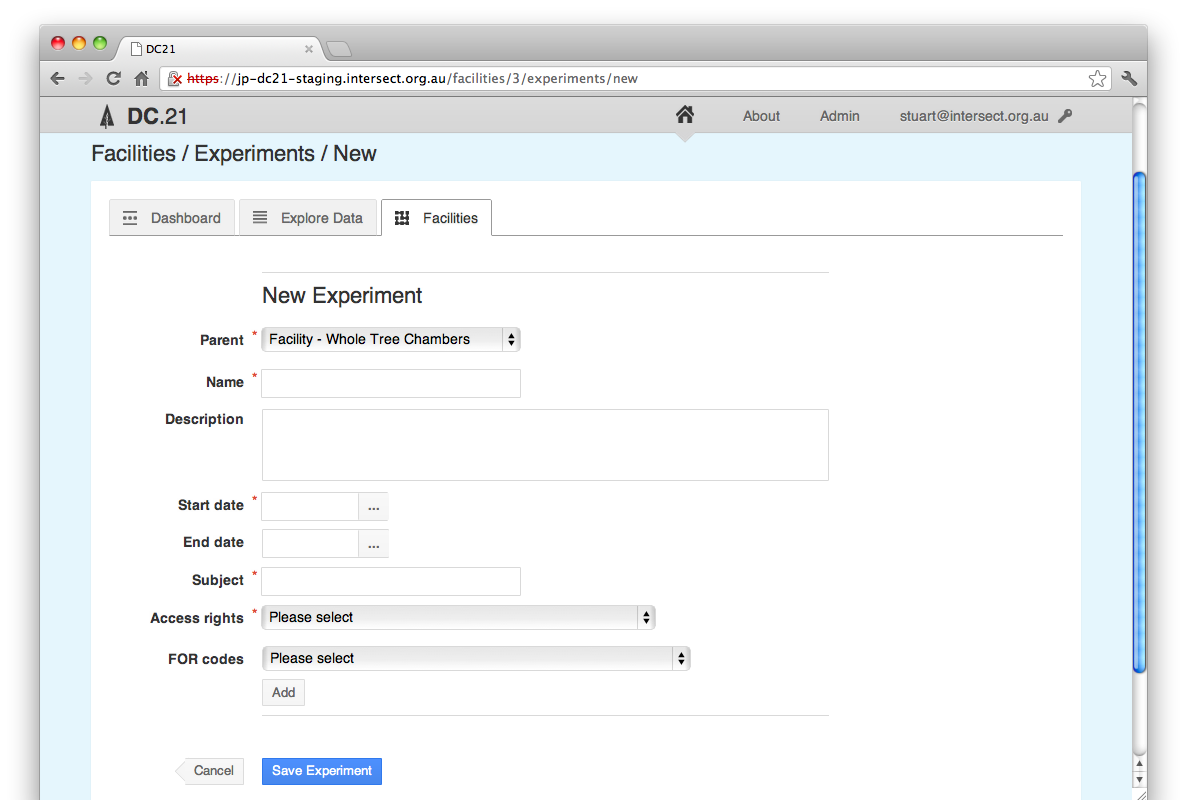
* From the Facility tab, click on the Facility Name of the entry you wish to edit. The Facility details will be displayed.
* Click on Edit Facility to open the details edit screen, which is the same as the one described in above.
* Change the details as required.
* Click on the Update button so save your changes.
  1. Creating an Experiment Entry

Note Take care. Once created, an Experiment entry cannot be deleted. This prevents data files which reference this Experiment from becoming invalid.

To create and Experiment, on the Facilities tab view the details of the Facility by clicking on the Facility name. All of the Experiments are listed below the details of the Facility.



Then click on the New Experiment button at the top left of the Experiment list to show the Experiment parameters entry dialog.



Enter the details for the Experiment.

|  |  |
| --- | --- |
| Parent | The Parent for an Experiment is either the current Facility, or another Experiment running at that Facility. If an Experiment is selected, the new Experiment is considered a sub-experiment of the one selected. |
| Name | The Name for the Experiment should be short, but descriptive enough to uniquely identify the Experiment, including distinguishing an Experiment from those that are likely to come in the future. |
| Description | The Description for the Experiment should describe the purpose of the Experiment and the techniques employed. Particular focus should be given to aspects of the Experiment that produce data that is stored in this system. |
| Start date | The Start date for the Experiment is the date that Experiment was first considered to be active. |
| End date | The End date for the Experiment is the date that the Experiment concluded. This field should be left blank for Experiments that are currently active. |
| Subject | The Subject for the Experiment is a short phrase describing the Experiment's main research area. The Subject is primarily recorded to support publication to [ANDS](http://www.ands.org.au/guides/cpguide/cpgsubject.html) and in their own words, “A subject is a term, keyword, classification code or phrase representing the primary topic or topics covered by a registry object.” |
| Access rights | The Access rights drop down list box provides a selection of licenses to release the data from this Experiment under. It is preferred in Australia that data is released under a [Creative Commons](http://creativecommons.org.au/learn-more/licences) licence. |
| FOR codes | Each experiment can have one or more Fields of Research (FOR) Codes. These FOR Codes are the Australian and New Zealand Standard Research Classification (ANZSRC) codes. More information can be found at <http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/4AE1B46AE2048A28CA25741800044242?opendocument>.  The FOR Code is a hierarchical classification with three levels:   * Divisions (two digits), which are based on a broad discipline * Groups (two digits) within each Division share the same broad methodology, techniques and/or perspective as others in the Division. * Fields of Research (two digits) within each Group are related. Fields of Research are categorised to the Divisions sharing the same methodology rather than the Division they support.   A unique number identifies each level. FOR Codes can be specified to Division, Group or Field of Research level.  FORCodes  FOR Codes are selected two digits at a time. After the Division is selected, a dropdown list box is displayed to allow selection from the Groups which are relevant to the selected Division. Similarly, after the Group is selected, a dropdown list box is shown to allow selection from the Fields of Research which are relevant to the selected Group.  Click the Add button to add the selected FOR Code to the list for this Experiment, which will appear below the FOR Code selection boxes.  FOR Codes can be removed from this experiment by clicking the Delete button to the right of the FOR Code in the list. |

After you have correctly entered the Experiment details, click Save Experiment at the bottom of the page.

To abandon creating the Experiment, click Cancel.

* 1. Editing an Experiment Entry

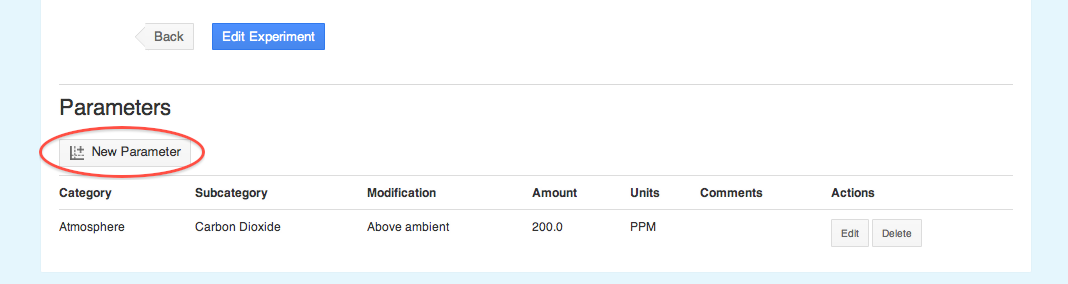
To modify any of the details for an Experiment entry:

* From the Facility tab, click on the Facility Name of the Facility which hosts the Experiment. The Facility details and Experiments for the Facility will be displayed.
* Click on the Experiment name for the Experiment you wish to modify. The Experiment details will be displayed.
* Click on Edit Experiment to open the Experiment detail entry screen, which is the same as the one described in above.
* Change the details as required.
* Click on the Update button so save your changes.
  1. Setting Up Experiment Parameters

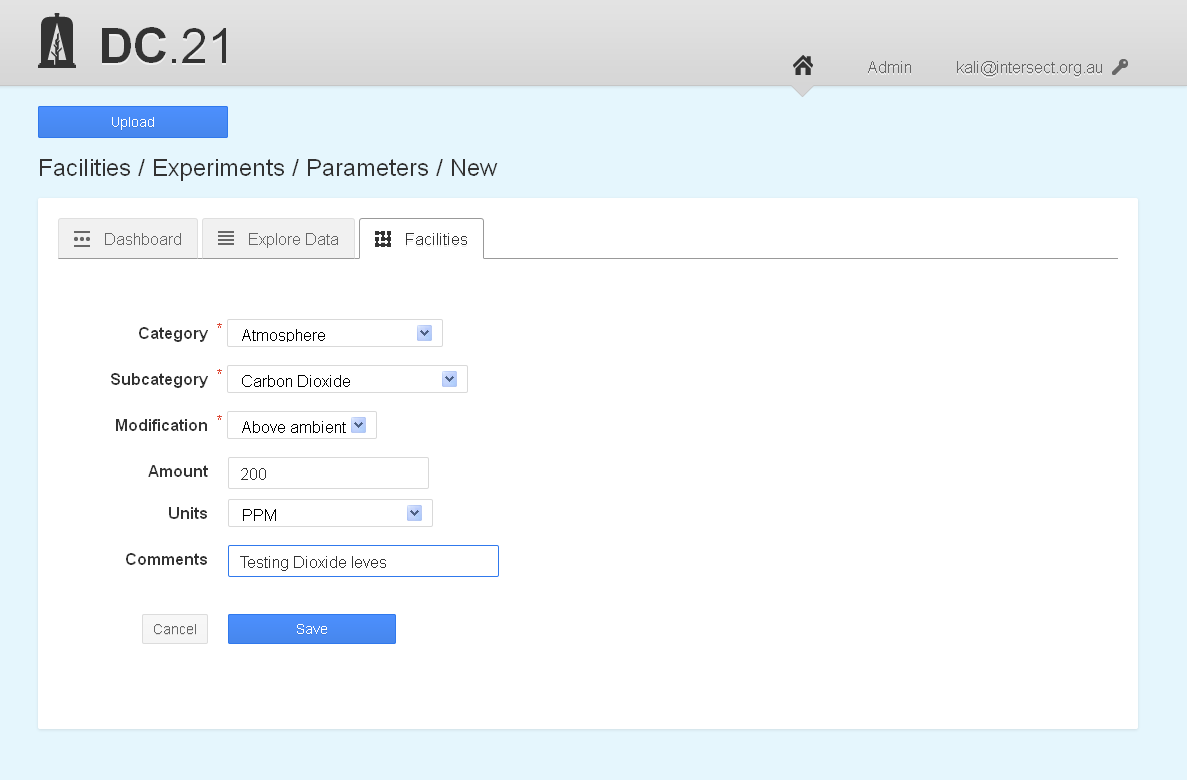
Experiments can optionally have one or more Experiment Parameters. These Parameters provide a structured way to describe Experimental treatments. They are documentary only. An Experiment can have multiple Parameters or none at all. Examples may be the details of the raising the CO2 within the tree chambers.

Because Parameters are not directly referenced by data files, the option to delete then is provided.

Experiment Parameters are added by clicking the New Parameter button directly below the Experiment details:



This button will display the form below:

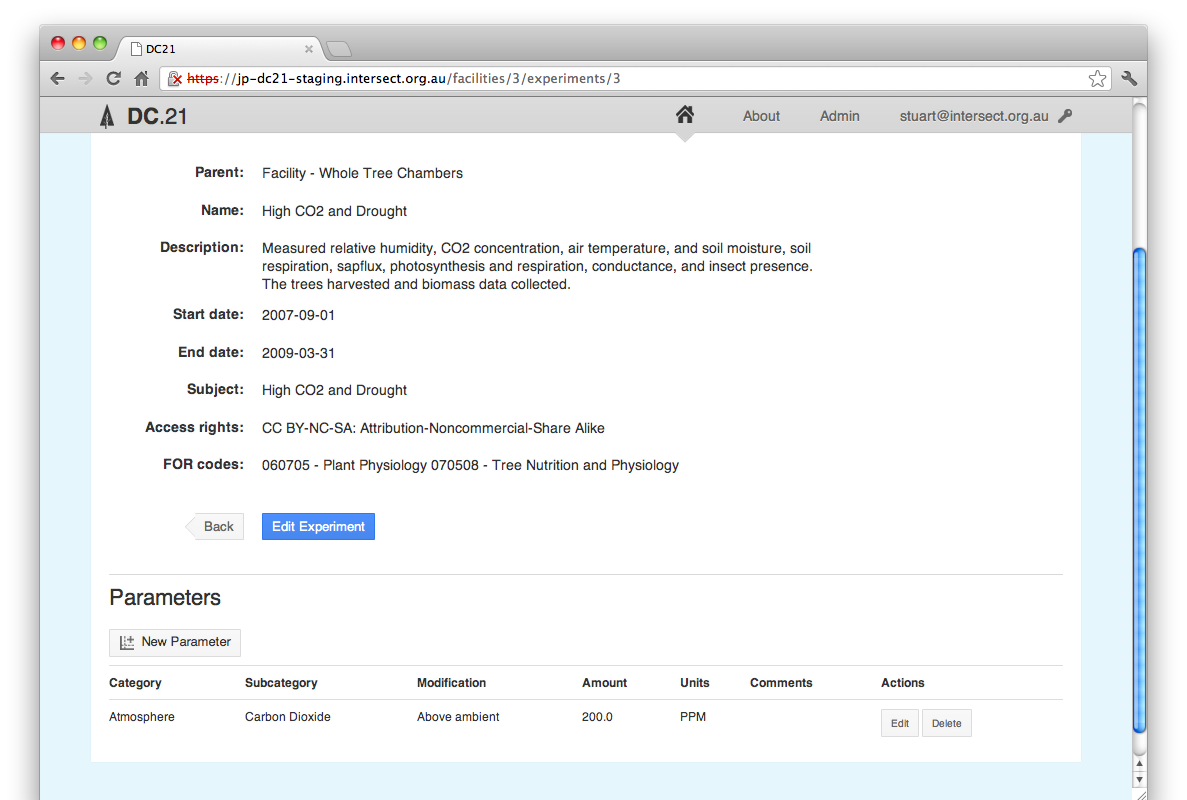


The system administrator configures the values available in the dropdown list boxes for the questions in this form.

|  |  |
| --- | --- |
| Category Subcategory | These fields allow you to select the medium that is being modified from the dropdown lists. These fields are mandatory. |
| Modification | The Modification indicates the general way in which the medium has been modified. This field is mandatory. |
| Amount Units | These optional fields allow more specific information to be recorded about modification. |
| Comments | This field can be used to record any unstructured, plain-text information you would like to record about the treatment. |

To finish, click the blue **Save** button at the bottom of the form.

