DC21 User Manual

For software version 2.0

31 January 2014

There has been a significant rise in the number of projects collecting research data in recent years. This growth has brought with it the challenge of managing the increased data acquisition infrastructure and the increased amount of data produced by this infrastructure.

DC21 was developed as a general purpose data capture application to address these challenges and specifically aims to:

- Allow tailoring to support different applications or research projects.

- Ensure raw data is never lost and can be used and interpreted in the future

- Allow researchers to make linkages between different types of data

- Allow researchers to package and archive data, assign a global unique ID and to cite data in research publications

- Make it easier for researchers to demonstrate that their research is reproducible

- Make data more discoverable so it is easier for researchers to get access to the data they need

- Make it easier for researchers and technical officers to distribute the data

- Make it easier for researchers to comply with institutional and government requirements with regard to research data management.

As a web application, a DC21 Implementation is easily accessed from a variety of locations and platforms with any modern web browser.

The DC21 product was initially developed for the Hawkesbury Institute for the Environment (HIE) at the University of Western Sydney, with funding provided by ANDS under their Data Capture program. The National eResearch Tools and Resource (NeCTAR) project has also contributed funding. It was installed under the name HIEv. Subsequent implementations of DC21 include DIVER, with funding from the Macquarie University and used for marketing data within the Faculty of Business and Economics.

DC21 is open source and available on GitHub in the dc21 and dc21-doc repositories.

Contents

1   Overview

A DC21 Implementation is designed to act as a central repository for research data. Technicians can configure field PCs to automatically push time-series data from sensors or other data acquisition devices into DC21. Time-series data and other data, such as photographs, videos, sound recordings, spreadsheets and other files can be uploaded manually. Researchers can use the system to discover and download the latest data available. Rich Metadata is stored for the individual Data Files, as well as for infrastructure and groupings of Data Files to support discovery and interpretation.

Because DC21 is open source, it is possible for other users to augment functionality and add support for further file formats.

Version 1.0 of DC21 dealt primarily with sensor networks which utilise Campbell Scientific data loggers, and provided rich metadata for Data Files which are uploaded in the TOA5 file format.

Version 2.0 of DC21 has been expanded to provide support for more general Data File categories, including image and video data. It has been implemented for the Faculty of Business and Economics at Macquarie University and is known as the DIVER system on that site. As an example, the following is the Context Diagram for that system.

DIVER CONTEXT DIAGRAM
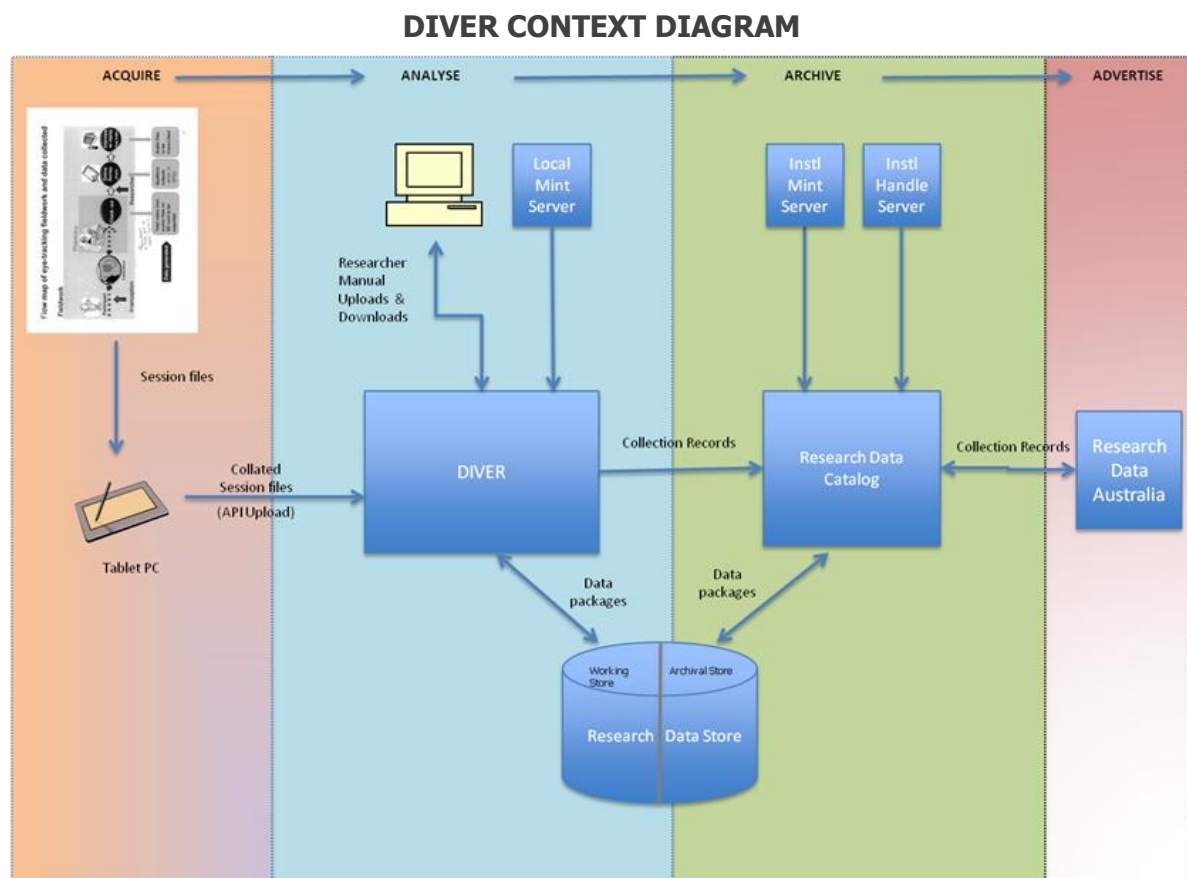
At Macquarie University, the Research Data Catalog in the above Context Diagram is implemented using ReDBox. See http://www.redboxresearchdata.com.au/ for more information.

All DC21 implementations take advantage of AAAA Data Management concepts.

All Data Files in DC21 are grouped according to a two level grouping structure based on Organisational Unit and Project, providing a convenient way to organise related files. See *Chapter 5 Organisational Units and Projects* for details. In addition, when Data Files are created from other Data Files, DC21's Parent-Child relationships can be used to keep track of that information.

The processing status of Data Files can be tracked through the steps of uploading, cleansing, analysis and publishing using the Type field for each Data File.

Once finalised, Packages of data can be defined, described, published to a local institutional repository and then published to ANDS as a RIF-CS data collection. This makes them available for discovery via the OAI-PMH protocol. See *Appendix B - RIF-CS* for more information. This enables researchers from outside the organisation that produced the data to discover it, and to request access to download a copy.

1.1 Installing and Tailoring DC21 for your Organisation

Potential DC21 users should contact Intersect to set up a hosting arrangement.

Alternatively, DC21 can be installed on an institution's local servers. Instructions for installing and upgrading to a new version of the DC21 can be found in DC21's documentation GitHub archive https://github.com/IntersectAustralia/dc21-doc/blob/master/README.md. At this link, you can find the following links and information.

- A list of available DC21 versions, including the last stable version.

- A link to the downloadable documentation ZIP file for each version. That ZIP file contains PDF versions of at least the User Guide (this document) and the Release Notes. The Release Notes for each version in turn contain links to the related installation and deployment instructions.

- A link to the tagged GitHub repository for each released version.

- Links to the GitHub repositories for the various DC21 add-on tools, such as the DC21-eyetracker-packager and the DC21 RESTful API Uploader.

Typically, a system administrator with Linux skills should do the DC21 installation.

After installation, the system must be tailored to your organisation's needs. Refer to section *11.5 Tailoring DC21 for Your Organisation's Needs* to see details of the settings available. To avoid confusion, tailoring the system to your organisation's needs should be done prior to any user being authorised to access the system.

2 Glossary

AAF Australian Access Federation - AAF provides the means of allowing a participating institution and/or a service provider to trust the information it receives from another participating institution. In particular, it allows researchers to use their home institution Login to access a growing number of participating services and resources. See http://aaf.edu.au/

ANDS Australian National Data Service – ANDS is building the Australian Research Data Commons: a cohesive collection of research resources from all research institutions, to make better use of Australia's research data outputs. See http://www.ands.org.au/

API Application Program Interface. For DC21, this is an HTTP interface which provides programmatic access to DC21.

Bagit A general purpose container file format. See *Appendix A - The Bagit format*.

FOR Field of Research code – 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/Products/1297.0~2008~Main+Features~Chapter+3,Fields+of+Research?OpenDocument#112714291310995051

Metadata Data about a file of data. Typically, it may include information such as the date and time to which the Data File relates, who collected it, where it was collected, why it was collected, explanation of columns in the Data File, or any other information about the data in the file.

MIME Type MIME (Multipurpose Internet Mail Extensions) Type is a standard way of identifying files according to their nature and format. MIME Types are managed by the Internet Assigned Numbers Authority (IANA). MIME Types were originally defined for use with email attachments, but are now widely across many different contexts.

The MIME Type string consists of two parts separated by "/". The first part indicates a logical format grouping. Examples are "image", "text", "application", "audio", etc. The second part indicates the specific format. Example MIME Type strings are "image/jpeg" for a JPEG image file, "text/plain" for a simple text file, "audio/mpeg3" for an MP3 audio file. MIME Type strings are generally easily understood.

If a string is prefixed by "x-", it is a non-standard type which has not be registered with the IANA. If a string is prefixed by "vnd-", it is vendor specific.

Mint See ReDBox Mint in this glossary.

Optical Character Recognition (OCR)	The process of extracting text information from an image file and writing into a file which can be accessed by a text editor, word processor or other text processing program.
RDA	Research Data Australia. See http://researchdata.ands.org.au/.
ReDBox	Research Data Box – ReDBox is a metadata registry application for describing research data. It is managed and supported by Queensland Cyber Infrastructure Foundation (QCIF). The project received initial funding support from the Australian National Data Service (ANDS) and the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) program and the Education Investment Fund (EIF) Super Science Initiative. See http://www.redboxresearchdata.com.au/
ReDBoX Mint	The Mint is a name-authority and vocabulary service that complements ReDBox. RedBox Mint is used to supply FOR codes for DC21. It is Open Source software. See http://www.redboxresearchdata.com.au/
RIF-CS	Registry Interchange Format - Collections and Services. See *Appendix B - RIF-CS*.
Ruby on Rails	The programming language in which DC21 is developed. See http://rubyonrails.org/ for more information.
Speech Recognition (SR)	The process of interpreting the spoken word in audio or video recordings and extracting it as text information. It is written into a text file which can be accessed by a text editor, word processor or other text processing program.
TOA5	TOA5 format files are produced by the Campbell Scientific LoggerNet program. They are column formatted Data Files containing header information which describes the data in each column.

3 Logging in to a DC21 Implementation

To begin using a DC21 Implementation, enter the DC21 Implementation's URL into your web browser. See your system's Application Administrator to find out your URL.

3.1 Classes of Users

DC21 defines three classes of Users. You will be assigned to one of these classes by the person who authorises your request for Sign In to the system.

Class of User	Characteristics of Class
Researcher	Has no access to the Admin tab and its functions. Cannot edit the Metadata of files uploaded or created by other Users. Cannot delete files uploaded or created by other Users. Cannot Publish files. (But can create Packages.)
API Uploader	Has exactly the same permission restrictions as the Reseacher Class of Users. Is intended for non-personal accounts that will be used for automated data uploading using the DC21 API. Therefore, the API Token is of most relevance to this class of user, even though it can be used by all Classes of Users.
Administrator	Has permission to perform all functions in this DC21 Implementation, including authorising new Users' requests for access to this DC21 Implementation.

3.2 Choosing your Login Method

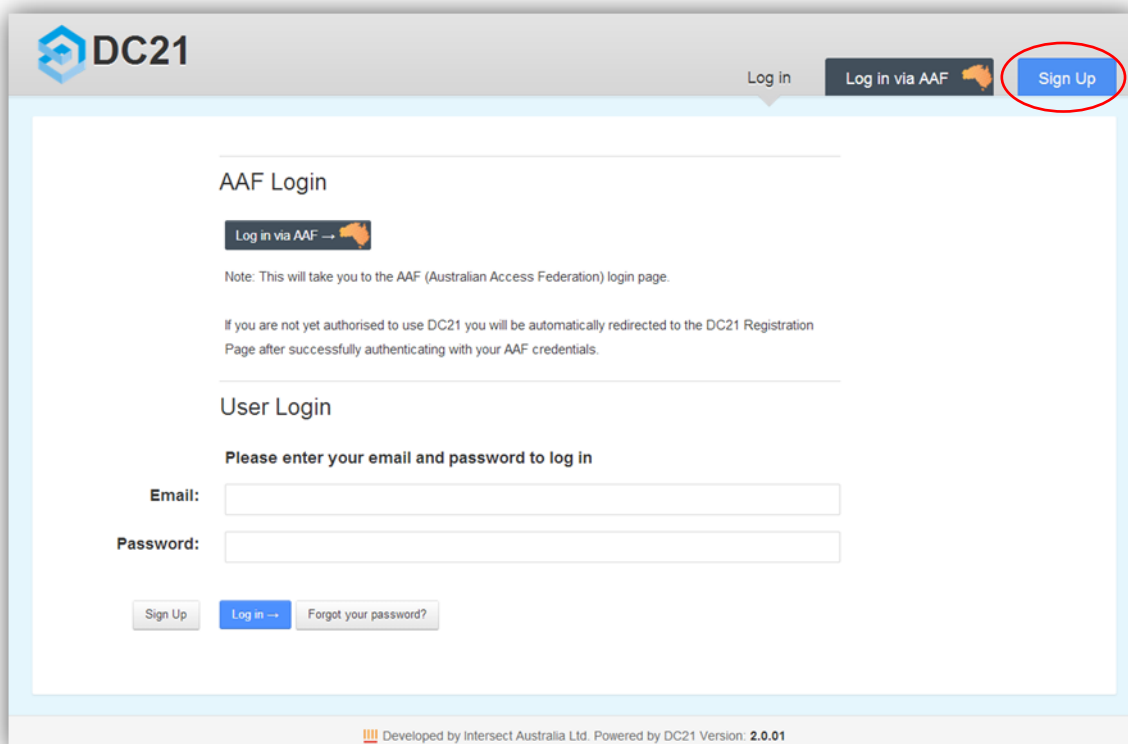There are two methods available for signing in to DC21.

- Australian Access Federation (AAF) authentication allows you to use the one login account to access many services. This is the preferred method if your institution supports it.
- DC21's standard authentication does not rely on any other facilities.

Even if you have set up AAF Authentication, you can still use DC21's standard authentication to access your DC21 account.

3.3 Standard DC21 Authentication

3.3.1 Signing Up

Before you can login you are required to have an account. You can apply for an account by clicking Sign Up on the top right of the login screen.

This will display a form where you will be requested to enter your first name, last name, email address and chosen password. 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.
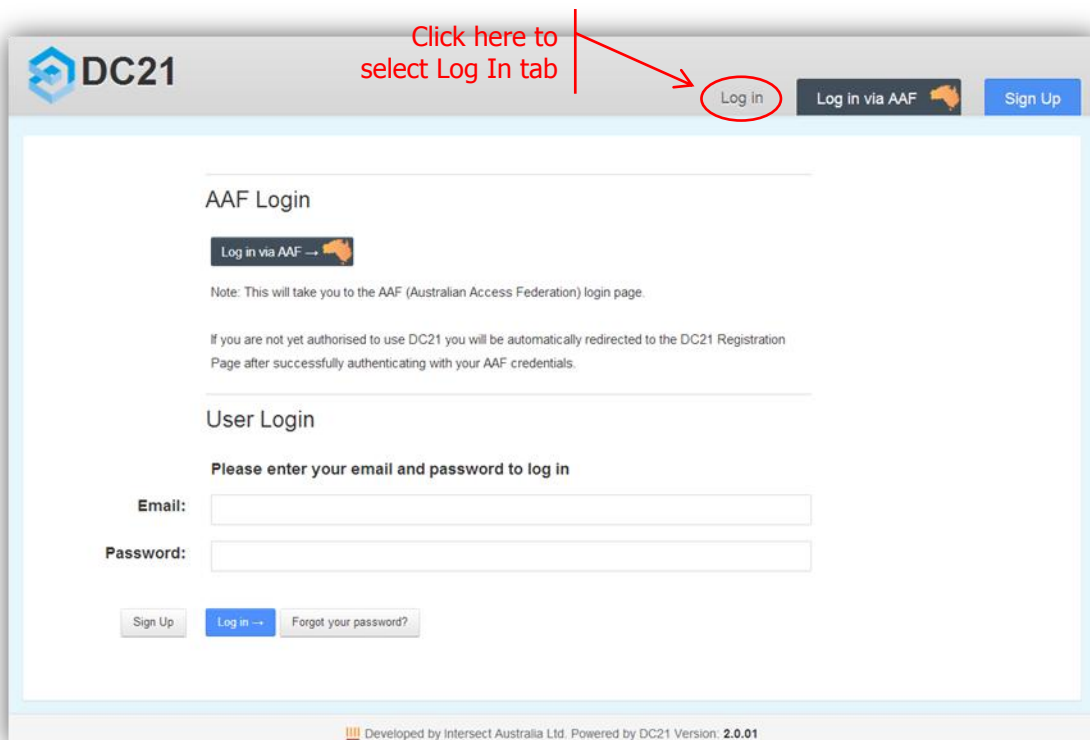
Once you have filled out this form and clicked on `Submit Request` an email will be sent to your system's Application Administrator who will either approve or deny your request for access. Your Application Administrator will typically be the Business Owner or Data Manager. (This will probably not be your System Administrator, who would typically be a technical IT person.) If your request is approved, you will receive an email informing you that you can now login using the password you entered in the original sign up form.

3.3.2   Standard Login

To see the login form, make sure you have the **Log in** tab selected on the top right. Enter your Email address and password, and click on `Log in →`.

Once you have logged in you will be taken to the main screen for the application. See section *4.1 The DC21 Main Screen*.

3.4   AAF Login

To authenticate using AAF, click `Log in via AAF` at the top of the DC21 log in screen or click the `Log in via AAF →` button in the body of the screen. You will be redirected to the AAF log in screen, into which you should enter your normal AAF credentials. Once you have successfully authenticated using AAF, you will be redirected to the DC21 main screen shown in section *4.1 The DC21 Main Screen*.

If you do not have a DC21 account, you will instead be redirected to the DC21 Sign Up screen as shown above in section *3.3.1 Signing Up*. Your name and email address will be pre-filled with your AAF details. You must enter a password in this screen. It is more secure

to use a different password from your AAF password. When you click Submit Request, an account creation request for your DC21 system will be raised and you will be returned to the DC21 Log In screen.

As described in section *3.3.1 Signing Up*, you will receive an email when your account is approved. You can then log in to DC21 using AAF authentication.

You can also log in using DC21's standard authentication method by directly using the User Login part of the DC21 log in screen as described above in section *3.3.2 Standard Login*. You must enter the email address that is specified in AAF and the password which you entered into the DC21 Sign Up screen when you created your DC21 account.
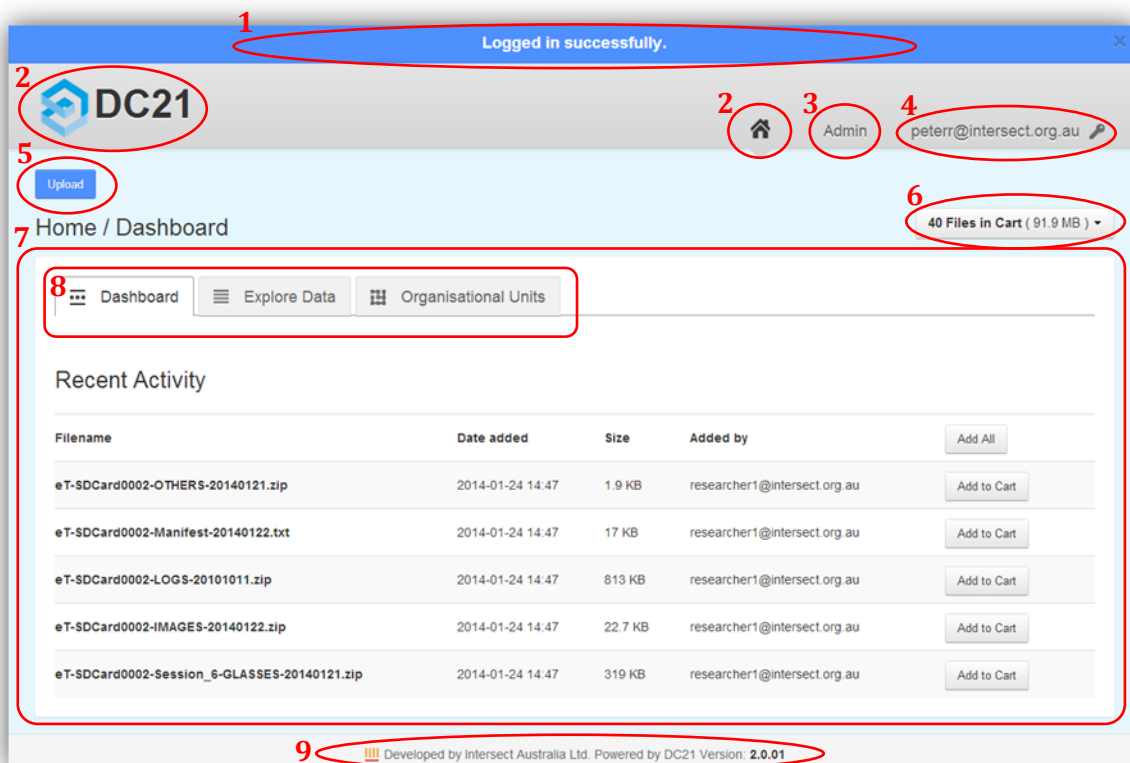
Note	If you already have a user account for your DC21 system, you can log in using AAF provided that your AAF email address and your DC21 email address are identical.

4 General Operation

This chapter describes aspects of the operation of DC21 which are common across a number of screens.

4.1 The DC21 Main Screen

The Main Screen consists of the following parts:

Most of these components are common to many of the screens in DC21.

1	Message bar	This blue area will appear when DC21 displays an error or informative message.
2	Home ⌂ button and System Logo	Click either of these to show the Home/Dashboard (that is, the view shown above). The screen shot above shows "DC21", but your system will show the name of your system and your system's logo. These are configurable values (see section *11.5 Tailoring DC21 for Your Organisation's Needs*).
3	**Admin** button	Click to access the Administration functions (see Chapter 11 *DC21 Administration*). This button is only present if you have Administrator permissions.
4	Login ID	This is your login email address. Click to open a dropdown menu of user

operations. (See sections *4.2* and *4.5*.)

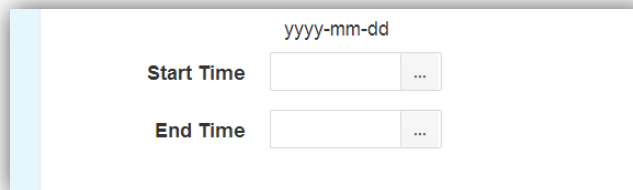| 5	Action button	In many screens, there is an action button at the top left corner. It is often `Upload` , which allows you to upload new Data Files to DC21. (See Chapter 7 *Uploading Data Files* for more information.) However, for some screens it will be a button for another function which is more relevant to the data being displayed in that screen.
6	Cart status box	The DC21 web interface allows you to add files to a Cart, which operates like an e-Commerce shopping cart. Click in this Cart status box to open a dropdown menu of Cart functions. (See section *8.3 The Cart)*
7	Working area	The content of this work area changes as you perform DC21 operations.
8	Tabs	Click on a tab to go to the default view for that tab. If you click on the tab which is already displayed, it will return to the default view for that tab.
9	Version indicator	This shows the version of DC21 which was used for the DC21 implementation you are accessing.

4.2   Entering Dates and Times

There are a number of places where dates and times may be entered.

Date and time entry fields appear with an ellipsis following the data entry boxes.
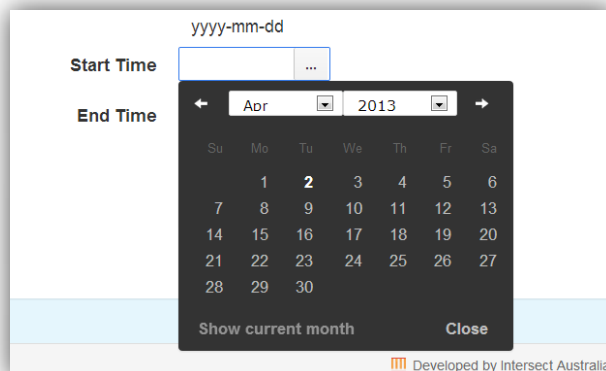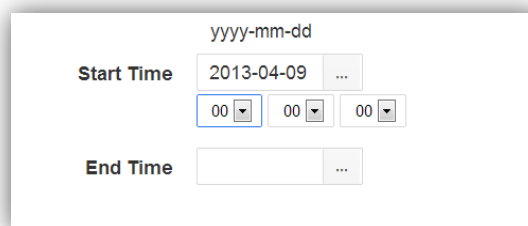

Click on the ellipsis to show a date picker dialog.