Once an Experiment Parameter has been created, it will appear in the Parameter list below the main description of the Experiment:

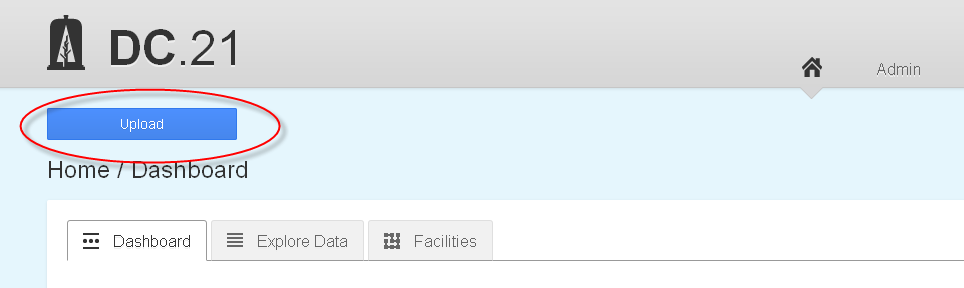


%%% This screen dump would be better with a number of parameters shown and highlighted.

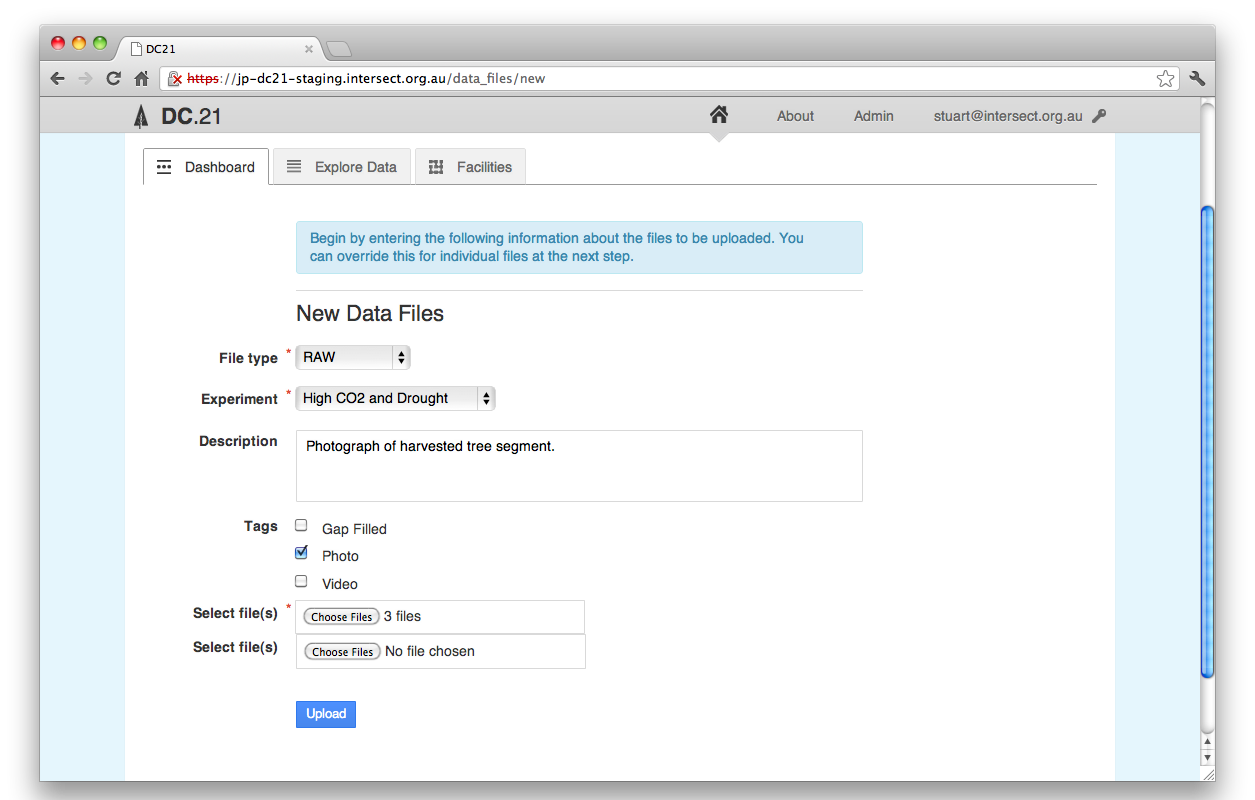
Existing Experiment Parameters can be edited or deleted using the appropriate button to the right of the parameter in the **Actions** column.

1. Uploading Data files

New files are added to the system using the blue **Upload** button at the top left of the screen:



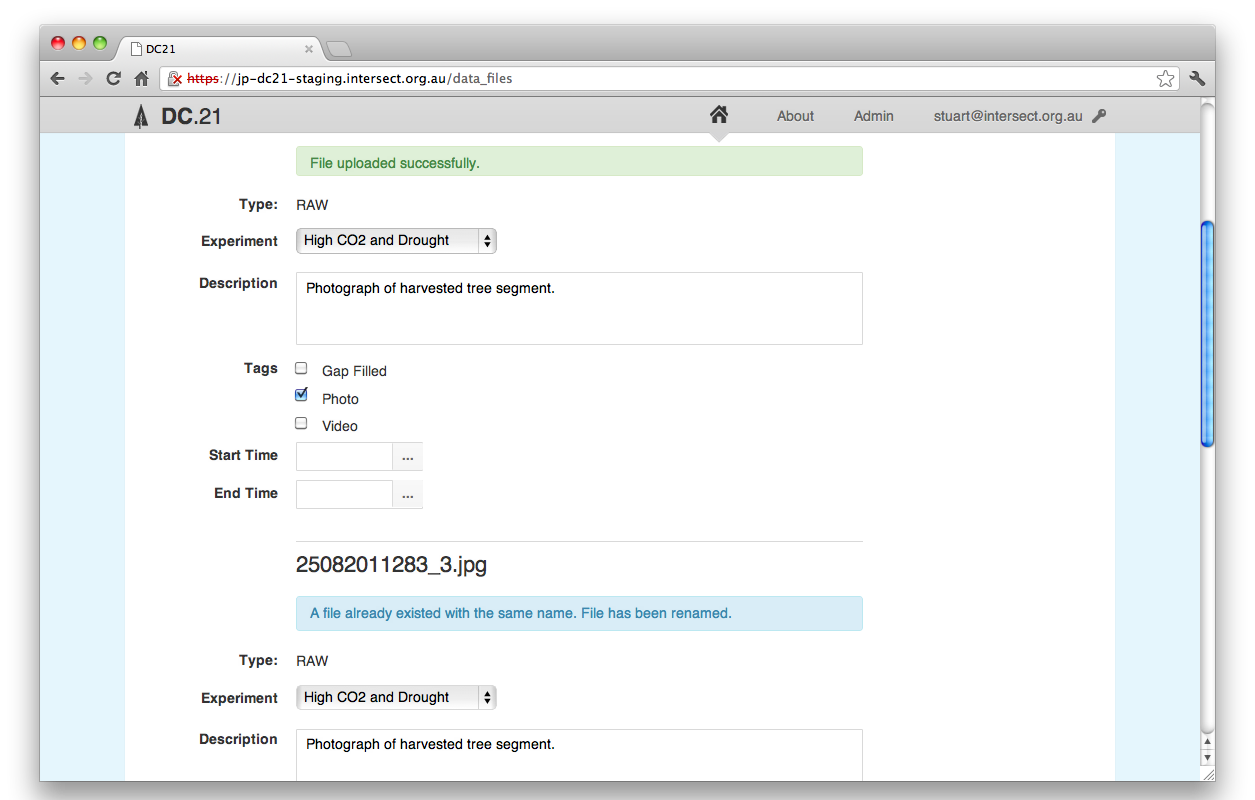
Clicking this button will take you to the New Data Files form:



On this form you will enter all the metadata for the new file or files and also select the files to be uploaded. See the section 4.1 Basic Information for information about the meaning of Metadata fields.

Note Files that are uploaded with a File Type of **RAW** and are also detected as containing valid TOA5 header information are treated as a special case. See the end of this section for more information.

After the file or files have successfully uploaded, the supplied metadata will be applied to all uploaded files and you will be presented with a screen on which you can edit all of the uploaded files’ metadata individually. This is useful, for example, if you wish to give all files the same description but add an extra tag to one of the files.



If the start and end dates and times for the data cannot be automatically extracted for the file, the above screen will give you with the opportunity to enter this information manually. Dates can be typed into the **Start Time** and **End Time** fields in *YYYY-MM-DD* format or by clicking on the ellipsis to the right and selecting a day from the calendar that is displayed.

If an uploaded file has the same filename as another data file that already exists within the system, HIEv will automatically suffix a unique number to the end of the original filename, before the file extension.

* 1. Uploading RAW TOA5 data files

When a TOA5 CSV file is uploaded with the **Type** of RAW, it is considered to become part of the canonical stream of data for that data logger. As a result, there will only ever be a single file with a **Type** of RAW that contains any given sample from a TOA5 data logger.

This has the effect that:

1. If TOA5 file is uploaded with a **Type** of RAW, and the file being uploaded is a complete superset of another file (or files) that are also RAW TOA5 files from the same data logger, the subset files will be replaced with this new file, regardless of the file names.
2. If a TOA5 file is uploaded with a **Type** of RAW that only partially overlaps an existing file of RAW data from the same data logger, the file will be uploaded, but its **Type** changed to ERROR and the original file(s) left in place.
3. If a TOA5 file is uploaded with a **Type** of RAW that overlaps an existing file of RAW data, but does not pass a sample-by-sample comparison with the original file(s), the file will be uploaded, but its **Type** changed to ERROR and the original file(s) left in place.

The Start and End Dates for the observations in a TOA5 file are stored in the file. Therefore, when TOA5 files are uploaded as RAW data, HIEv does not permit the user to enter Start and End Dates.

* 1. Manual Data Upload Action Summary

| **IF...** | | | | **THEN...** | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Type selected is** | **File type is** | **File overlap is** | **And File** | **Resulting type will be** | **Resulting file name will be** | **Resulting messages to user** |
| RAW | TOA5 | None | does not already exist | RAW | as per uploaded | success |
| RAW | TOA5 | None | already exists | RAW | suffixed - see (1) | filename change (3) |
| RAW | TOA5 | Safe | does not already exist | RAW | as per uploaded | safe replacement (2) |
| RAW | TOA5 | Safe | same as file being replaced | RAW | as per uploaded | safe replacement (2) |
| RAW | TOA5 | Safe | already exists (but is not the file being replaced) | RAW | suffixed - see (1) | safe replacement (2), filename change (3) |
| RAW | TOA5 | Unsafe | does not already exist | ERROR | as per uploaded | bad overlap (4) |
| RAW | TOA5 | Unsafe | already exists | ERROR | suffixed - see (1) | bad overlap (4), filename change (3) |
| RAW | Non-TOA5 | N/A - only for TOA5 | does not already exist | RAW | as per uploaded | success |
| RAW | Non-TOA5 | N/A - only for TOA5 | already exists | RAW | suffixed - see (1) | filename change (3) |
| Not RAW | TOA5 | N/A - we don't check unless RAW | does not already exist | as specified | as per uploaded | success |
| Not RAW | Non-TOA5 | N/A - only for TOA5 | already exists | as specified | suffixed - see (1) | filename change (3) |

Notes:

1. suffixed means appending \_1 (or the next available number) - e.g. blah.dat becomes blah\_1.dat (or blah\_2.dat if blah\_1.dat already exists)
2. MESSAGE: The file replaced one or more other files with similar data. Replaced files: <filenames here>
3. MESSAGE: A file already existed with the same name. File has been renamed.
4. MESSAGE: File cannot safely replace existing files. File has been saved with type ERROR. Overlaps with <filenames here>
   1. Automating the upload of data to HIEv