Click on the left and right arrows or Month and Year dropdown menus to select the required month. Click `Show current month` to display the calendar for the current month. Click on

the date to select the date and close the date picker dialog. Clicking `Close` will close the date picker without selecting a date.

Alternatively, click in the date entry box and type in a date in YYYY-MM-DD format. After you have entered the date, use the Tab key to move to the next field or click in another field.

Note	Do not press Enter, as this will activate the data entry screen's Save button.
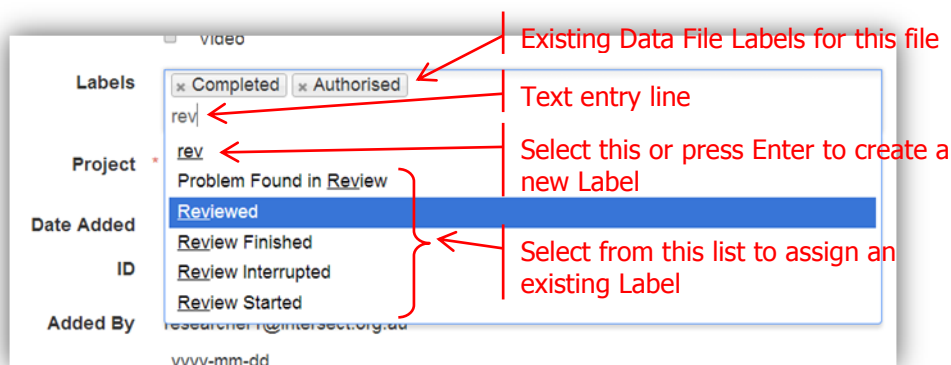
When the date has been selected, three extra numeric fields will be displayed under the date (as shown below) to permit entry of the Hour, Minute and Second for the time. Use the dropdown menus for these fields to enter the desired time. You cannot type the time in using your computer's keypad.

4.3   Entering Labels

Labels are user-defined character strings which can be assigned to Data Files to help identify or describe the Data File. Multiple Labels can be assigned to any Data File.

You can easily enter Labels which have already been assigned to other Data Files. As you type the characters for a Label, all matching existing Labels are displayed in a drop down list. The characters you have typed also appear as the first entry in this list.

Select the Label you require from the dropdown list using the mouse or the arrow and Enter keys.

You can enter a new Label which is not yet used for any Data File. That Label will subsequently appear in the matching dropdown list when Labels are entered, even for other Users of your system.

Click again in the Labels entry box and follow the same procedure to add two or more Labels.

An existing Label can be removed from the Data File by clicking the X associated with the Label.

4.4   Signing Out

Click on your login email address at the top right of the screen to see a dropdown menu.
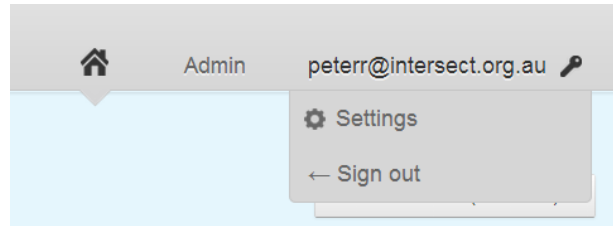
Click on ← **Sign Out** to finish your DC21 session.

4.5   Changing Your User Settings

Click on your login email address at the top right of the screen to see the dropdown menu. Click on **Settings** to access the three tabs described in the following sub-sections.

When finished, click on the Home 🏠 button to return to the DC21 Main Screen.

4.5.1   Overview Tab

This tab displays a summary of your User information.
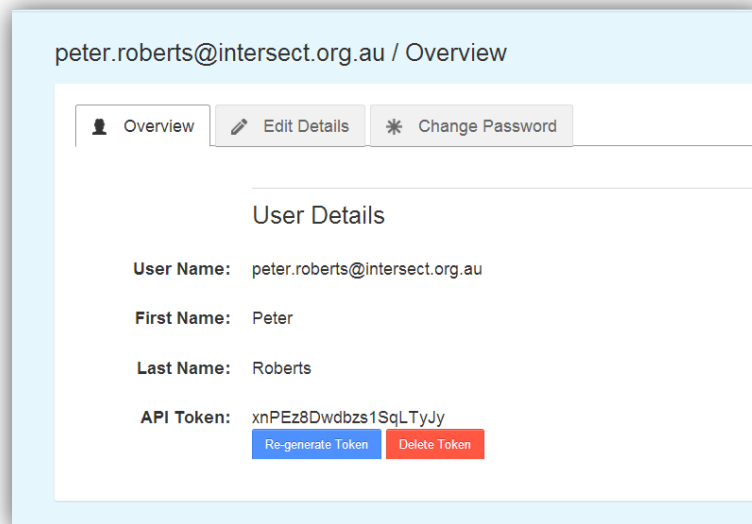
User Name	Your valid email address which you use for logging on.
First Name	Your name.
Last Name	

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 DC21. The HTML API can be used to set up an automatic upload of data from a field PC. For instructions on using the HTML API, see the API definition on the GitHUB repository by going to https://github.com/IntersectAustralia/dc21-doc/blob/master/README.md and selecting the API Uploader link for your DC21 version. From that page, select the File Upload API definition link.

Initially, no token is displayed and only a single `Generate Token` button is displayed. Clicking `Generate Token` will cause a token to be displayed in this field. Copy and paste it into the required place in your API scripts.
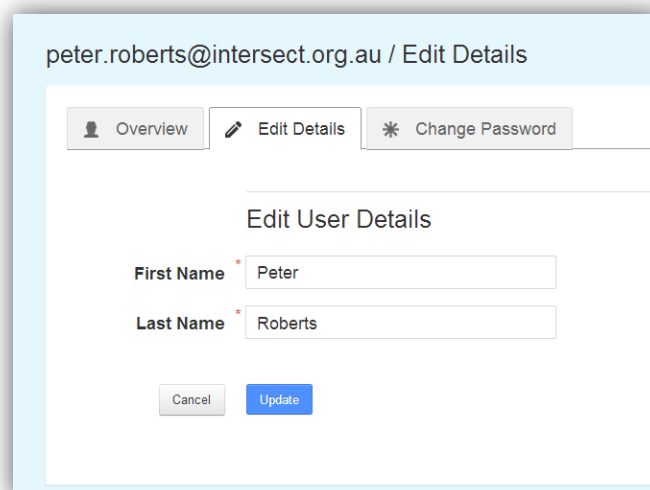
Clicking `Re-generate Token` will cause the current token to be invalidated and a new token to be generated and displayed. You must replace the token value in your API scripts with this new token so that your API scripts continue to work. This button appears instead of `Generate Token` if a valid token is available.

> Note   For security reasons, from time to time you should regenerate your Token and update it in your API scripts.

Clicking `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.

4.5.2   Edit Details Tab

The Edit Details tab allows you to update your First Name and Last Name.

peter.roberts@intersect.org.au / Edit Details

- Overview
- Edit Details
- Change Password

Edit User Details

First Name *   Peter

Last Name *   Roberts

Cancel   Update
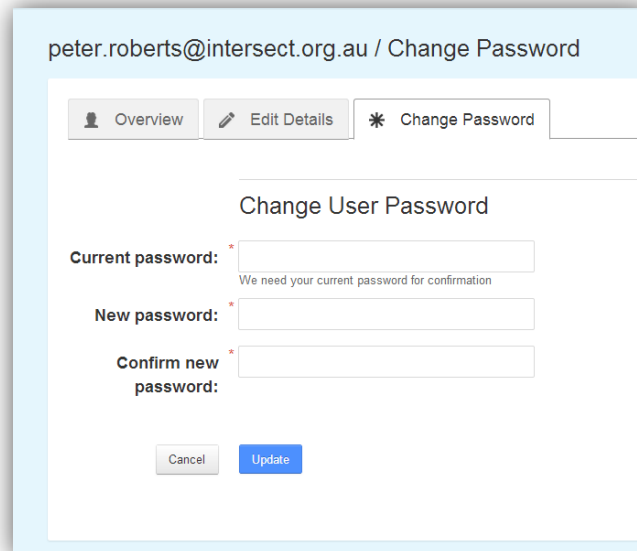
There is no function to change your email address. If you need to do this, you may need to create a new account, or contact your system's Application Administrator.

Click `Cancel` to return to the Overview Tab without accepting changes, and click `Update` to store the changed values you've entered.

4.5.3 Change Password Tab

Use this tab to change your logon password.

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.

You must correctly enter your current password and the strings 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.

5   Organisational Units and Projects

DC21 identifies each Data File according to the Organisational Unit and Project to which it belongs.

Projects
: A Project is a collection of tasks or work which is performed for a specific purpose and is typically funded as a unit.

: All Data Files uploaded to DC21 must be assigned to a Project. There can be many Data Files for each Project.

: The Project information is copied when the Data Files are packaged and Published to the ANDS system.

Organisational Units
: An Organisational Unit is a group or facility within your organisation which is responsible for a number of Projects.

: Each Project is owned by an Organisational Unit. There can be many Projects in each Organisational Unit.

DC21 allows the terms which describe the nature of the Organisational Units and Projects in your organisation to be tailored. For example, you may refer to your Organisational Units as "Research Groups", or perhaps they are "Facilities". You may refer to your Projects as "Experiments", or you may use the term "Projects". See your Application Administrator to learn the terms which have been assigned for Organisational Units and Projects in your DC21 implementation and how they should be used. Instructions for tailoring these terms to your organisation's needs can be found in section *11.5.1 System Configuration parameters*.
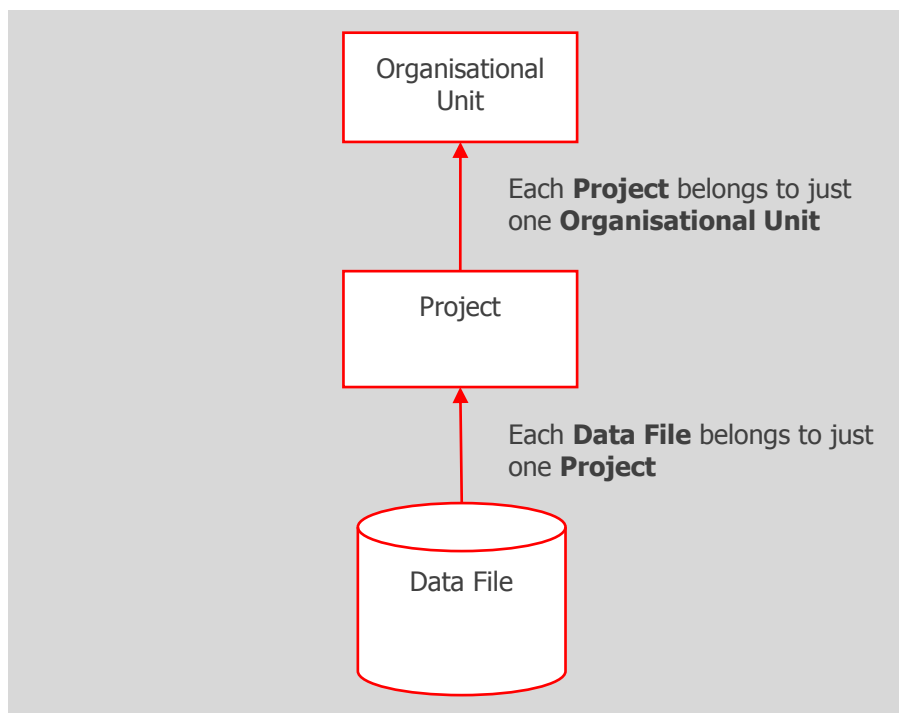
> **Note** Throughout this document, the terms "Organisational Unit" and "Project" are used. However, the screens which DC21 displays use the terms which are tailored for your DC21 Implementation. For example, this document will describe a tab named **Organisational Units**, but on your system it will be named with your tailored term. Similarly, the button New Organisational Unit , which creates an entry for a new Organisational Unit in DC21's table of Organisational Units, will be named with your tailored term.

Organisational Unit entries are stored in a table, as are Project entries.

Before a Data File is uploaded, the Organisational Unit and Project entries it uses must already exist.