As well as via the web interface, data can be uploaded to the HIEv system using an HTTP-based API. The upload of data into the system is facilitated through a Ruby script. This aspect of HIEv operation is beyond the scope of this manual. Instructions and a download for using this script can be found in the HIEv WiKI documentation on GitHUB at <https://github.com/IntersectAustralia/dc21/wiki/Setting-Up-Automated-Load-From-PC> %%% Check this URL

1. Managing Data Files

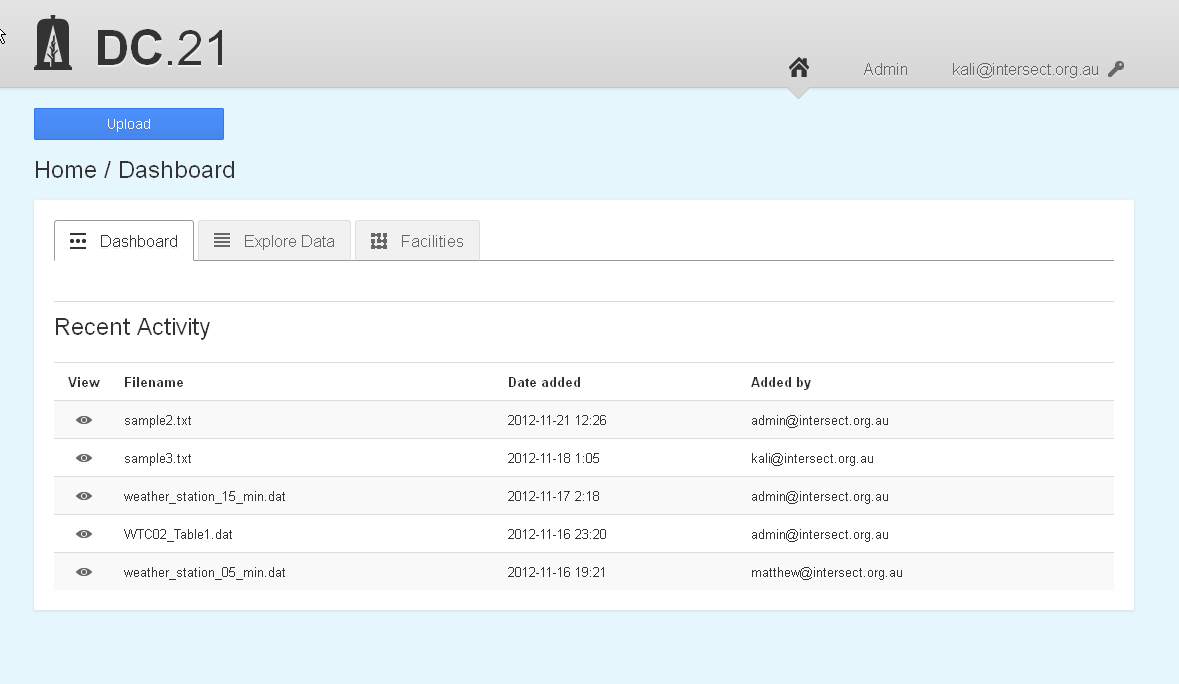
The key views of the files loaded into the HIEv database are the Dashboard and Explore Data views. These views allow you to perform the key functions of the HIEv system.

Access these views by clicking on their respective tabs on the HIEv Home Screen.

* 1. The Dashboard Tab

The default tab on the Home Screen is the Dashboard tab. It shows a list of the five files which have been uploaded or packaged most recently by the users of the HIEv system.

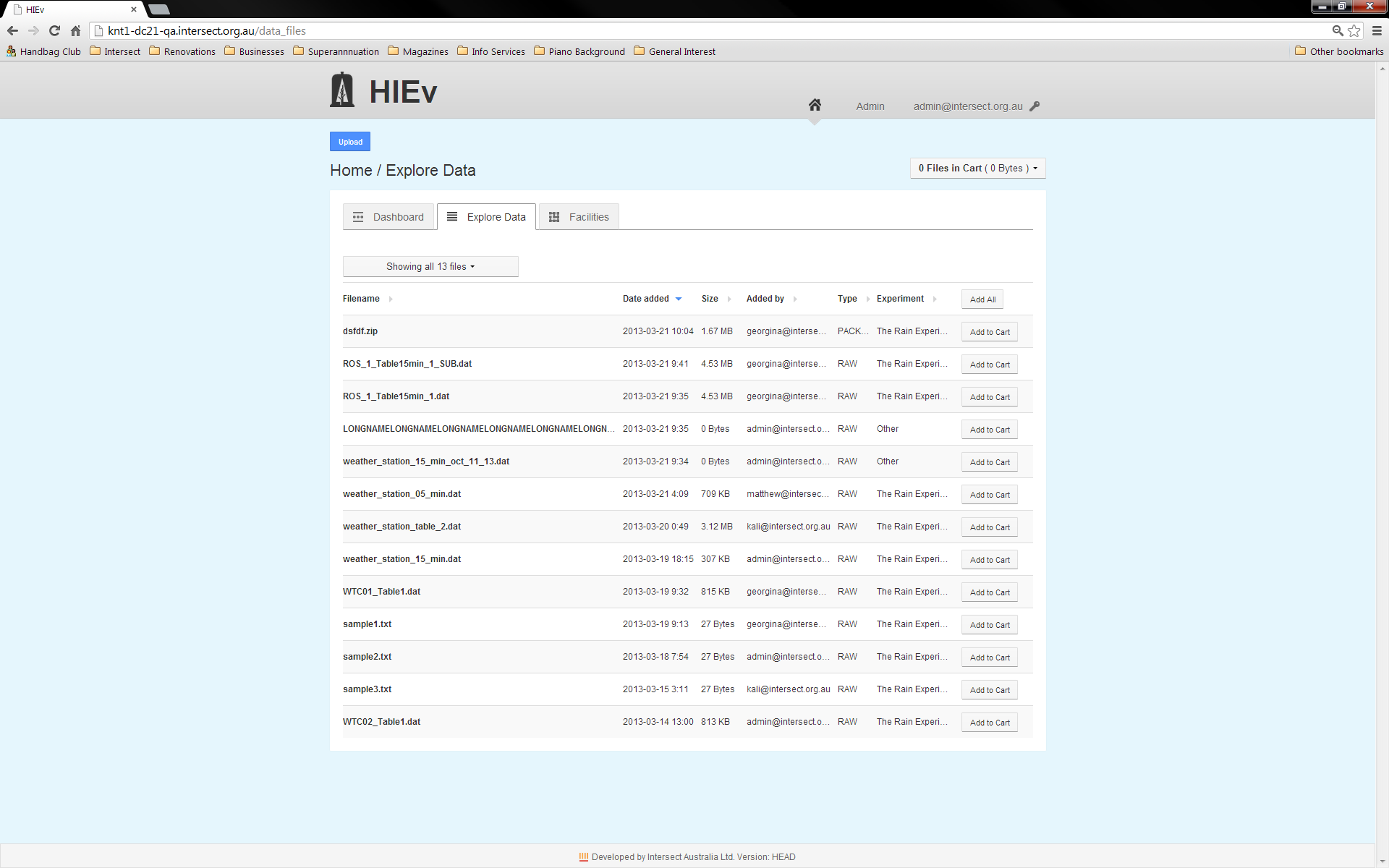
The operation of the Dashboard tab is similar to the operation of the Explore Data tab, except that searching and sorting functions are not supported. Therefore, please see the description of the Explore Data tab below for more information.



%%% Are there ever other sections besides “Recent Activity”? DC21-141(?) may change this. Check later.

* 1. The Explore Data Tab and File Searching

The Explore Data tab provides the main data management functions of the HIEv system. The initial view shows all data files which have been uploaded. If there are more files than fit on one screen, only the first 30 files will be shown, and the subsequent files can be shown by paging through the data using the page number buttons.



%%% Annotated screen dump, with more than 30 files in the database so page buttons show.

* + 1. Sorting

Click on the heading of any column in the file list in this view to sort the files into increasing order by that column. Click again to reverse the sort order.

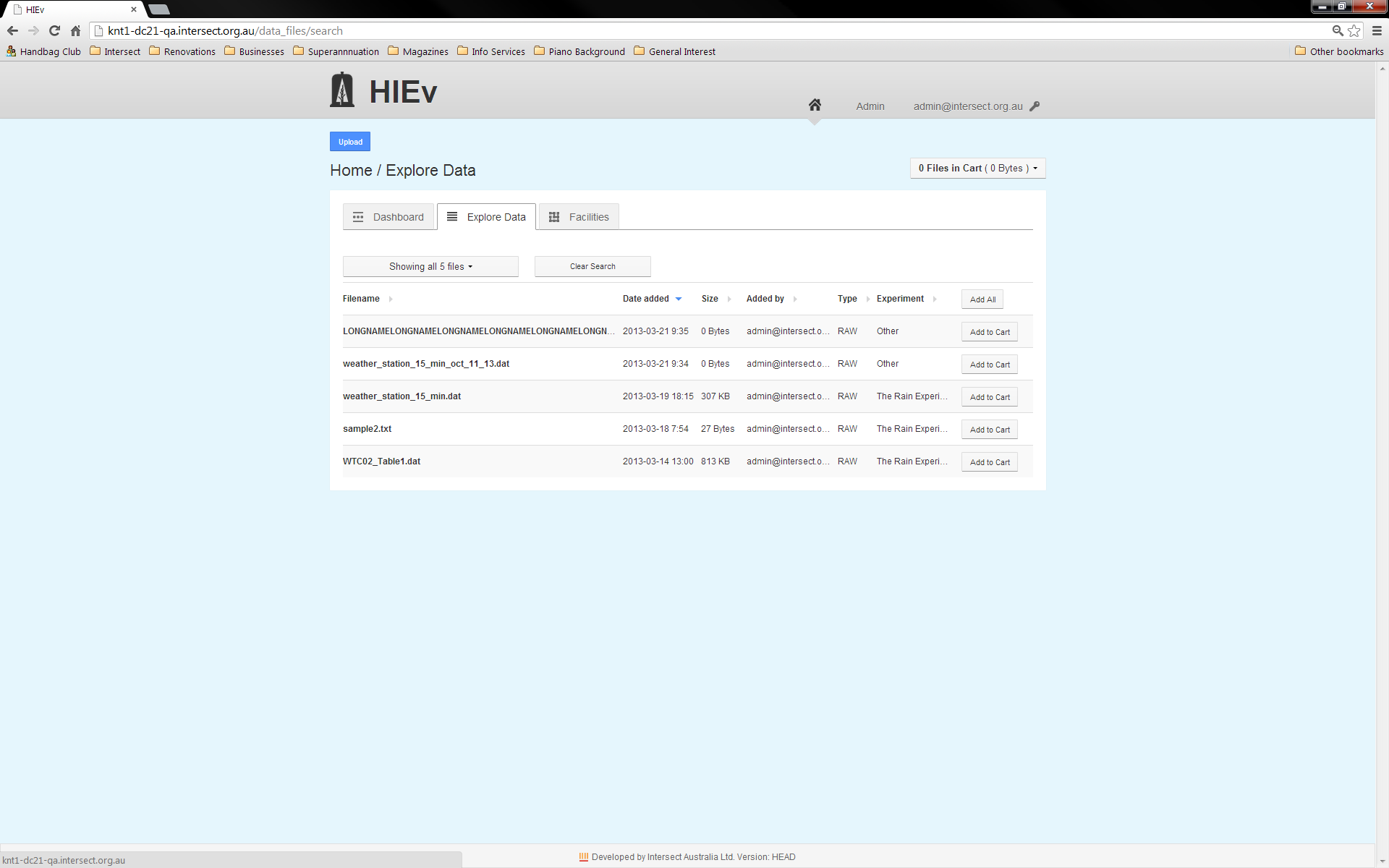
A triangular up or down arrow to right of any file list column heading indicates the active sort order. A grey right-pointing arrow indicates that the file list is not sorted by that column.

Re-sorting the data always resets the display to the first page of the file list.

Sorting can be done by only one file list column at a time.