Organisational Unit and Project entries are created and updated using the **Organisational Units** tab (which will be named with your system's tailored name) on the main DC21 screen.
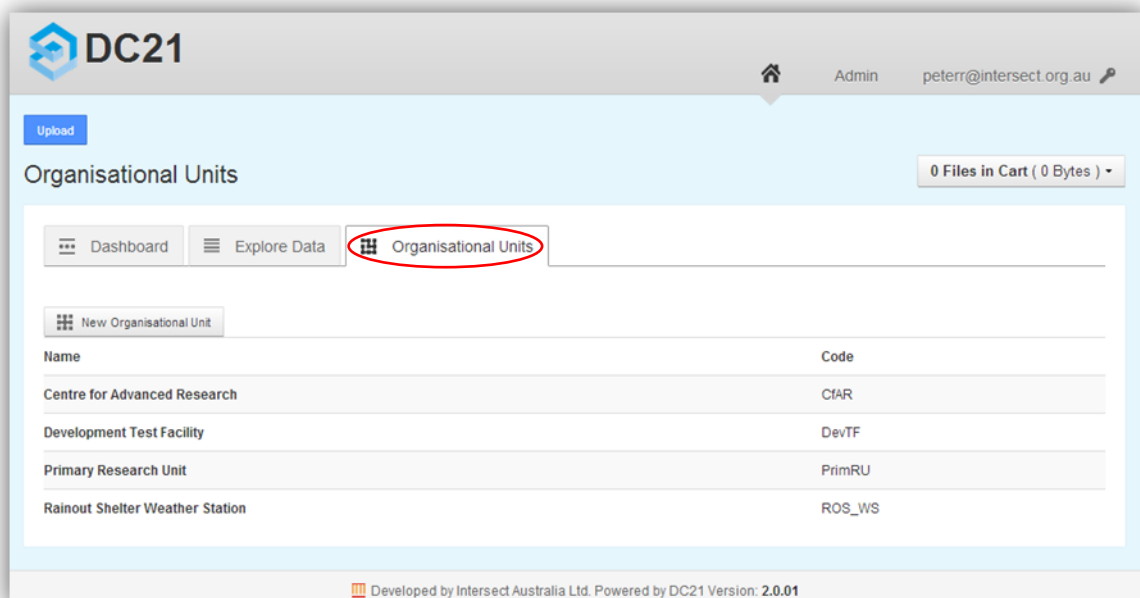
The **Organisational Units** tab lists all of the Organisational Units currently defined in DC21.

5.1   Creating a Organisational Unit Entry
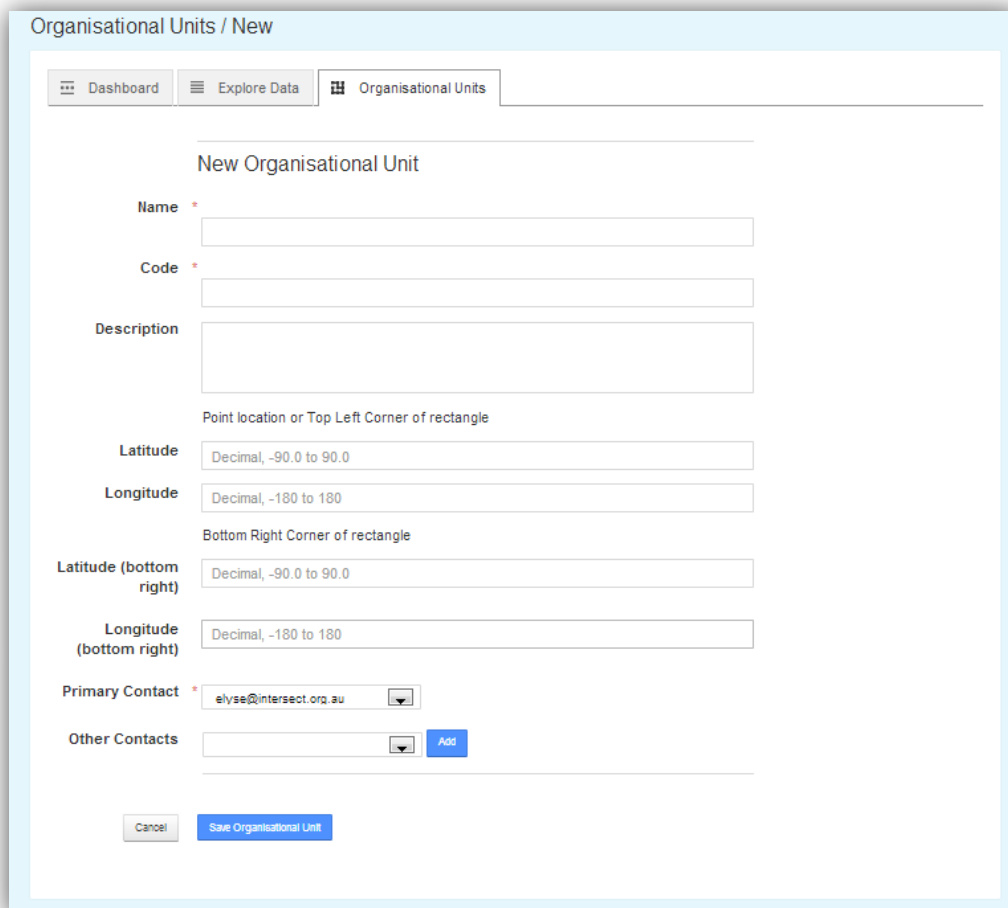
> **Note** Take care. Once created, an Organisational Unit entry cannot be deleted. This restriction prevents Projects which reference an Organisational Unit from becoming invalid if the Organisational Unit were to be deleted.

To create an Organisational Unit entry:

- Click the New Organisational Unit button (which will be named with your system's tailored term) at the top-left of the Organisational Units tab.
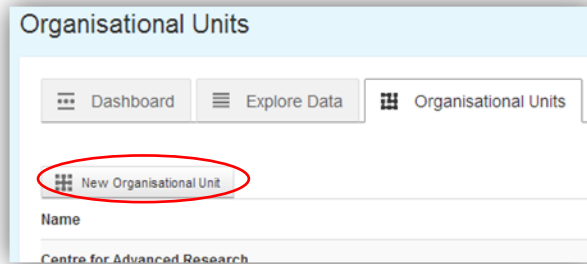
This will display a form that allows you to add details about the Organisational Unit you would like to create. (The words "Organisational Unit" will be replaced by your system's tailored term).
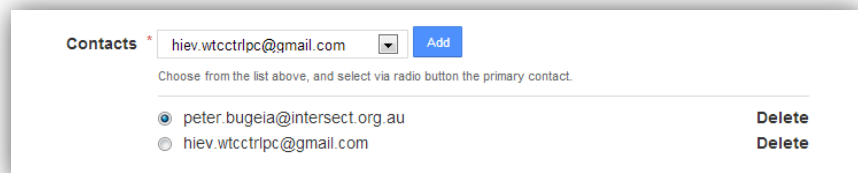
- Enter the details for the Organisational Unit.

Name	The Name for the Organisational Unit is a short, plain-English title that will be used in the application interface to refer to the Organisational Unit. The Name must be unique.
Code	The Code for the Organisational Unit is a short identifying string. The Code must be unique.

Description	The Description of the Organisational Unit should be as comprehensive as possible, describing details that would help a researcher both discover the Organisational Unit when searching and assist that researcher in being able to interpret the data that is produced by the Organisational Unit. These details would include things such as: • The purpose of the Organisational Unit • Types of sensors installed at the Organisational Unit, if appropriate • Location of the sensors within the Organisational Unit, if appropriate
Latitude and Longitude fields	If appropriate, the Latitude and Longitude for the Organisational Unit are expressed in Decimal Degrees (http://en.wikipedia.org/wiki/Decimal_degrees) and can be taken directly from Google Maps. If inappropriate, these fields can be left blank. If a single set of co-ordinates is given, it is considered to be the central point for the Organisational Unit. 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 Organisational Unit. Enter the northwest corner in the first two fields and the southeast corner in the fields labelled "bottom right".
Contacts	The Contacts for the Organisational Unit are selected from the Users registered within DC21. There must be at least one Contact for each Organisational Unit.

Select a Contact from the email addresses shown in the dropdown list and click Add to add it to the Contact list shown below this question.

To remove an incorrectly added Contact, click the word **Delete** 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 Save Organisational Unit to save the Organisational Unit's details and return to the Organisational Units list.

To abandon creating the Organisational Unit entry, click Cancel .

Once Organisational Units have been created they will appear in the list on the Organisational Units tab.

5.2 Editing an Organisational Unit Entry

To modify any of the parameters for an Organisational Unit entry:
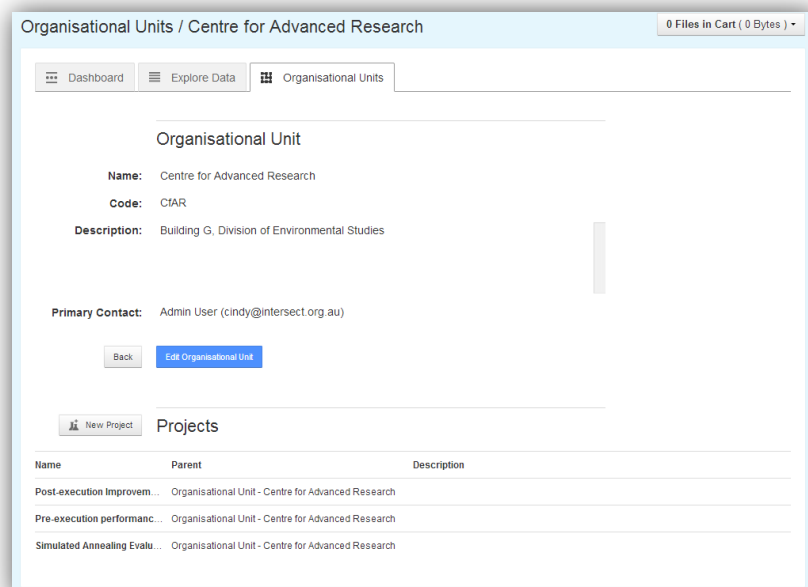
- From the **Organisational Units** tab, click on the name of the Organisational Unit you wish to edit. The Organisational Unit's details will be displayed.
- Click Edit Organisational Unit to open the details edit screen, which is the same as the one described in *5.1 Creating a Organisational Unit Entry* above.
- Change the details as required.
- Click Update to save your changes.

5.3 Creating a Project Entry

Note	Take care. Once created, a Project entry cannot be deleted. This prevents Data Files which reference this Project from becoming invalid if the Project were to be deleted.

To create a Project entry:

- On the **Organisational Units** tab, view the details of the Organisational Unit to which the Project applies by clicking on the Organisational Unit name. The Organisational Unit details will be displayed and all of the Projects registered for that Organisational Unit will be listed below the details of the Organisational Unit itself.

- Click on the New Project button at the top left of the Project list to show the Project parameters entry dialog.

- Enter the details for the Project.

Parent	The Parent for a Project is either the current Organisational Unit or another Project for the current Organisational Unit. If a Project is selected, the new Project is considered a sub-Project of the one selected. (Note that this is different from the general Parent-Child Relationships which are defined between Data Files.)
Name	The Name for the Project should be short, but descriptive enough to uniquely identify the Project, including distinguishing a Project from those that are likely to come in the future. The Name should be unique for this Organisational Unit.
Description	The Description for the Project should describe the purpose of the Project and the techniques employed. Particular focus should be given to aspects of the Project that produce data that are stored in this system.
Start date End date	The Start and End dates for the Project. These dates are not checked against the dates in the Data Files for the project.
Subject	The Subject for the Project is a short phrase describing the Project's main research area. The Subject is primarily recorded to support data publication such as to Research Data Australia, who requests *"A subject is a term, keyword, classification code or phrase representing the primary topic or topics covered by a*

registry object."
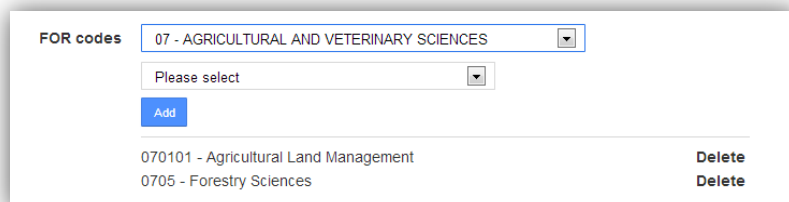
Access rights

The Access rights dropdown list box provides a selection of licences under which the data from this Project will be released. In Australia it is preferred that data is released under a [Creative Commons](#) licence. The list of licences shown in the dropdown list is hard-coded in DC21. More information about these licences can be found at http://creativecommons.org/licenses/.