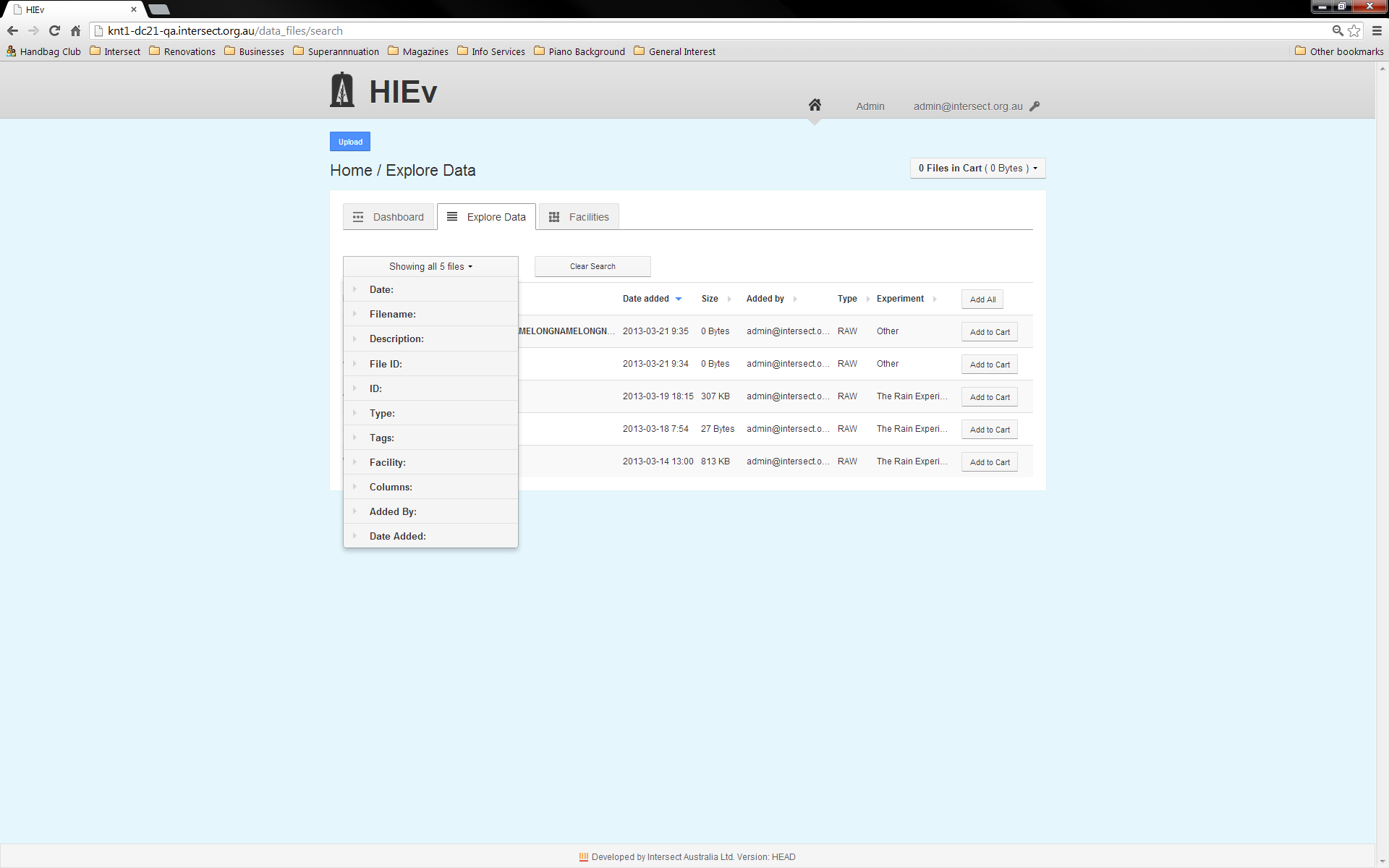
* + 1. Searching

When the number of files uploaded to your system becomes large, finding the file you are interested in may become difficult. The Search facility is provided to assist. It supports searching for data files using the metadata that was supplied at the time each file was uploaded.



The screen above shows the Explore Data tab when a search is active, restricting the number of files shown. Note the Clear Search button, which is only present when a search is active. Click this button to return to displaying all files.

To change the search conditions, click on the Showing ... files button, which will show a dropdown list of search parameters. The exact text of this button changes, depending on how many files are presently shown in the file list.



If there is a search active, one or more of the search categories in this menu will be expanded when you first display it. These are the search categories which have active search data.

Click on the metadata field you wish to search by to expand its display its search parameters.

If you specify more than one search condition using more than one metadata field, the file list will display only those files which satisfy all of the conditions you specify.

* + - 1. Regular Expressions

Regular expressions are used for searching for specific substrings in general text. They are used widely across many computer systems.

HIEv uses regular expressions to provide comprehensive search functionality for Filenames, Descriptions and IDs.

A few of the more useful functions of regular expressions are described briefly below. However, a comprehensive description of regular expressions is beyond the scope of this manual. Users can read a thorough description at <http://www.regular-expressions.info/reference.html>

HIEv’s use of regular expressions is not case sensitive. Therefore, you can enter either upper or lower case characters and get the same result.

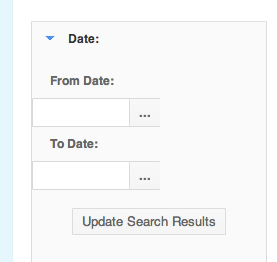
The following few examples are provided as a quick start to using regular expressions.

|  |  |  |  |
| --- | --- | --- | --- |
| Simple string |  | A simple string containing no special characters will match that string, regardless of where within the target string it occurs. | mpl will match the string example at the fourth character. |
| Start of string | ^ | The character ^ will match the beginning of the string. | ^exa will match example, but ^xa will not. |
| End of string | $ | The character $ will match the end of the string. | le$ will match example, but pl$ will not. |
| Any character | . | The period character will match any character. | a.c will match abc, aac, adc, a7c and a-c but it will not match ac or abbc. |
| Repeated character | \* | Asterisk causes the matching to zero or more repetitions of the previous character. | ab\*c will match ac, abc, abbc or abbbc, but will not match a7c or ahc. It will match aac at the second character and acc at the first character, because there are no characters between the a and c in those strings. |
| Repeated characters | + | The plus sign causes the matching to one or more repetitions of the previous character. | ab+c will match abc, abbc or abbbc but will not match ac. |
| Alternate characters | [ ] | Strings enclosed within square brackets will match any one of the characters within the brackets. | a[123]b will match a1b, a2b or a3b only. It will not match ab or any other substring. |
| Character ranges | [-] | Use – between [] to match one of range of characters. | [0-9] matches any digit.  [a-z] matches any letter.  [a-z0-9] matches any digit or letter. |
| Escape character | \ | In order to match a special character, precede it with the backslash character.  Special characters are [\^$.|?\*+(){}  Putting \ before other characters often has a special meaning, so should be avoided. | \\ will match \  \. will match .  \\* will match \*  \[ will match [ |
| Combinations |  | Any of the above search methods can be combined. | ^.c will match any string with c as its second character.  ^abc$ will match the string abc only. abcd or aabc will not be matched.  1[abcd]+2 will match any combination of the characters a, b, c or d which occur between the digits 1 and 2.  [\[\]] will match either [ or ].  \.+ will match any run of periods.  [0-9]+ will match any integer number.  [0-9]+\.[0-9]\* will match any number with a decimal point. |

If an invalid regular expression is entered, HIEv will place an error message at the top of the screen, clear the search field and ignore the regular expression.

* + - 1. Restricting by Data Date

The **Date** field allows you to search for files based on the start and end date specified in the file’s metadata.



A date can be entered in either the **From Date**, **To Date** or both. If only a **From Date** is specified, all files containing data for after that date will be included. If only a **To Date** is specified, all files containing data for before that date will be included.

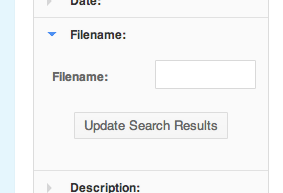
If you restrict by date, files which have no dates in their metadata will not be displayed.

Dates can be typed into the **From Date** and **To Date** fields in *YYYY-MM-DD* format or by clicking on the ellipsis to the right and selecting a day from the calendar that is displayed.

For TOA5 files, this search option checks the Start and End Dates in the Information from the File (see section ) and for all other files, this search option checks the Basic metadata Information (see section ).

* + - 1. Restricting by Filename substring

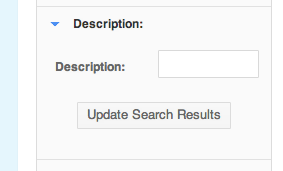
The **Filename** field allows you to search for files based on their filename:



HIEv treats the search string as a regular expression. See section for more information.

* + - 1. Restricting by Description substring

The **Description** field allows you to search for files based on their free-form text descriptions:



HIEv treats the search string as a regular expression. See section for more information.

Use the search string ^$ to search for files without any Description.

* + - 1. Restricting by File ID

Enter an integer number into the Field ID search field to display only the file with that File ID. It is not usual to use this search method with any other search method, as this method will always display exactly one file, or no files if there is no file with the entered File ID.

%%% Screen dump

* + - 1. Restricting by ID

The **ID** field allows you to search for files based on their entered ID.

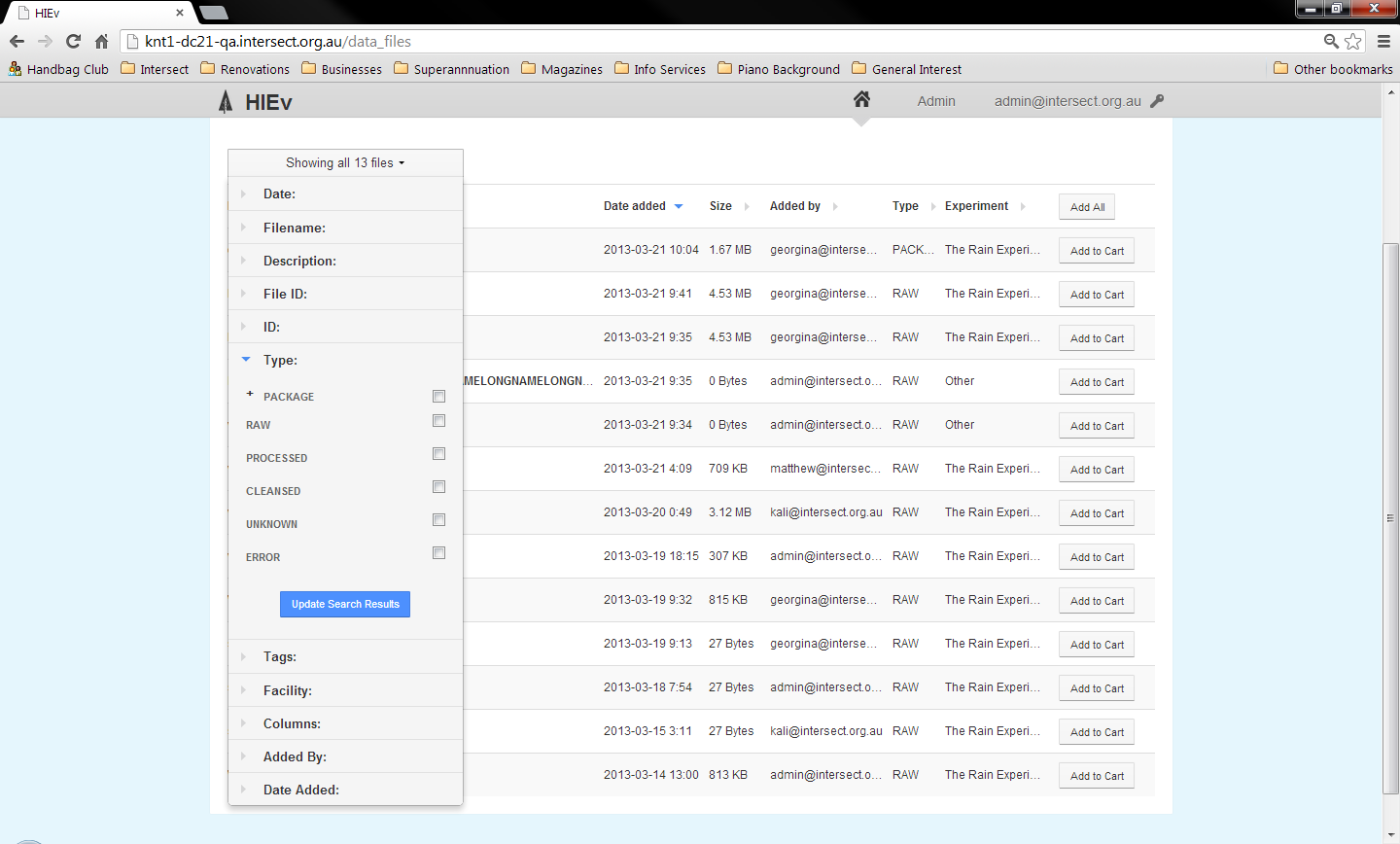
%%% Screen dump

HIEv treats the search string as a regular expression. See section for more information.

Use the search string ^$ to search for files without any ID, or ^ \*$ to search for files with an emplty ID or one consisting of just spaces.

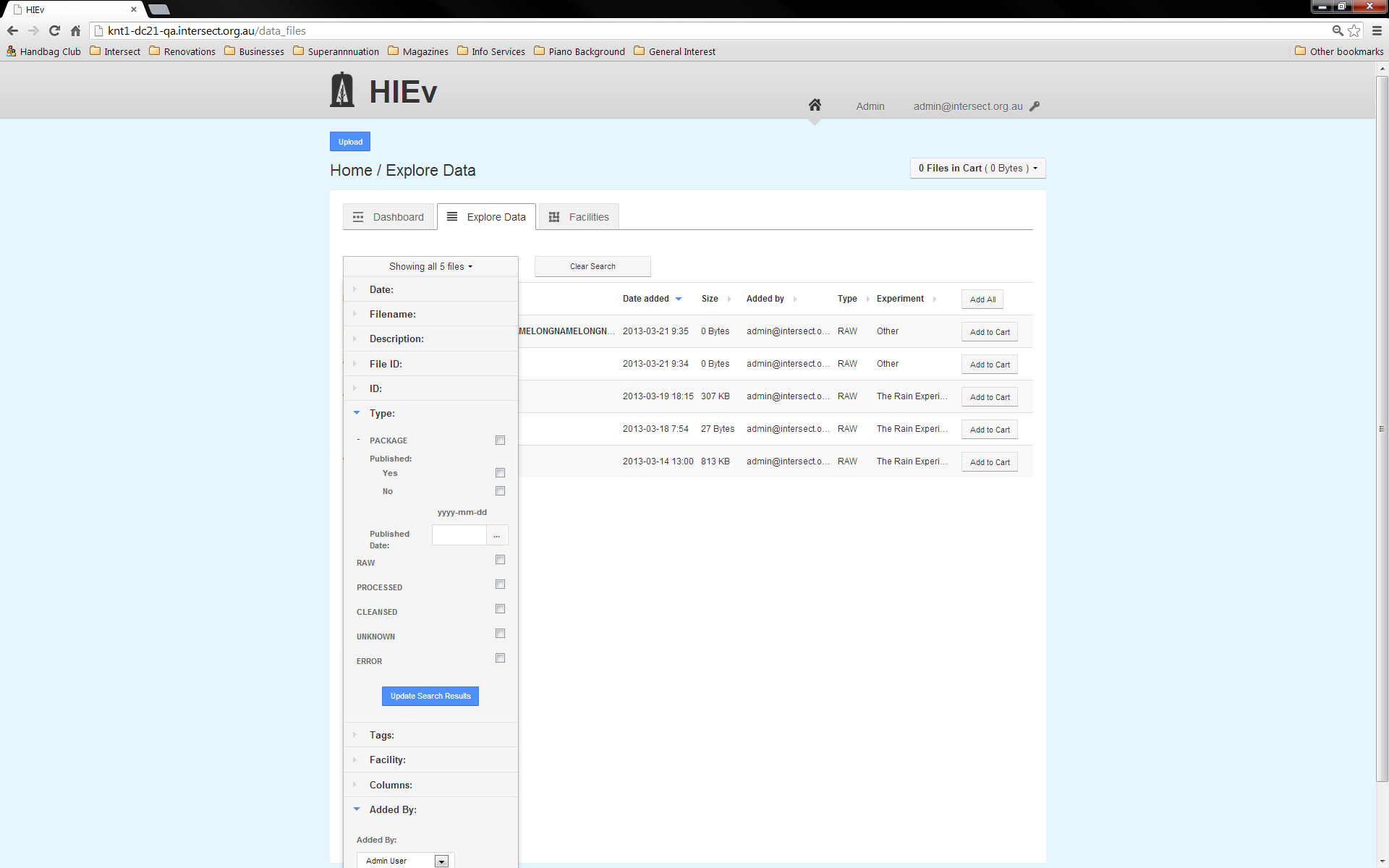
* + - 1. Restricting by File Type

The **Type** search parameters allows you to search for files based on their specified type.



The set of possible types is displayed as a list of checkboxes. Selecting none of the checkboxes is the same as selecting them all - files will not be filtered based on their type. Once at least one checkbox has been selected, only files of that type will be returned in the search results. More than one type can be selected.

Click on the + sign to the left of the Package option to open further search conditions for Packages. (Clicking on the word Package sets its checkbox.) You can click on the minus sign to close it again.

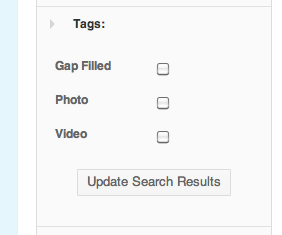


Select the Yes or No checkbox to display only files which are, or are not, Published. To show a Package file regardless of its Publish status, leave both checkboxes unchecked. If a Published Date is entered, then only files Published on that date will be displayed. Again, click on the ellipses to the right of the Published Date box to select a date from a calendar.

%%% The operation of this package selection functionality is buggy, and it’s not clear if it will be fixed prior to end of Sprint 8. Recheck after sprint 8. DC21-525

* + - 1. Restricting by Tags

The **Tags** interface allows you to search on the tags that have been assigned to a file.

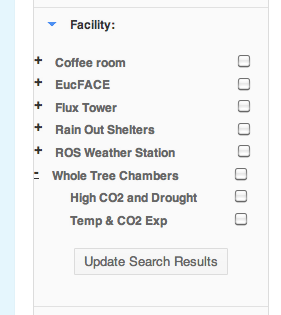


Like the Type interface, selecting none of the available checkboxes means that files will be returned in the search results regardless of the tags they have. Once a checkbox is selected, only files that have the corresponding tag will listed. More than one tag can be selected at any given time.

It is not possible to search for files which do not have a specific tag.

* + - 1. Restricting by Facility

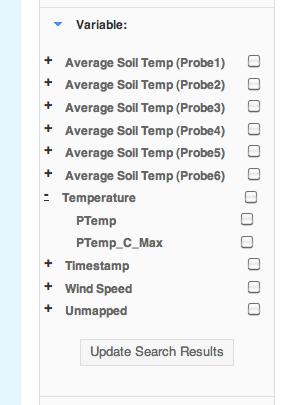
The **Facility** interface allows you to search for files based on the facility or experiment that produced the file.



The interface is a two-level hierarchy of checkboxes. The top level shows all the facilities defined in the system and the second level shows the experiments that are defined for those facilities. Selecting a facility selects all of the experiments for that facility. If only specific experiments are required, clicking on the plus sign to the left of a facility will expand the hierarchy and allow individual experiments to be selected or deselected.

* + - 1. Restricting by Data File Columns

The **Columns** interface allows you to search for TOA5 format data files that contain specified columns.



Like the Facility interface, this shows a two-level hierarchy of checkboxes. The top level contains all the Names from that are used in the Columns Mapping table. The last top level group is an extra group called **Unmapped** that contains all the TOA5 column headings that are not mapped to a standard Name in the Column Mappings table. See for more information.

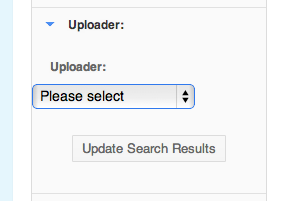
Selecting the checkbox for a standardised top-level Name will select all the TOA5 column headings that are mapped to it. Clicking the plus sign to the left of the top-level Name will show all the TOA5 column headings mapped to it and allow you to select them individually.

Selecting more than one checkbox will cause any file which has any one or more of those corresponding TOA5 column headings to be listed.

If you set any checkbox in this search function, only TOA5 data files will be listed.

* + - 1. Restricting by person who added the file

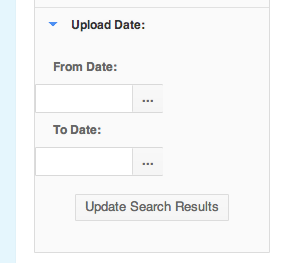
The **Added By** field allows you to search for files that were uploaded by a specific user:



The user must be selected from the list of a users registered in the system.

* + - 1. Restricting by Upload Date

The **Date Added** interface allows you to search for files that were uploaded between a specified set of dates:



Like the **Date** field, the **Date Added** interface allows you to select a **From date** and a **To Date**. If both dates are specified only files uploaded between those dates will be included in the search results. If only a **From Date** is specified, all files uploaded after that date will be included. If only a **To Date** is specified, all uploaded before that date will be included.

* 1. The Cart

The Cart operates like an e-Commerce shopping cart. HIEv provides functions for adding files to the Cart and for doing operations, such as Downloading and Publishing, on all files in the Cart. See more information about Downloading in Chapter and Publishing in Chapter .

The content of the Cart persists between login sessions.

If a file is deleted from the system or is replaced by a new upload, that file will disappear from all users’ Carts.

Note There is one Cart per user account. If two people simultaneously use the same login account, their operations on the Cart will interfere with one another.

Add a file to the Cart by clicking on any Add to Cart button for that file. There are Add to Cart buttons in multiple places, including the Dashboard file list and the Explore Data file lists.

The Cart status box shows the number of files in your Cart and the total size of these files. Click on this Cart status box, which appears at the top right of many HIEv screens, to show a dropdown menu of operations which can be performed on the files in the Cart. These operations are:

|  |  |
| --- | --- |
| Download | Click on this option to download data files to your local computer. See Chapter for instructions on using this feature. |
| Package | Click on this option to create a publishable Package containing all files in the Cart. See section for instructions on using this feature. |
| Clear cart | Click on this option to remove all files from the Cart. It does not delete the files themselves. |
| Edit cart | Click on this option to view a list of the Cart contents, remove individual files from the Cart, download all files in the Cart or create a publishable Package using all files in the Cart. See section for more details. |

* + 1. Editing the Cart Contents

Selecting the Edit Cart option will display the following screen:

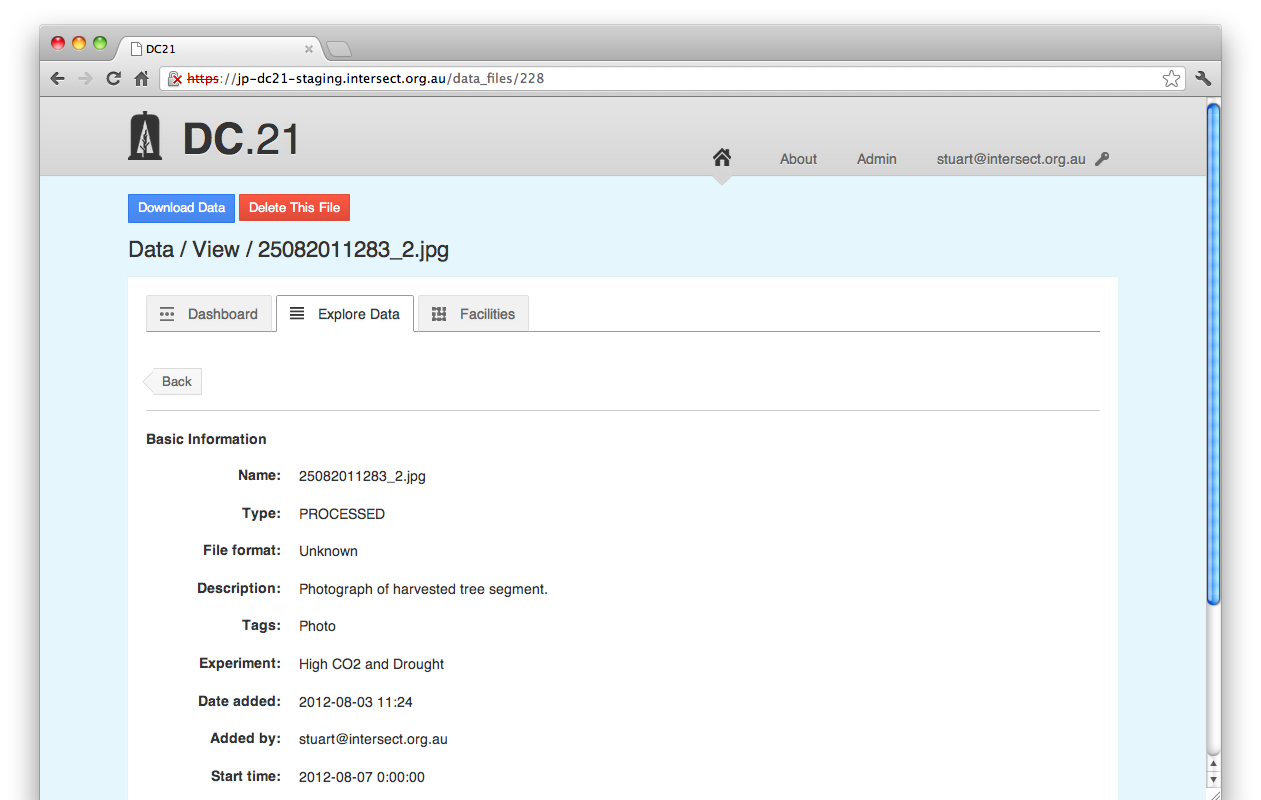
%%% Screen dump

There is one line in the table on this screen for each file in your Cart. These files can be removed individually by clicking on the Remove button on the relevant line. If the Remove All button is clicked, the Cart will be emptied and further Cart operations cannot be performed until further files are again added to the Cart.

Package and Download buttons are also available on this screen. See section and Chapter for more information about the operations initiated by these buttons.

* 1. Viewing and Editing a File's Metadata

Clicking on any filename in the Filename column of the Dashboard tab, Explore Data tab or Edit Cart view will display the metadata for that file in a screen similar to the following.