FOR codes

Each Project 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/Products/1297.0~2008~Main+Features~Chapter+3,Fields+of+Research?OpenDocument#112714291310995051, which says, in part,

The FOR is a hierarchical classification with three levels, namely Divisions (2 digits), Groups (4 digits) and Fields (6 digits). Each level is identified by a unique number.

Each Division is based on a broad discipline. Groups within each Division are those which share the same broad methodology, techniques and/or perspective as others in the Division. Each Group is a collection of related Fields of research. Groups and 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.

The 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 Add to add the selected FOR code to the list for this Project, which will appear below the FOR code selection boxes.

You can click Add after selecting just the Division, the Division and Group, or all three levels, as appropriate for your Project.

FOR codes can be removed from this Project by clicking the word **Delete** to the right of the FOR code in the list.

> Note   DC21 obtains FOR codes from a separate MINT server, the details of which were set up during DC21 installation. (See the Context Diagram in Chapter 1 *Overview*.) If this server is not available at the time you are creating or updating an experiment, you will not be able to select FOR codes. If this occurs, you will receive an error message which you should report to your system administrator. You should still be able to save your experiment and then add FOR codes later when the FOR code Mint server is available again.

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

To abandon creating the Project, click Cancel .

5.4   Editing a Project Entry

To modify any of the details for a Project entry:

- From the Organisational Units tab, click on the name of the Organisational Unit which hosts the Project. The Organisational Unit details and Projects for the Organisational Unit will be displayed.

- Click on the name of the Project you wish to modify. The Project details will be displayed.

- Click on Edit Project to open the Project detail entry screen, which is the same as the one described in *5.3 Creating a Project Entry* above.

- Change the details as required.

- Click on the Save Project button to save your changes.

5.5   Setting Up Project Parameters

Projects can optionally have one or more Project Parameters. These Parameters provide a structured way to describe experimental treatments. They are documentary only. Projects can have multiple Parameters or none at all.

Because Parameters are not directly referenced by Data Files, the option to delete them is provided.
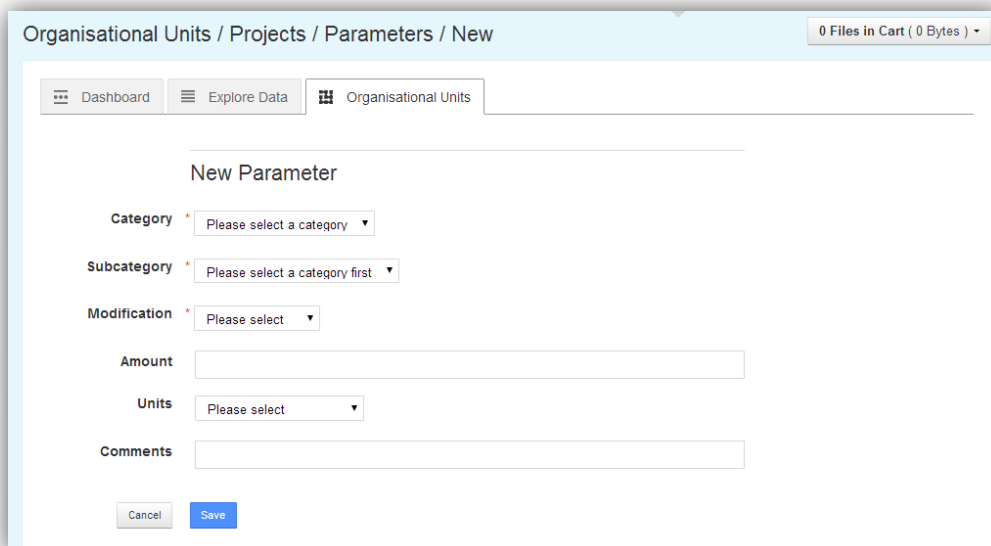
To add a Project Parameter:

- From the Organisational Units tab, click on the name of the Organisational Unit hosting the Project and then the name of the Project to which you wish to add a Parameter. The Project details will be shown.

- Click the New Parameter button directly below the Project details.

This button will display the form below.

- Enter the details of the Parameter. The System Administrator configures the values available in the dropdown list boxes for the questions in this form.

Category and 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 and Units	These optional fields allow more specific information to be recorded about the Modification.
Comments	This field can be used to record any unstructured, plain-text information you would like to record about the treatment.

- Click the Save button at the bottom of the form to save your Parameter details and return to the Project details screen.

Once a Project Parameter has been created, it will appear in the Parameters list below the Project details.

Existing Project Parameters can be edited or deleted using the `Edit` and `Delete` buttons to the right of the parameter in the **Actions** column.

6   Data File Storage and Metadata

DC21 stores uploaded Data Files using a database structure.

The Data Files stored within DC21 can only be accessed by using the DC21 system.

6.1   Data File Types

DC21 supports uploading and storage of any format Data File, regardless of its content or purpose.

When Data Files are uploaded, DC21 examines the file content to determine its format and sets its MIME Type into the Data File's Metadata based on the format discovered.

There are some Data File types for which DC21 provides special processing options.

TOA5	TOA5 Data Files are produced by the Campbell Scientific LoggerNet program. Not all DC21 installations will use this file format.
	DC21 processes these files specifically, as described in section *7.3 Uploading RAW TOA5 Data Files*.
Images with Text	DC21 can be configured to extract the text from certain image files and write it to a text file. See *7.1 Uploading Image Files* and *11.5.2 OCR Processing parameters* for more information.
Audio/video with speech	DC21 can be configured to interpret the speech in certain video and audio files, convert it to text and write it to a text file. See *7.2 Uploading Video and Audio Files* and *11.5.3 Speech Recognition Processing parameters* for more information.
Package Files	Package Files are generated by DC21 and contain data and metadata ready for Publishing to ANDS. Package Files are in BAGIT format. See *Appendix A -The Bagit format* for info on this format, and Chapter 9 *Publishing Your Data* for instructions to create and Publish a Package File.

6.2   Metadata

In addition to storing the files themselves, DC21 also stores information about each file, known as "Metadata". This Metadata falls into four categories.

6.2.1   Basic Information

The Basic Information is Metadata entered by the user when the Data File is uploaded. It consists of the following fields.

Name	The Name to be used for the file which stores the data in DC21 Working Store.

Type	The Type of the Data 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 Application Administrator. The file's Type is generally aimed at tracking data through its various stages of processing.

There are three predefined data Types which are built into the system and cannot be changed:

RAW   indicates a file that will be tested for known upload File Formats on file upload.

ERROR   indicates a file that has failed to process or upload correctly. Errors typically occur on uploading TOA5 files or on OCR or SR processing. For OCR and SR processing, the output file will have this Type set. (See *7.3 Uploading RAW TOA5 Data Files* for more information about TOA5 Data Files and sections *11.5.2* and *11.5.3* for information on OCR and SR processing.)

PACKAGE   indicates a file containing a collection of Data Files which is intended for Publishing. See Chapter 9 *Publishing Your Data* for more information.

If the Type field has been set to one of the above predefined Types, it cannot be changed. If it has been set to one of the non-predefined Types, it can be changed, but only to another non-predefined Type.

File Format	This field will display the MIME Type of the Data File in the standard format, such as **text/plain** or **image/jpeg**.

If the file is known to be a specific format which is handled by DC21, one of these following specific values may be shown.

TOA5   The file was inspected on upload and discovered to be TOA5 format. TOA5 files are processed differently on file upload. See section *7.3 Uploading RAW TOA5 Data Files* for more information.

BAGIT   The file is a Package which is formatted as a BAGIT ZIP file. See *Appendix A - The Bagit format* for more information on BAGIT files.

If the file format cannot be determined, this field will display **UNKNOWN**.

The File Format is set automatically when the Data File is uploaded or created.

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 Data File from a constrained list of possibilities defined by your Application Administrator. A Data File may be assigned multiple Tags or no Tags at all.
	It is recommended that for new work you use the newer Labels functionality instead of Tags. Labels provide the same functionality in a more flexible and comprehensive way.
Labels	Labels are user-defined character strings which can be assigned to Data Files to help identify or describe the Data File. Once defined, a new Label is available for use by all Users across all files, Projects and Organisational Units.
	Multiple Labels can be assigned to any Data File.
	This field lists all Labels which are assigned to this Data File.
Project (this name will be replaced by your system's tailored term)	This field indicates which Project produced the Data File. A Data File can be associated with only one Project. Any User can create Projects. See Chapter 5 *Organisational Units and Projects* for more information.
Organisational Unit (this name will be replaced by your system's tailored term)	The Organisational Unit field indicates which Organisational Unit is associated with the above Project. Each Data File must be associated with exactly one Organisational Unit. Any User can create Organisational Units. The Organisational Unit field cannot be directly set. Its value is copied from the Organisational Unit field in the selected Project. See Chapter 5 *Organisational Units and Projects* for more information.
Date added	The date on which this Data File was added to the database. For Packages, it's the date on which the package was created. This field is set automatically by the system.
Creation Status	This field tracks the progress of background file creation, such as the creation of Package .ZIP files or the processing of OCR and SR. Valid Status values are QUEUED, WORKING, COMPLETED and FAILED. For Package .ZIP file creation, it works together with the "Packaging Progress Field". For OCR and SR processing, this Status is applied to the output file.
Size	The size of this file. This field is set automatically by the system when a file is uploaded or a Package is created. For Packages, OCR output and SR output, this field will show zero until the processing has completed without error.
Packaging progress	This is a temporary field that is only available while a Package is being created. This field tracks the progress of Package creation by showing the progressive number of bytes that have been packaged so far. See section *9.1 Creating a Package* for more information.

File ID	File IDs are unique integers which are assigned and used internally by DC21 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 external ID which has been used outside of DC21 to identify this data. It is a character string. For Data Files, the field is input by the user and it can be used to provide an ID to identify an individual Data File. For DC21 Packages, the field is read-only and auto-generated on Package creation. The ID field, along with other Metadata, is copied into the RIF-CS file when the file is copied into a Package. No two files in DC21 can have the same non-null ID. See sections *9.1 Creating a Package* and *9.3 Managing Published Packages* for more information about the use of this ID field.
Added by	This field indicates the User who uploaded the file to DC21. For Packages, it's the User who Packaged it. For OCR and SR output .TXT files, it's the User who ran the processing. This field is set automatically by DC21.
Published	This applies to Packages only and indicates whether the Package has been Published or not. This field is set automatically by DC21.
Published date	This applies only to Packages which have been Published and indicates whether the date on which the Package was Published. This field is set automatically by DC21.
Start time End time	These fields only apply to Packages and non-TOA5 Data Files. They hold the dates and times which were manually entered when the file was uploaded or created. These times can be specified with a precision of one second and 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.)

6.2.2   File Relationships

Parent-Child File Relationships indicate the origin of Data Files or the use of Data Files. The existence of a relationship indicates that the Parent Data File was used to generate or create the Child Data File. A file can have multiple relationships, either as Parent or Child.

If the Metadata of a particular Data File reports a Child Data File, then that Child Data File's Metadata will show this particular Data File as one of its Parents.

There is no restriction on the number of Parents or Children that any Data File can have.

Parent-Child File relationships can be created manually or automatically. Examples of automatic creation include:

- When a text file is generated from an image file by using Optical Character Recognition (OCR), the image file is set as a Parent of the generated text file.

- When a text file is generated from an audio or video file by using Speech Recognition (SR), the audio or video file is set as a Parent of the generated text file.

- When a Package file is constructed from a number of Data Files, all of those Data Files are set as Parents of the newly constructed Package file. (There can be very many Parent Data Files of a package file.) See Chapter 9 *Publishing Your Data* for more information on creating Packages.

Manual creation of Parent-Child relationships is done using the Metadata Edit screen. You can use this to manually indicate files which are derived in some manner from another Data File. See section *8.4 Viewing and Editing a File's Metadata*.

The Metadata View screen shows two fields:

Parents	All files which are recorded as Parents of the current Data File. The files listed here will have the current Data File listed as their Child.
	Clicking on the name of a file in this list will open the Metadata View screen for that Data File.
Children	All files which are recorded as Children of the current Data File. The files listed here will have the current Data File listed as their Parent.
	Clicking on the name of a file in this list will open the Metadata View screen for that Data File.

Note	The Metadata Editor checks the validity of Parent-Child relationships and will not allow a Data File to be its own Child, nor its own grandchild. However, it does not check for relationship loops longer than that. It is possible to create a set of relationships in which a Data File is its own great-grandchild. This is not recommended.