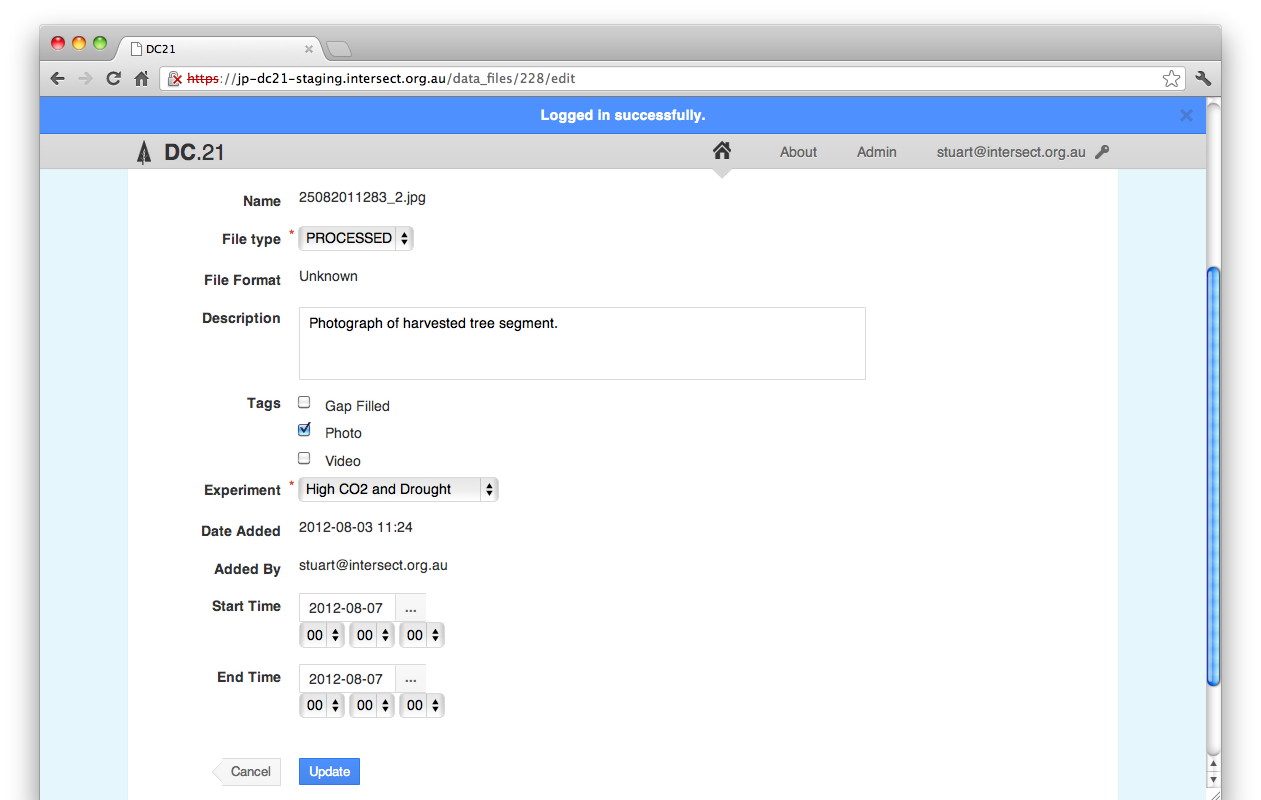
Files of all types will display the Basic Information section showing the Basic metadata, although some fields are not shown for some file types. TOA5 files also show **Information From The File** and **Columns** sections. See Chapter for more information.

There is a button which says Add to Cart. Click this button to add this file to your Cart. If the file is already in your Cart, then the button will change to say Remove from Cart, and you can use it to remove the file from your Cart.

At the bottom of the screen is a button to edit the file's metadata.

%%% Screen dump, if it doesn’t show on the screen dump above.

This button will take you to a form that allows you to modify the file's metadata.



Once you have finished editing the metadata, click the "Update" button to save your changes. If all your changes are valid, the metadata will be updated and you will be returned to the metadata view.

Note You will only have permission to edit the metadata of this file if you are logged in with the user credentials which were used when the file was uploaded, or you have Administrator permission.

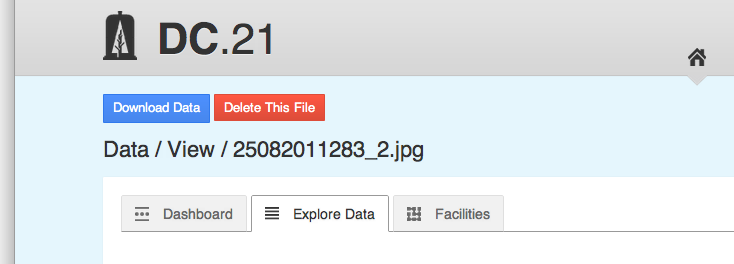
* 1. Deleting a Data File

Note Deleting files removes the file completely from the HIEv system. It is irreversible.

In order to delete a file from the HIEv database, first view the metadata for that file by clicking on the file name in either the Dashboard or Explore Data views. If you have permission to delete that file, a red Delete This File button will appear at the top of the Metadata View screen.

Note You will only have permission to delete this file if you are logged in with the user credentials which were used when the file was uploaded, or you have Administrator permission.

You cannot delete a file which has a non-null ID. In order to delete it, you must clear the ID first. This functionality makes it more difficult to inadvertently delete a Package which has been Published and has had its Published ID entered.



When you click on the Delete This File button, you will be asked to confirm that you do wish to delete the file.

1. Publishing Your Data

Once an Experiment is complete, the data collected can be Published to the [Australian Research Data Commons](http://www.ands.org.au/about/approach.html#ardc).

Before Publishing, the data files to be Published must be combined together into a single Package file. Package files are ZIP files which use the Bagit format, which is described in .

Once created, Package files are shown sorted with the other data files in the file list which is displayed on the Explore Data tab.

Exactly which files should be included in a Package intended for publishing is largely dependent on what is meaningful for the data and research discipline in question. It is entirely valid to have the same data appear in multiple Package files if that will aid discovery and reuse for other researchers. For example, a large set of data could be divided into two smaller, but overlapping, sets of data that represent different lenses (research problems) that the data could be seen through.

Once a Package has been published the metadata describing the Package will be made available for harvesting by the UWS Metadata Store. After this has occurred and the Package has been approved, it will become discoverable in [Research Data Australia](http://researchdata.ands.org.au/).

The process of Publishing involves a few steps:

* The user creates a Package which contains the data files to be Published using the HIEv Publish function. This step copies the selected data files’ metadata into a RIF-CS file and copies the data files themselves into a Package ZIP file. These two files are logically linked together.
* The user Publishes the Package using HIEv’s Publish function. This copies the RIF-CS file into a location so that it can be harvested by the OAI-PHM harvester. It also sets the Published flag and Published Date field in the ZIP file’s metadata.
* At some subsequent time, the OAI-PHM harvester will discover the RIF-CS file. The harvester copies the RIF-CS file and the Packaged ZIP file it refers to into the Published data store. There is no indication in the Package’s HIEv metadata when or if this has occurred.

Note See the section Adding a Package’s External ID

Once a Package has been Published and harvested, an ID is generated for it by the external Published data store.

To edit a Package’s metadata:

* Navigate to the Package file using the Explore Data tab.
* Click on the Package file’s filename to open the metadata view screen.
* Click on the Edit Metadata button to open the metadata edit screen.

%%% Edit Package Metadata screen dump

* Enter the new ID into the metadata ID field.
* Click Update to save the change.

See section 7.4 Viewing and Editing a File's Metadata for more information on using this edit function.

Note Editing most fields of a Package’s metadata is not recommended. It does not modify the data in the matching RIF-CS file. Please see section 8.3 **Error! Not a valid bookmark self-reference.** for more information about restrictions on modifying Packages’ metadata.

Managing Published Packages for important information about managing this process.

* 1. Creating a Package

Creating a Package creates two related components:

|  |  |
| --- | --- |
| ZIP Bagit File | This ZIP file contains a snapshot of all the data for this Package. This includes copies of the data files and a Readme.HTML file which contains a copy of the all of the data files’ metadata. |
| Matching RIF-CS File | This file contains a copy of the Package metadata which is entered at the time the Package is created. |

To create a Package containing one or more files:

* Add the required file of files to your Cart, ensuring that the Cart contains only those files you wish to include in your Package. See Chapter for instructions on using the HIEv Cart.
* Click on the Cart status box to open the Cart dropdown menu.
* Select Package from the dropdown menu. Alternatively, selected Edit Cart from the dropdown menu and then click on the Package button on that screen, as shown below. If this method is used, the Cart can be reviewed prior to Packaging.
* The New Package screen will be displayed. Enter the metadata associated with your Package. See below for details.
* Click on Save to cause your Package file to be created and saved. The Package file can now be viewed in the Explore Data tab. If you click on Back, you will be returned to the Explore Data tab and the Package will not be created.

%%% New Package screen dump.

Note It is very important to check this metadata closely before clicking Save. This metadata is copied into the RIF-CS file immediately after pressing this button. The data in the RIF-CS file cannot be edited. If it is wrong, the Package must be deleted and re-created.

It is possible to create a Package which contains other Packages. There may be circumstances when this is meaningful.

* 1. Publishing a Package

When a Package has been created and its metadata is correct, it can be Published.

To Publish a Package:

* View the Package’s metadata by clicking on its filename on either the Dashboard tab or Explore Data tab. Review its metadata to ensure you have selected the correct Package file and that it is ready to Publish.

%%% Metadata view screen dump with Publish button highlighted.

* Click on the Publish action button at the top of the metadata screen. This Publish button will not appear if the Package has already been Published. A Package cannot be Published twice.
* A dialog box is shown for you to indicate that you are sure you wish to proceed. Click on OK. If you click on the Cancel button, you will returned to the Package metadata screen.
  1. Adding a Package’s External ID

Once a Package has been Published and harvested, an ID is generated for it by the external Published data store.

To edit a Package’s metadata:

* Navigate to the Package file using the Explore Data tab.
* Click on the Package file’s filename to open the metadata view screen.
* Click on the Edit Metadata button to open the metadata edit screen.

%%% Edit Package Metadata screen dump

* Enter the new ID into the metadata ID field.
* Click Update to save the change.

See section for more information on using this edit function.

Note Editing most fields of a Package’s metadata is not recommended. It does not modify the data in the matching RIF-CS file. Please see section **Error! Not a valid bookmark self-reference.** for more information about restrictions on modifying Packages’ metadata.

* 1. Managing Published Packages

Note Considerable care must be taken when managing Published Packages. You should have a thorough understanding of the way the harvesting and storage of Published Packages is configured on your system before using the functions in this section.

Once Packages are Published using the HIEv Publish function, they are available for harvesting, but they are not necessarily harvested promptly. Depending on how your system is configured, it may take some time for them to be harvested, perhaps even days. In addition, you cannot tell from HIEv if the Package has already been harvested or not.

* + 1. Publishing a second time

HIEv prevents you Publishing a Package more than once. However, Publishing a second time is possible if a new package with the same data is created. Such Publishing is generally harmless, even if the originally harvested version has already been harvested. The exact results will depend on the way your system is configured.

* + 1. Deleting Published Packages

It is possible to delete any Package file that you have created. (If you have administration privileges, you can also delete the Package files created by others.) However, Packages with their metadata ID field set to a non-empty value cannot be directly deleted. To delete them, first clear the ID field. This helps prevent inadvertent deletion of Published Packages.

Before harvesting, if you delete a Package using the HIEv file delete function, its RIF-CS file will also be deleted. This effectively undoes the Publish function and the Package will never be harvested.

After harvesting, if you delete a Package, it will not affect any already harvested version of this Package.

* + 1. Editing Published Packages

As explained above, at the time a Package is created its matching RIF-CS file is also created. (The action of Publishing only copies that RIF-CS file to a discoverable location for harvesting.)

Therefore, even though editing the metadata of a Package is possible, it will not alter the metadata already stored in the RIF-CS file. For this reason, editing a Package’s metadata is not recommended.

The only effective way to change the metadata of a Package is to delete the Package and recreate it. Prior to Publishing, this is always a safe and reliable way to update a Package’s metadata.

After Publishing, the effect of Package deletion depends on whether harvesting has already occurred (see previous section).

* + 1. Correcting Published Packages

If a Package is Published incorrectly and has not been harvested, deleting it is sufficient to avoid the Publishing of the incorrect data.

If it has been harvested, deleting it will have no downstream effect. It will be necessary for you to contact the administrator of your Published data store and ask for its removal.

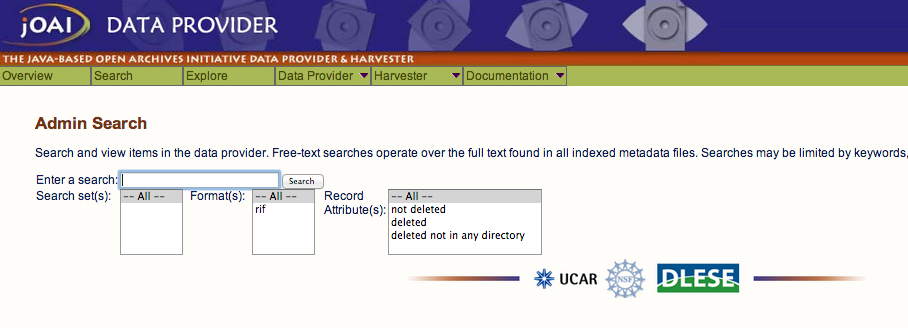
If the metadata of a Published Package is not correct, you can delete the Package, recreated it with correct metadata and re-Publish. If it has already been harvested, some systems may correctly overwrite the old data with the new. However, it is generally best to contact the administrator of your data store and ask for the removal of the original package.

* 1. Viewing Published data

%%% PeterB suggests it may be better to remove this section, although there’s presently no other way. There is a chance of a RIF-CS view option in a later sprint. All HIEv applications must use jOAI if they implement Publishing.

HIEv implementations use the Java Open Archives Initiative (jOAI) to harvest Published data. This tool can be configured in many ways, so it is best to understand its operation specific to your site.

The descriptions of published Packages can be viewed by going to the jOAI web interface at **http://<***your.HIEv.sever>***/oai/admin/query.do** and performing a search:



Clicking the search button with the search field blank will show all published Packages:



1. Downloading files

HIEv allows you to download any data file, or multiple data files, to your local computer.

If you download a single file, it will be saved on your computer in its usual format. If you simultaneously download more than one file, the files will be combined into a ZIP file and that ZIP file will be downloaded to your computer. (This ZIP file is not a Bagit format ZIP file.)