6.2.3   Summary 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 DC21 Users.
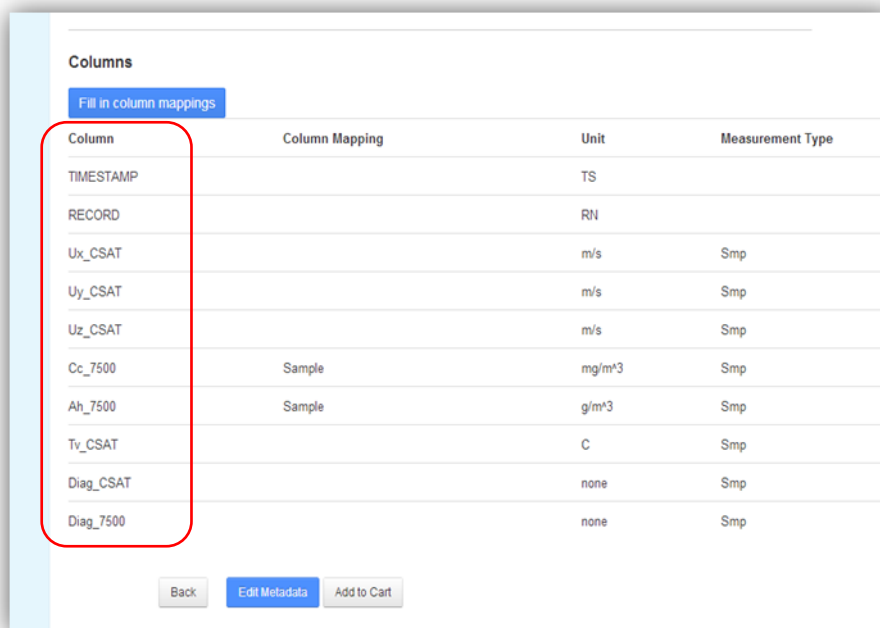
Start time End time	The first and last times of the observations found in the TOA5 Data File.
Sample interval field	The frequency of samples in the Data File, if relevant.
Datalogger model	The model of data logger used to generate the TOA5 file.
Station name	Designates the Data Logger from which this data was generated and is checked during the upload of a TOA5 file to determine whether an existing TOA5 file or files needs to be replaced.

Serial number	These five fields are also checked during the upload of a TOA5 file to
Os version	determine whether an existing TOA5 file or files needs to be replaced.
Dld name	
Dld signature	
Table name	

6.2.4   Column Information for TOA5 Files

Column Information Metadata is collected automatically from TOA5 files when they are uploaded. This automatically collected Metadata for each TOA5 Data File is displayed whenever the TOA5 Data File's Metadata is displayed. It cannot be changed by DC21 Users.


In addition to the Column Information from the TOA5 file there is one extra column called Column Mapping, circled in red above. The value in this column is assigned by matching the value in the TOA5 column heading against the values in DC21's Column Mapping Table. It is used to simplify searching by TOA5 column heading. See section _11.3 Managing Column Mappings_ for more information about setting up the Column Mapping table.
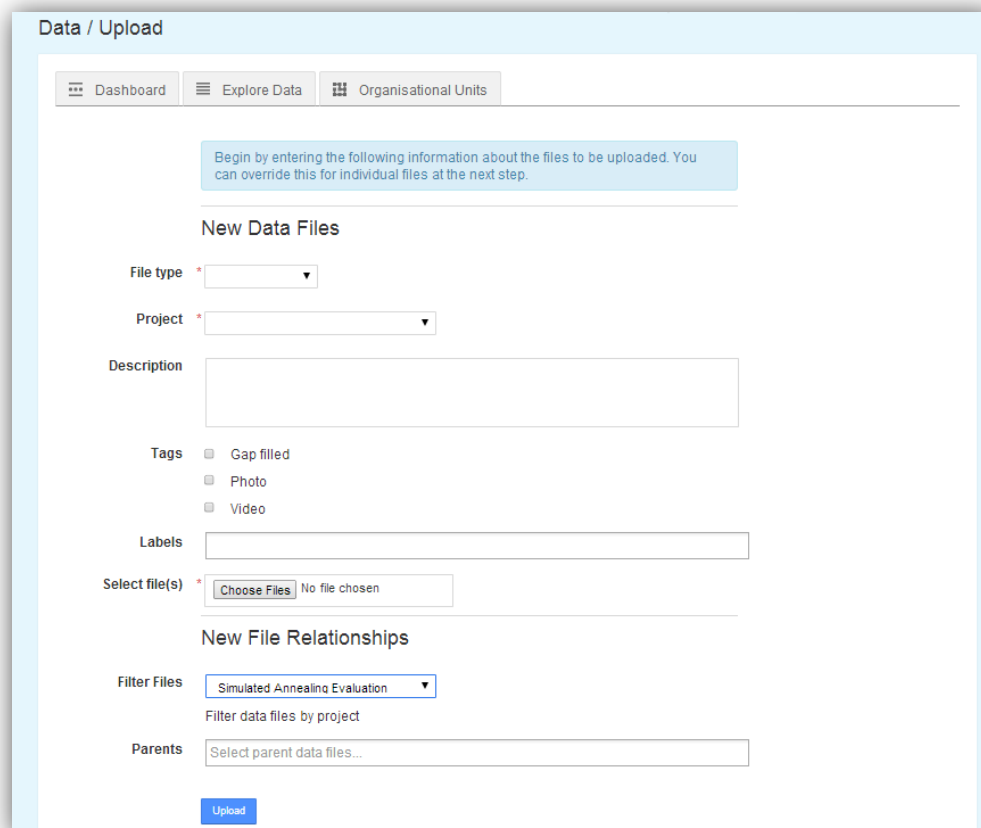
7   Uploading Data Files

To upload one or more new Data Files:

- Click the Upload button which is displayed at the top left of the main screen.
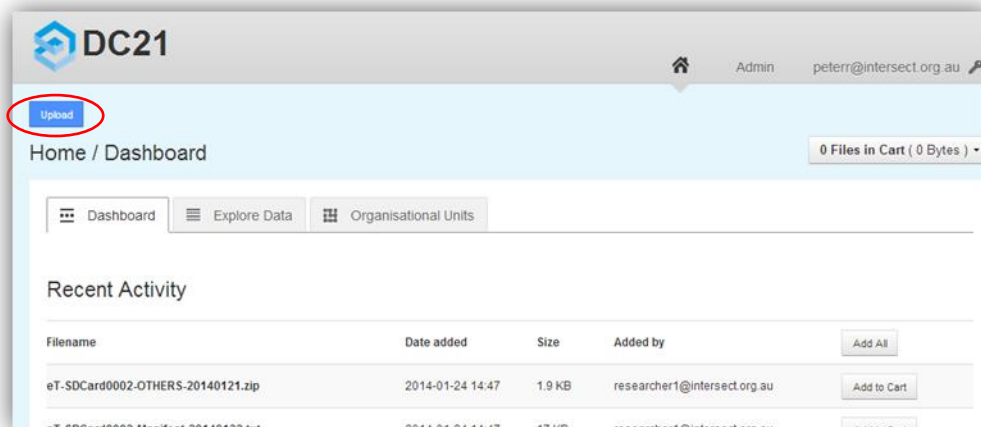
Clicking this button will take you to the New Data Files form.

- Enter all the Metadata that you want applied to the new file or files. See the section *6.2.1 Basic Information* for information about the meaning of Metadata fields. See section *4.3 Entering Labels* for information on Labels and how to enter them.
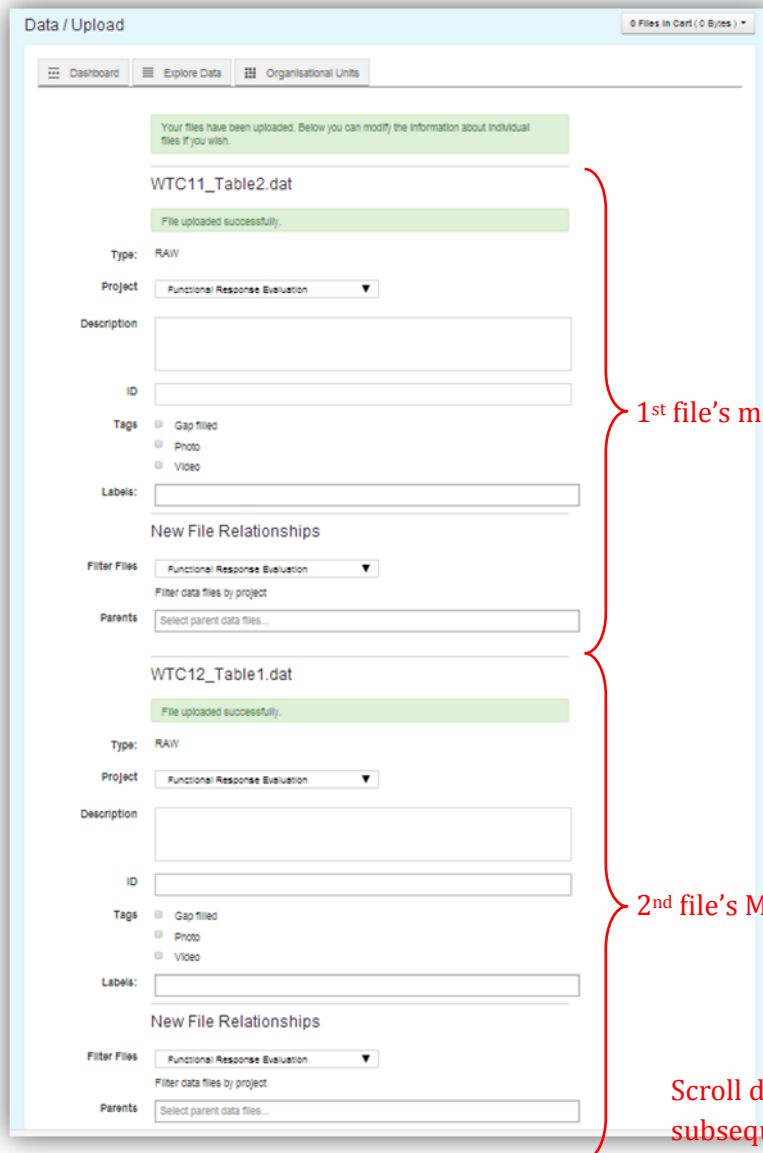
Note	Files that are uploaded with a File Type of **RAW** and are also detected as containing valid TOA5 header information are treated as a special case. See section *7.3 Uploading RAW TOA5 Data Files* for more information.

- Click on Choose Files and use the file dialog which opens to select one or more files to be uploaded. When you confirm your file choice and return to the New Data Files dialog, a second Choose Files button will appear below the first.

- If you require more files to be uploaded, click on the new Choose Files button and use the file dialog to select one or more further files to be uploaded. Repeat this step using the new Choose Files button which appears after each file selection dialog is closed until you have selected all the files you require to be uploaded.

- Click on Upload at the bottom of the screen to cause the files to be uploaded.

After the file or files have successfully uploaded, a screen will be displayed with a separate section for each uploaded file displaying the Metadata applied to that file.

- For each file shown on this screen, adjust the Metadata for that file as you require.

Data / Upload

Your files have been uploaded. Below you can modify the information about individual files if you wish.

WTC11_Table2.dat

File uploaded successfully.

Type:	RAW
Project	Functional Response Evaluation ▼
Description	
ID	
Tags	☐ Gap filled ☐ Photo ☐ Video
Labels:	

New File Relationships

Filter Files	Functional Response Evaluation ▼
	Filter data files by project
Parents	Select parent data files...

1st file's metadata

WTC12_Table1.dat

File uploaded successfully.

Type:	RAW
Project	Functional Response Evaluation ▼
Description	
ID	
Tags	☐ Gap filled ☐ Photo ☐ Video
Labels:	

New File Relationships

Filter Files	Functional Response Evaluation ▼
	Filter data files by project
Parents	Select parent data files...

2nd file's Metadata

Scroll down to see subsequent files' Metadata...

In particular, if the start and end dates and times for the data were not automatically extracted for the file, enter them now if required. See *4.2 Entering Dates and Times* for instructions on entering dates and times.

Also ensure the Labels are correct using the instructions in section *4.3 Entering Labels*.

- Click on the Update button at the bottom of this screen to apply the modified Metadata to the files and complete the upload process.

If an uploaded file has the same filename as another Data File that already exists within the system, DC21 will automatically suffix a unique number to the end of the filename, before the file extension, to avoid the conflict.

7.1 Uploading Image Files

For all uploaded Data Files, DC21 checks if the file should undergo Optical Character Recognition (OCR) processing. To be processed, the following must all be true.

- **Auto OCR on Upload** must be enabled.
- The Data File's MIME Type must match any one of the configured **OCR Supported MIME Types**.
- The Data File name must match the **Auto OCR Regular Expression**.

See *11.5.2 OCR Processing parameters* for information on setting the configuration parameters mentioned above.

If all of the above are true, an empty output text Data File will be created with File Format text/plain and File Type PROCESSED. Its name will be based on the uploaded file with .TXT extension. OCR background processing will be queued immediately after the Update button is clicked.

You will be advised by email when the background OCR process completes. A Parent-Child relationship will be defined between the two files, with the original image or video file set to the Parent and the text file as the Child.

7.2 Uploading Video and Audio Files

For all uploaded Data Files, DC21 checks if the file should undergo Speech Recognition (SR) processing. To be processed, the following must all be true.

- **Auto SR on Upload** must be enabled.
- The Data File's MIME Type must match any one of the configured **SR Supported MIME Types**.
- The Data File name must match the **Auto SR Regular Expression**.

See *11.5.3 Speech Recognition Processing parameters* for information on setting the configuration parameters mentioned above.

If all of the above are true, an empty output text Data File will be created with File Format text/plain and File Type PROCESSED. Its name will be based on the uploaded file with .TXT extension. SR background processing will be queued immediately after the Update button is clicked.

You will be advised by email when the background SR process completes. A Parent-Child relationship will be defined between the two files, with the original image or video file set to the Parent and the text file as the Child.

7.3 Uploading RAW TOA5 Data Files

When a TOA5 CSV file is uploaded and the **Type** is specified as RAW, it will be integrated into the canonical stream of data for the data logger that generated it. 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.

The system automatically determines the start time, end time and data logger that generated the data from the TOA5's header information and uses it to control processing of this file. Other metadata is also extracted from the TOA5's header information.

As a result of this processing:

1. If a 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 extracted and stored as Metadata for the file. Therefore, when TOA5 files are uploaded as RAW data, DC21 does not permit the User to enter Start and End Dates.

DC21 attempts to preserve the Parent-Child relationships of TOA5 files by transferring existing relationships to any replacement file. However, if this would cause a conflict, relationships may be deleted. Greatest priority is given to preserving relationships in which the uploaded TOA5 file is a Parent.

7.3.1 TOA5 Data Upload Action Summary

The following table summarises the upload processing that will be done and the results for various combinations of file type and existing data.

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

7.4   Automating the upload of data to DC21

As well as via the web interface, data can be uploaded to DC21 using an HTTP-based API. The upload of data into the system is facilitated through a Ruby script. On Windows, there is also a .BAT script which wraps this Ruby script.

When Data Files are uploaded to DC21 using the HTTP-based API, the OCR, SR and TOA5 processing is exactly the same as is done for manual uploading. It is dependent on the same settings as described in section _11.5.2 OCR Processing parameters_ and section _11.5.3 Speech Recognition Processing parameters_.

Instructions and downloadable scripts for Windows can be found in the DC21 WIKI documentation on GitHub at https://github.com/IntersectAustralia/dc21-doc/blob/master/README.md. Select the documentation ZIP file for your DC21 version and see the Release Notes in that ZIP file. Also see section _11.5 Tailoring DC21 for Your Organisation's Needs_.

8 Managing Data Files

The key views of the files uploaded into the DC21 database are the Dashboard and Explore Data views. These views allow you to perform the key functions of DC21.

Access these views by clicking on their respective tabs on the DC21 Home Screen.
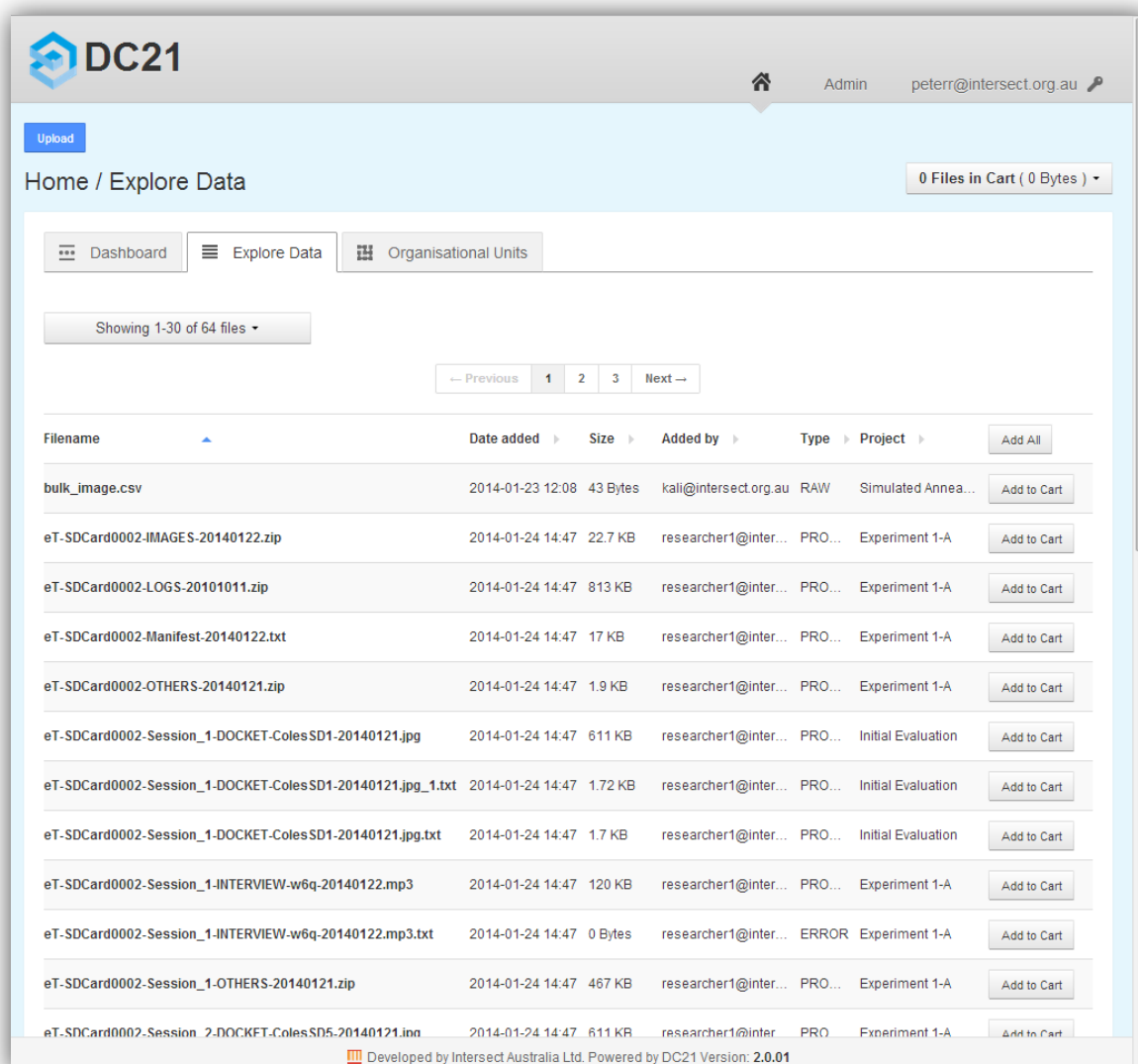
8.1 The Dashboard Tab

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 all Users of DC21.
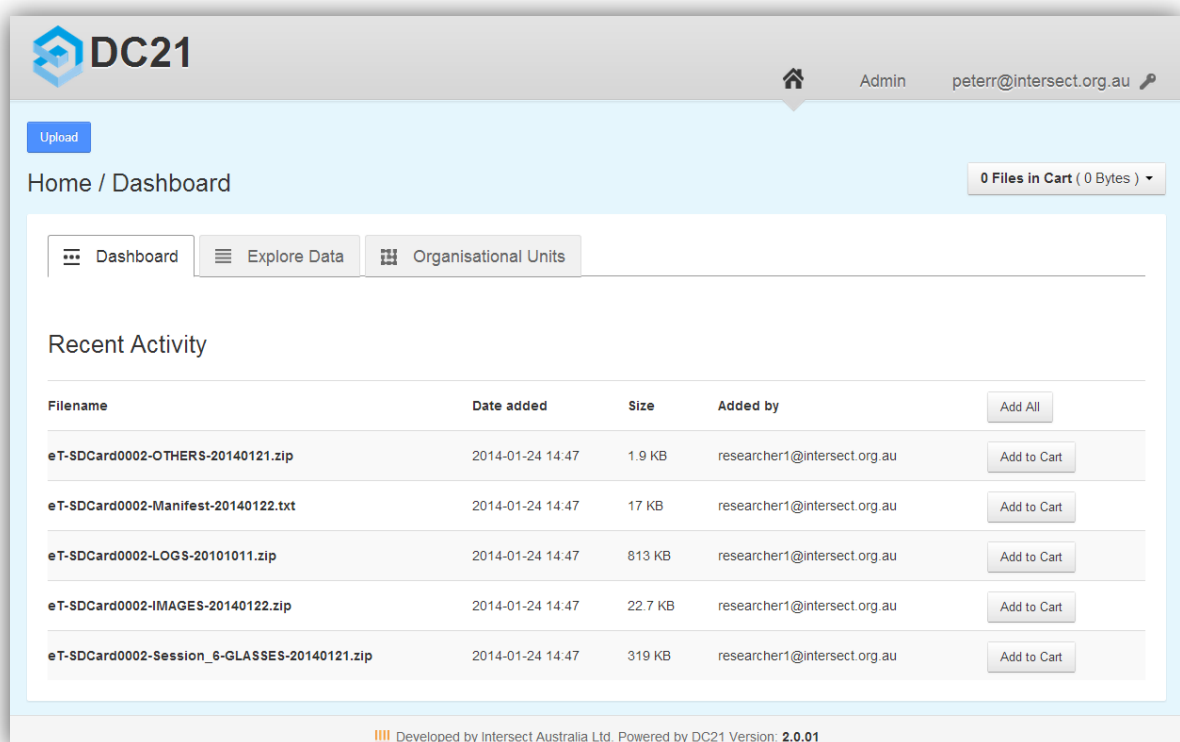
The operation of the Dashboard tab is similar to the operation of the Explore Data tab, except that searching and sorting functions are not supported. See the description of the Explore Data tab below for more information.

8.2 The Explore Data Tab and File Searching

The Explore Data tab provides the main data management functions of DC21. 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.

8.2.1 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 the right of any column heading indicates the active sort order. A grey right-pointing arrow indicates that the file list is not sorted by that column.
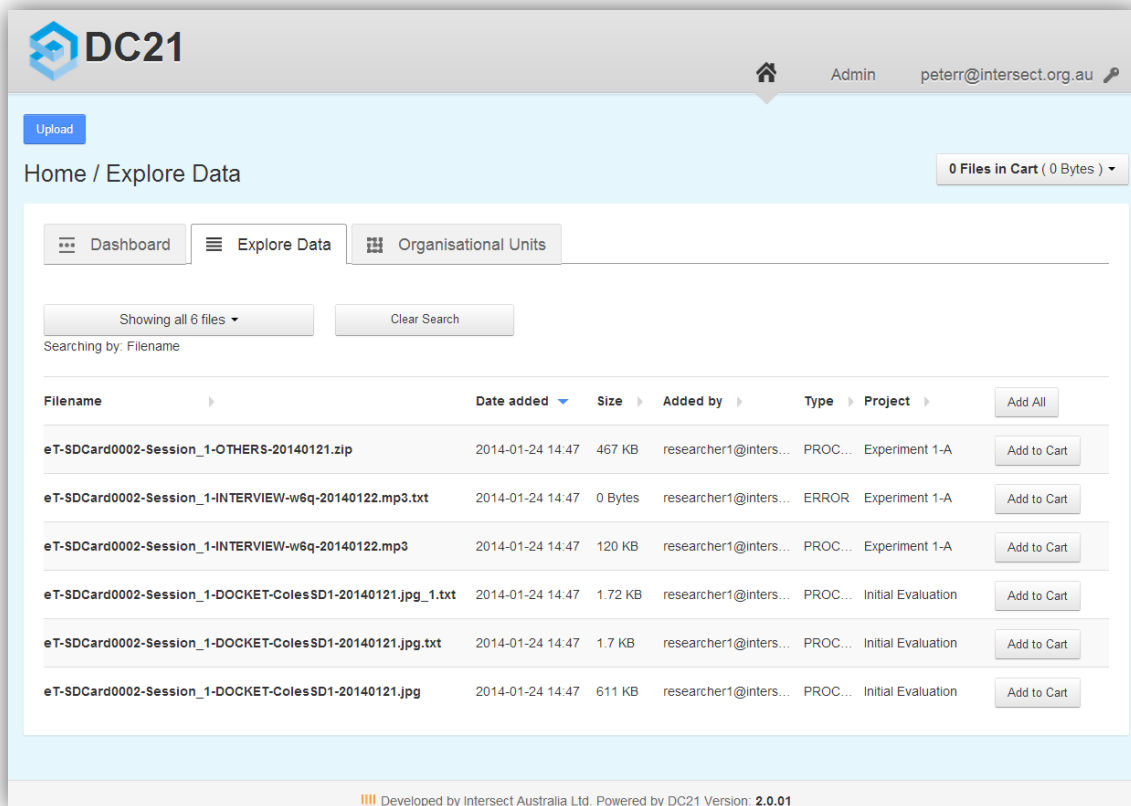
Re-sorting the list always resets the display to the first page of the file list.