To download one or more files:

* Add those files to your Cart, ensuring that the Cart contains only those files you wish to download. See section Chapter for instructions on how to add data files to your Cart.
* Click on the Cart status box to open the Cart dropdown menu.
* Select Download from the dropdown menu. Alternatively, select Edit Cart from the dropdown menu and then click on the Package button on that screen, as shown below. If this method is used, the Cart can be reviewed prior to Packaging.
* A file dialog box will open. Navigate to the sub-directory into which you wish to save the downloaded data and select the name you wish to use for the downloaded data file. The file dialog will be for the one data file if only one file is in your Cart, or it will be for one ZIP file if multiple files are in your Cart.

When downloading data files, only the data files themselves are downloaded. Metadata is not downloaded.

If you download a Packaged ZIP file, you can access the metadata for each of the files, which is included in that Packaged ZIP file. See section for instructions on creating a Packaged ZIP file and for details of the Bagit format, which is used for Packaged ZIP files.

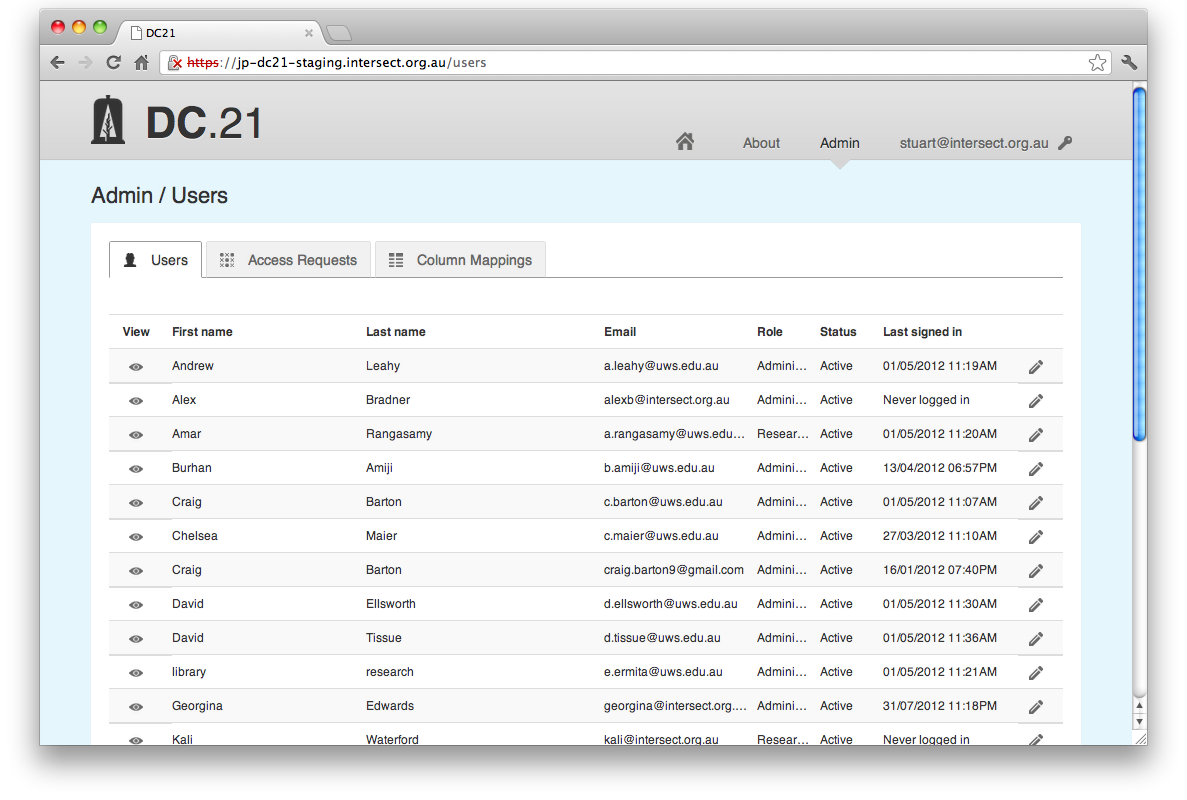
1. System Administration

When a user is created they are given a role within the HIEv system. This role dictates what permissions the user has within the system. The most powerful role a user can be given within the system is that of the Administrator.

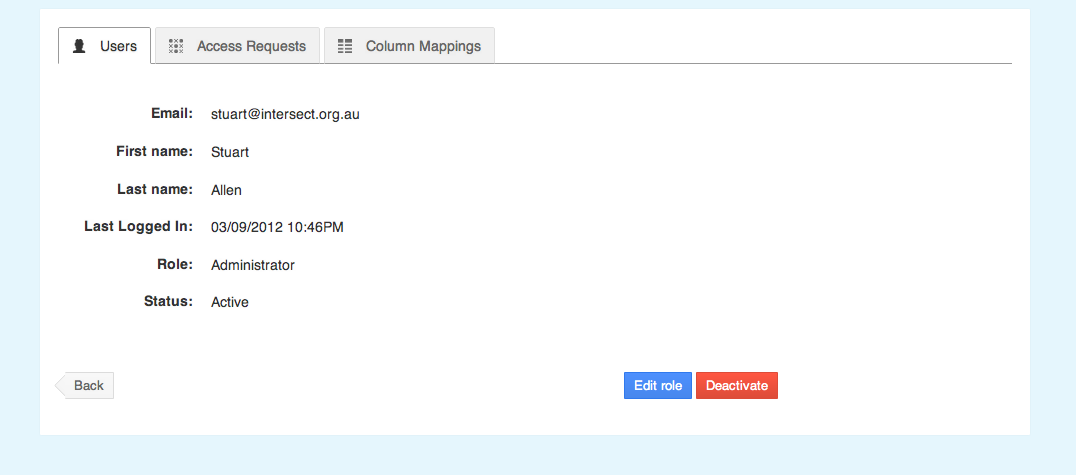
All Administrators have access to the Admin section accessed from the **Admin** link at the top right of the screen (only visible for this role). The Admin section has three tabs: the Users and Access Requests tabs are used for managing the details of users who can access the system and Column Mappings tab is one method of managing the list of defined Column Mappings.

* 1. Managing Users’ Details

The **Users** tab lists all the users that are registered within the system:

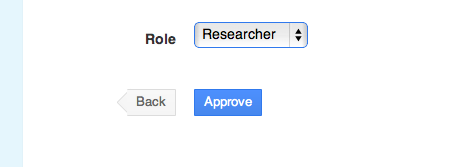


Click on a User’s email address to open a screen showing that User’s details.



Two functions are provided:

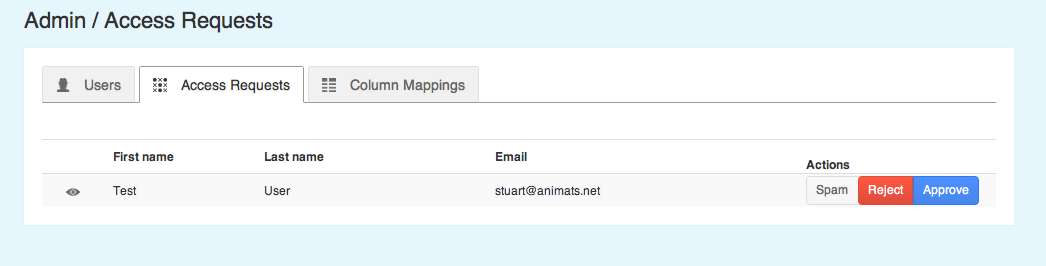
|  |  |
| --- | --- |
| **Deactivate** | Disable the account from being used to login to the system. No data uploaded by the user will be deleted. |
| **Edit role** | Change the role that will be assigned to the user for future logins. Clicking this button opens the following screen that allows the User’s role to be changed. See section for information on roles. |



It is not possible to delete a User’s entry. This ensures that historical information relating to that User remains meaningful. Instead of deletion, a User’s login account should be Deactivated.

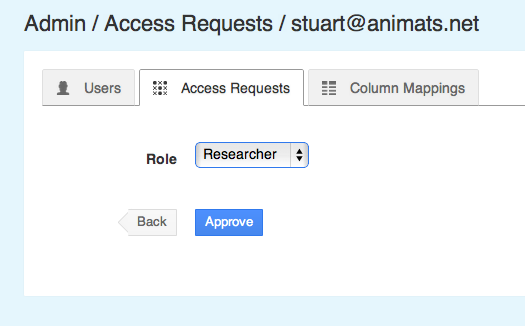
* 1. Authorising New Users – The Access Requests Tab

The **Access Requests** tab is where an administrator can approve or deny requests for a user account in the system:



Each access request line in this table has three buttons:

|  |  |
| --- | --- |
| **Spam** | Click on this button to ignore the account request and remove it totally from the system. |
| **Reject** | Click on this button to reject the access request and send the User an email informing the user that his or her request for an account has been rejected. |
| **Approve** | Click on this button to accept the User’s access request. It will take you to a screen where you must select a role for the user in the system. Completing this approval process results in a confirmation email being sent to the User. See section for information on the permissions of the three available roles. |



Clicking any of these three buttons will remove the request from the Access Requests table.

* 1. Managing Column Mappings

Column Mappings are a way of defining a relationship between the column headings in TOA5 data files (the "Code" part of the mapping) to a standard name from a defined ontology (the "Name" part of the mapping.)

The Column Mappings are stored once for the whole system and all users share the one set of mappings.

A basic set of Column Mappings is defined as part of the configuration of the HIEv System at installation. In addition, further mappings can be added as they are needed.

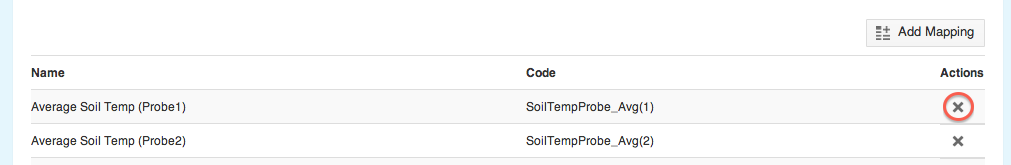
The Column Mappings can be added or modified only by users with Administration permission.

These Columns Mappings are used by HIEv for two purposes:

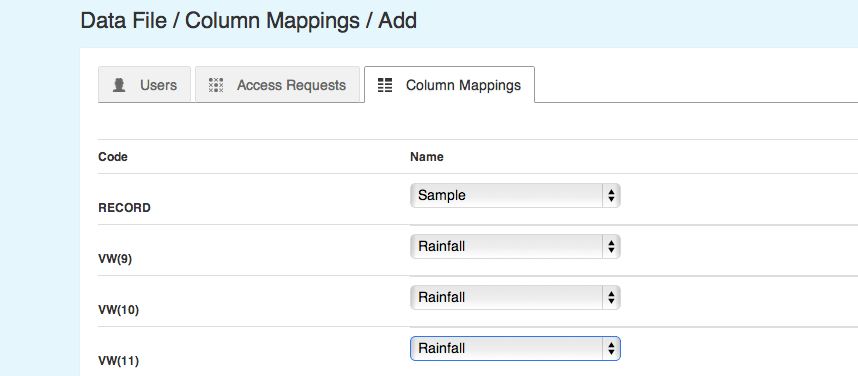
* When the Metadata for a TOA5 file is displayed, the Column Mappings table is checked against the column names in the TOA5. If any match is found, the Name from any matching entry in the Columns Mapping table is shown in the metadata display in the Name column of the Columns information.
* When searching for data files by Columns, the options in the search parameters are the column headings from all TOA5 data files stored in the HIEv system. They are grouped and sorted by their matching Name from this table. Those without matching Names are shown in an Unmapped list as the final item.
  + 1. The Column Mappings tab

This tab allows users with administrator permissions to add and delete column mappings.

To **delete** an existing mapping, click the cross in the far right **Actions** column of the table for the mapping you wish to delete:

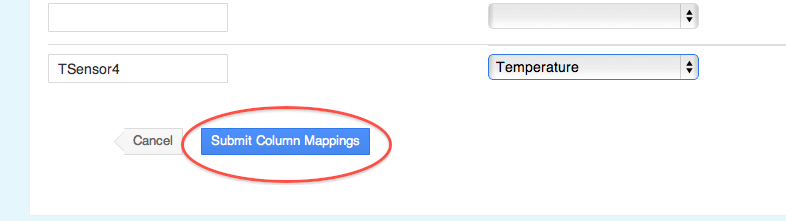


To **add** more mappings click the **Add Mappings** button at the top left of the tab. This will display a form where the mapping pairs can be defined.



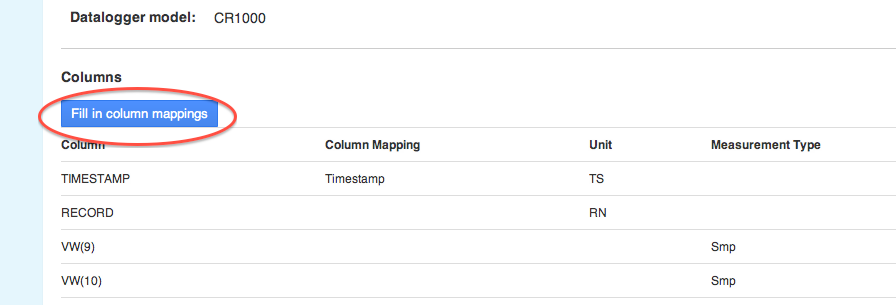
On the left of each row the code from a TOA5 column header can be entered and on the right the standard Name to map to can be selected from a drop-down list.

Once the mappings are defined, click the **Submit Column Mappings** button:

  
%%% Maybe this should be part of the screen dump above so that users are clear that it’s all on the one screen.

* + 1. Updating from the Explore Data tab

Column mappings may also be defined using the **Fill In Column Mappings** button on the **View Metadata** page for any TOA5 file. This button is only available to Users with Administrator permission. %%% This may not be true due to a bug. Or if it’s intended behavior, at least non-admin users can’t use the other tabs on this screen, so it would be OK-ish. Confirm what behavior will be in the final Sprint 8 release.



This is the preferred method for updating Column Mappings, as it avoids the need to manually type the column headers into the "Code" fields:

%%% Screen dump

The list of unmapped fields in this TOA5 data file is displayed in teh left hand column and dropdown boxes are shown in the right hand column. Use the dropdown boxes to set the equivalent names for any of the column headings shown. Click on Submit Column Mappings to cause these new mappings to be added to the system-wide Column Mapping table. You will observe that these new mappings are now listed in the Column Mapping column of the Columns section of the metadata display for that TOA5 file.

1. Configuring Tags, Column Mappings and Experiment Parameters

When HIEv is first installed, the available Tags, Column Mappings and Experiment Parameters are populated in the database from the configuration file: dc21app\_extra\_config.yml %%% Check file name. Once the system has been installed, the experiment parameters and tags can be modified by modifying the lookup tables that store these values. This can be done directly using a tool like PSQL. Another convenient way to do this is to use the Rails console. The instructions below use the Rails console to add rows to the tables.

If you wish to delete or modify existing rows in these tables, make sure you maintain referential integrity with existing records.

To begin, ssh to the server the HIEv system is running on. Once connected, the first step is to determine the Rails Environment the system is running as. This will generally be “production”, but you can check by looking for a RailsEnv line in the Apache Rails configuration (eg. <root>/etc/httpd/conf.d/rails\_dc21app.conf %%% Check this file name). This will look something like the following:

RailsEnv production

In this case, the Rails Environment is "production". Once the value of this setting has been determined for you system, navigate to the location of the application installation (e.g <root>/home/devel/dc21app/current %%% Check this file name) and enter the directory "dc21app/current" %%% Check this file name. From here you can start the Rails Console using the command:

RAILS\_ENV=<RailsEnv> bundle exec rails console

e.g.

RAILS\_ENV= production bundle exec rails console

This will give you a prompt similar to:

Loading production environment (Rails 3.1.1)

1.9.2p290 :001 >

From this prompt you can issue commands to add Tags and Experiment Parameters.

To add a Tag use the command:

Tag.create!(name: '<Tag name>')

e.g.

Tag.create!(name: 'Analysed')

This will result in output similar to:

(0.3ms) BEGIN

(1.3ms) SELECT 1 FROM "tags" WHERE LOWER("tags"."name") = LOWER('Analysed') LIMIT 1

SQL (8.7ms) INSERT INTO "tags" ("created\_at", "name", "updated\_at") VALUES ($1, $2, $3) RETURNING "id" [["created\_at", Fri, 14 Sep 2012 10:55:24 EST +10:00], ["name", "Analysed"], ["updated\_at", Fri, 14 Sep 2012 10:55:24 EST +10:00]]

(0.6ms) COMMIT

=> #<Tag id: 6, name: "Analysed", created\_at: "2012-09-14 00:55:24", updated\_at: "2012-09-14 00:55:24">

To add a Modification or a Unit for an Experiment Parameter, use the commands:

ParameterModification.create!(name: 'Above average')

ParameterUnit.create!(name: 'PSI')

Parameter Categories and Sub Categories require an extra step to define the relationship between the two:

parameter\_category = ParameterCategory.create(name: 'Light')

parameter\_category.parameter\_sub\_categories <<

ParameterSubCategory.create(name: 'Brightness')

This will result in output similar to:

1.9.2p290 :001 > parameter\_category = ParameterCategory.create(name: 'Light')

(0.4ms) BEGIN

SQL (121.5ms) INSERT INTO "parameter\_categories" ("created\_at", "name", "updated\_at") VALUES ($1, $2, $3) RETURNING "id" [["created\_at", Fri, 14 Sep 2012 16:14:26 EST +10:00], ["name", "Light"], ["updated\_at", Fri, 14 Sep 2012 16:14:26 EST +10:00]]

(0.5ms) COMMIT

=> #<ParameterCategory id: 8, name: "Light", created\_at: "2012-09-14 06:14:26", updated\_at: "2012-09-14 06:14:26">

1.9.2p290 :002 > parameter\_category.parameter\_sub\_categories <<

1.9.2p290 :003 > ParameterSubCategory.create(name: 'Brightness')

(0.3ms) BEGIN

(0.3ms) ROLLBACK

(0.2ms) BEGIN

ParameterCategory Load (0.7ms) SELECT "parameter\_categories".\* FROM "parameter\_categories" WHERE "parameter\_categories"."id" = 8 LIMIT 1

SQL (1.2ms) INSERT INTO "parameter\_sub\_categories" ("created\_at", "name", "parameter\_category\_id", "updated\_at") VALUES ($1, $2, $3, $4) RETURNING "id" [["created\_at", Fri, 14 Sep 2012 16:14:27 EST +10:00], ["name", "Brightness"], ["parameter\_category\_id", 8], ["updated\_at", Fri, 14 Sep 2012 16:14:27 EST +10:00]]

(0.5ms) COMMIT

ParameterSubCategory Load (0.7ms) SELECT "parameter\_sub\_categories".\* FROM "parameter\_sub\_categories" WHERE "parameter\_sub\_categories"."parameter\_category\_id" = 8 ORDER BY "parameter\_sub\_categories"."name"

=> [#<ParameterSubCategory id: 27, name: "Brightness", parameter\_category\_id: 8, created\_at: "2012-09-14 06:14:27", updated\_at: "2012-09-14 06:14:27">]

To add a Column Mapping name use the command:

ColumnMapping.create!(code:'<Code>', name:'<Name>')

This will result in output similar to:

1.9.2p290 :001 > ColumnMapping.create!(code:'VOL', name:'Volume')

(0.1ms) BEGIN

(1.0ms) SELECT 1 FROM "column\_mappings" WHERE LOWER("column\_mappings"."code") = LOWER('VOL') LIMIT 1

SQL (8.5ms) INSERT INTO "column\_mappings" ("code", "created\_at", "name", "updated\_at") VALUES ($1, $2, $3, $4) RETURNING "id" [["code", "VOL"], ["created\_at", Wed, 31 Oct 2012 14:18:53 EST +11:00], ["name", "Volume"], ["updated\_at", Wed, 31 Oct 2012 14:18:53 EST +11:00]]

(0.9ms) COMMIT

=> #<ColumnMapping id: 6, code: "VOL", name: "Volume", created\_at: "2012-10-31 03:18:53", updated\_at: "2012-10-31 03:18:53">

1. Migrating data to a new system

To restore a **pg\_dump** you pass the file to psql with an empty database. If you have an existing database with the same name, you need to drop it first and recreate it.

The command to drop the database is **dropdb**. So you 'su' to the **postgres** user and run the command:

$ sudo su - postgres

$ dropdb <database name>

$ createdb <database name>

Once you have done that, you can exit the **postgres** user, and restore the database dump:

$ exit

$ psql -U <user> <database name> < sql\_dump.sql

To restore the data, you need to untar it into your root directory. It is likely that your permission system won't allow you to create a directory under root, so you should create it manually, and assign the right permissions to it:

$ sudo mkdir /data

$ sudo chown <user>:<group> /data

$ cd /

$ tar xvf <tar file>

1. Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Version No. | Revision Date | Summary of Changes | Revised by |
| V1.0 | 15/11/12 | Initial | Stuart Allen |
| V1.1 | 15/12/12 | After internal Intersect review | Stuart Allen |
| V1.2 | %%% | %%% | Peter Roberts |
|  |  |  |  |

1. The Bagit format

BagIt is currently defined in an Internet Engineering Task Force ([IETF](http://en.wikipedia.org/wiki/IETF)) internet draft.

Quoting from the preamble of the Bagit entry on Wikipedia:

BagIt is a hierarchical file packaging format designed to support disk-based storage and network transfer of arbitrary digital content. A "bag" consists of a "payload" (the arbitrary content) and "tags", which are metadata files intended to document the storage and transfer of the bag. A required tag file contains a manifest listing every file in the payload together with its corresponding checksum. The name, BagIt, is inspired by the "enclose and deposit" method,[[1]](http://en.wikipedia.org/wiki/BagIt#cite_note-ENCDEP-1) sometimes referred to as "bag it and tag it".

Bags are ideal for digital content normally kept as a collection of files. They are also well-suited to the export, for archival purposes, of content normally kept in database structures that receiving parties are unlikely to support. Relying on cross-platform (Windows and Unix) filesystem naming conventions, a bag's payload may include any number of directories and sub-directories (folders and sub-folders). A bag can specify payload content indirectly via a "fetch.txt" file that lists URLs for content that can be fetched over the network to complete the bag; simple parallelization (e.g., running 10 instances of "wget") can exploit this feature to transfer large bags very quickly. Benefits of bags include

* Wide adoption in digital libraries (e.g., the Library of Congress).
* Easy to implement using ubiquitous and ordinary filesystem tools.
* Content that originates as files need only be copied to the payload directory.
* Compared to XML wrapping, content need not be encoded, saving time and storage space.
* Received content is ready-to-go in a familiar filesystem tree.
* Easy to implement fast network transfer by running ordinary transfer tools in parallel.

Further information about the Bagit hierarchical file packaging format can be found at various places on the Internet, including

Internet Engineering Task Force – <http://www.ietf.org>

Wikipedia – <http://en.wikipedia.org/wiki/BagIt>

Version 0.97 of the Bagit specification - <http://tools.ietf.org/html/draft-kunze-bagit-08>

1. RIF-CS

Quoting from the Global Registries website (<http://globalregistries.org/rifcs.html>):

The **Registry Interchange Format - Collections and Services** (RIF-CS) Schema was developed as a data interchange format for supporting the submission of metadata to a collections service registry. It is based on ISO2146 but only includes elements needed for a collection service registry and so is not a full binding to the standard.

A collection in the RIF-CS Schema context could be a repository, a registry, a collective work or an index/database. There are no hard and fast rules about what constitutes a collection and it is up to the data providers to consider what their collections are and what metadata should be provided. The RIF-CS schema also supports other registry object types, namely services, activities and parties. Any or all of these along with their relations to each other are able to be expressed in RIF-CS format.

The Australian National Data Service (ANDS – http:// <http://www.ands.org.au>) uses the RIF-CS standard for management of data in the Australian Research Data Commons. It provides a training resource for RIF-CS at <http://www.ands.org.au/training/rif-cs/index.html>.

ANDS uses the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH – see <http://www.openarchives.org/pmh/tools/tools.php>) to collect RIF-CS data.

1. Data File Upload Scenarios

%%% Maybe this table would be better included in the file upload section. Or at least, put a strong reference to this table there.

If a TOA5 format CSV data file is uploaded to HIEv, the TOA5 format is automatically detected and the processing is as shown in the table below.

| **IF...** | **Type selected is** | **File type is** | **File overlap is** | **File name is** | **THEN...** | **Resulting type will be** | **Resulting file name will be** | **Resulting messages to user** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | RAW | TOA5 | None | does not already exist |  | RAW | as per uploaded | success |
|  | RAW | TOA5 | None | already exists |  | RAW | suffixed - see (1) | filename change (3) |
|  | RAW | TOA5 | Safe | does not already exist |  | RAW | as per uploaded | safe replacement (2) |
|  | RAW | TOA5 | Safe | same as file being replaced |  | RAW | as per uploaded | safe replacement (2) |
|  | RAW | TOA5 | Safe | already exists (but is not the file being replaced) |  | RAW | suffixed - see (1) | safe replacement (2), filename change (3) |
|  | RAW | TOA5 | Unsafe | does not already exist |  | ERROR | as per uploaded | bad overlap (4) |
|  | RAW | TOA5 | Unsafe | already exists |  | ERROR | suffixed - see (1) | bad overlap (4), filename change (3) |
|  | RAW | Non-TOA5 | N/A - only for TOA5 | does not already exist |  | RAW | as per uploaded | success |
|  | RAW | Non-TOA5 | N/A - only for TOA5 | already exists |  | RAW | suffixed - see (1) | filename change (3) |
|  | Not RAW | TOA5 | N/A - we don't check unless RAW | does not already exist |  | as specified | as per uploaded | success |
|  | Not RAW | Non-TOA5 | N/A - only for TOA5 | already exists |  | as specified | suffixed - see (1) | filename change (3) |

(1) suffixed means appending \_1 (or the next available number) - e.g. blah.dat becomes blah\_1.dat (or blah\_2.dat if blah\_1.dat already exists)

(2) MESSAGE: The file replaced one or more other files with similar data. Replaced files: <filenames here>

(3) MESSAGE: A file already existed with the same name. File has been renamed.

(4) MESSAGE: File cannot safely replace existing files. File has been saved with type ERROR. Overlaps with <filenames here>