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

8.2.2 Searching

When the number of files uploaded to your system becomes large, finding the files 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 set or 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.
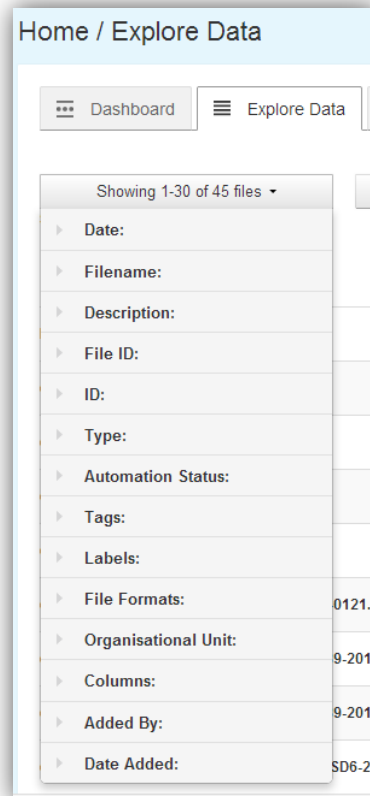
If there is a search active, one or more of the search categories in this menu will be expanded when you first display it. These are the search categories which have active search data.

Click on the Metadata field by which you wish to search to expand it.

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.

Click on any Update Search Results button to cause DC21 to redo the search based on all the criteria you have entered. These buttons will appear when you open any of the Metadata search fields.

8.2.2.1 Regular Expressions

Regular expressions are used for searching for specific substrings in general text. They are used widely across many computer systems. DC21 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.

DC21'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.
		Example: `mpl` will match the string `example` at the fourth character.
Start of string	^	The character ^ will match the beginning of the string.
		Example: `^exa` will match `example`, but `^xa` will not.
End of string	$	The character $ will match the end of the string.
		Example: `le$` will match `example`, but `pl$` will not.
Any character	.	The period character will match any character.
		Example: `a.c` will match `abc`, `aac`, `adc`, `a7c` and `a-c` but it will not match `ac` or `abbc`.
Repeated character	*	Asterisk causes matching to zero or more repetitions of the preceding character.
		Example: `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 zero characters between the `a` and `c` in those strings.
Repeated characters	+	The plus sign causes matching to one or more repetitions of the previous character.
		Example: `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.
		Example: `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.
		Examples: `[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.

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.

Examples: `\\` will match `\`

`\.` will match `.`

`\*` will match `*`

`\[` will match `[`

Combinations　　　Any of the above search methods can be combined.

Examples: `^.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 occurs between the digits `1` and `2`.

`[\[\]]` will match either `[` or `]`.

`\.+` will match any run of periods.

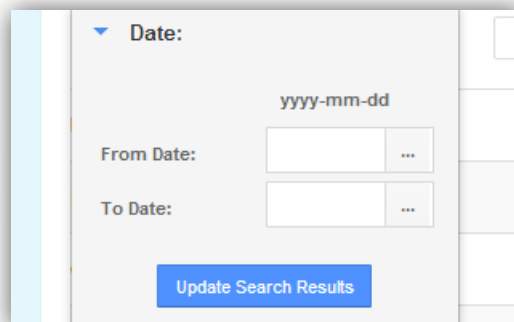`[0-9]+` will match any integer number.

`[0-9]+\.[0-9]*` will match any number with a decimal point.

If an invalid regular expression is entered, DC21 will place an error message at the top of the screen, clear the search field and ignore the regular expression.

8.2.2.2　Restricting by Date

The **Date** field allows you to search for files based on the start and end date specified in the file's Metadata.
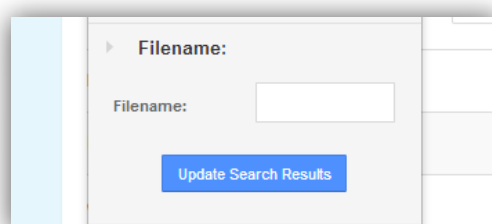
Dates can be entered in either the **From Date**, **To Date** or both. If only a **From Date** is specified, all files containing data on or after that date will be included. If only a **To Date** is specified, all files containing data for before or on that date will be included. See *4.2 Entering Dates and Times* for instructions on entering dates.

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

For TOA5 files, this search option checks the Start and End Dates in the Information from the File (see section *6.2.2 File Relationships*) and for all other files, this search option checks the Basic Metadata Information (see section *6.2.1 Basic Information*).

8.2.2.3   Restricting by Filename substring

The **Filename** field allows you to search for files based on their filename.
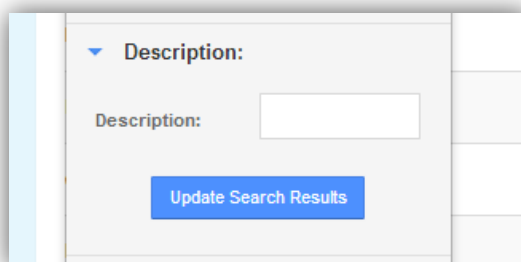
DC21 treats the search string as a regular expression. See section *8.2.2.1 Regular Expressions for more information.*

8.2.2.4   Restricting by Description substring

The **Description** field allows you to search for files based on their free-form text descriptions.
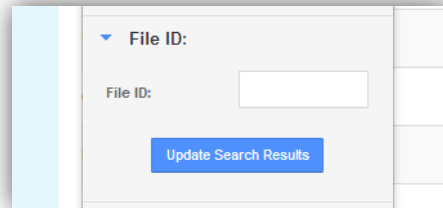
DC21 treats the search string as a regular expression. See section *8.2.2.1 Regular Expressions* for more information.

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

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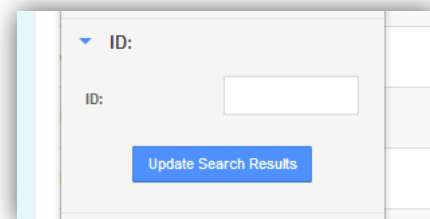
Enter an integer number into the File ID search field to display only the file with that File ID. It is not usual to use this search method with any other search criterion, as this method will always display exactly one file, or no files if there is no file with the entered File ID.

8.2.2.6 Restricting by ID

The **ID** field allows you to search for files based on their entered ID.
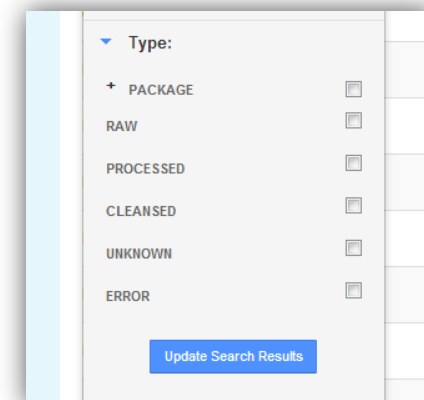
DC21 treats the search string as a regular expression. See section *8.2.2.1 Regular Expressions* for more information.

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

8.2.2.7 Restricting by File Type

The **Type** search parameters allow 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 – that is, files will not be filtered based on their Type. When 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** itself 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 Published status, leave both checkboxes unchecked. If a **Published Date** is entered, then only files Published on that date will be displayed. See *4.2 Entering Dates and Times* for instructions on entering dates.

8.2.2.8   Restricting by Automation Status

The Automation Status interface allows you to restrict your search to files with specific background processing status. Background processing is used for creating Package files, OCR processing and SR processing.

Like the Type interface, selecting none of the available checkboxes means that all files will be returned in the search results regardless of their Automation Status. Once a checkbox is selected, only files that have that status will be listed.

These status values apply only to the *output* files from background processing. That is, they apply only to output Package .ZIP files, OCR .TXT files and SR .TXT files. Files which were *input* to these processes are not selected by these options.

If you select any one of these options, files which were not the subject of background processing will not be listed. Package files which were executed without background processing will also not be listed.

The Status meanings are:

COMPLETE Those output files which have been queued for background processing and have completed successfully

FAILED Those output files which have been queued for background processing and have failed to completed successfully

WORKING Those output files which are currently being created in a background process

Files listed by this option will have a zero file size, which is not final.

The current implementation of DC21 supports only a single simple FIFO queue and processes only one job at a time, so this option will not list more than one Data File.

QUEUED Those output files which are still in the queue waiting for background processing

These files will have zero file size.

If more than one Status is selected, all files with any of the selected Statuses will be listed.

8.2.2.9   Restricting by Tags

The **Tags** interface allows you to search based on the tags that have been assigned to a file.
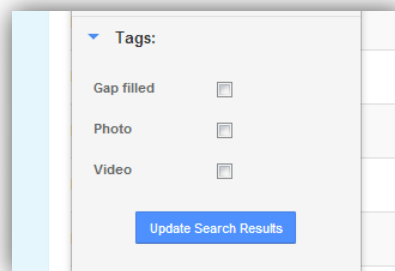
Like the Type interface, selecting none of the available checkboxes means that files will be returned in the search results regardless of the tags they have. Once a checkbox is selected, only files that have the corresponding tag will be listed.

If more than one Tag is selected, all files with any of those tags will be listed.

DC21 does not support searches for Data Files which do not have a specific Tag or Tags, or which do not have any Tags set.

8.2.2.10   Restricting by Label

The Labels interface allows you to search for files which have been assigned specific Labels.
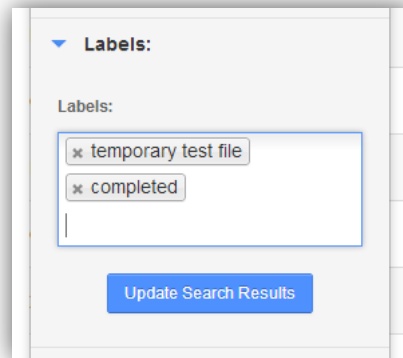
Clicking in the Labels search control displays a dropdown list of all Labels which are assigned to Data Files in your system's Working Store. Type characters to restrict this list to only those Labels which contain the characters you type. Select the Labels you want to add to your search using the mouse or the arrow and Tab keys. You can remove Labels from the control by clicking the X associated with them.

If you leave this box empty, the search will not be restricted by Label.

DC21 does not support searches for Data Files which do not have a specific Label or Labels, or which do not have any Label.

8.2.2.11 Restricting by File Format

The File Format interface allows you to search for files of specific MIME Types or DC21 File Formats, such as BAGIT, TOA5 or UNKNOWN.
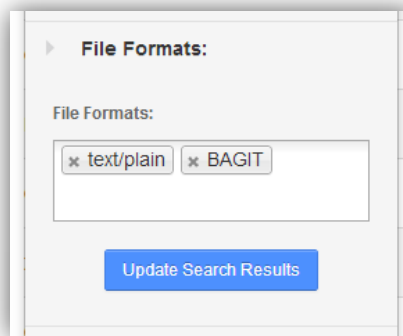
Clicking in the File Formats search control displays a dropdown list of MIME Types, augmented with the DC21 defined File Formats. See *6.2.1 Basic Information* for information about File Formats, which is part of the Metadata Basic Information. Select the File Types you want to add to your search. You can remove File Formats by clicking the X associated with them.

Leaving this box empty will include all File Formats in your search output.

8.2.2.12 Restricting by Organisational Unit

The **Organisational Unit** (which will be replaced by the term specific to your system) interface allows you to search for files based on the Organisational Unit or Project that produced the file.
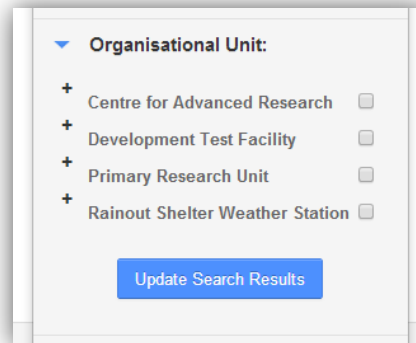
The interface is a two-level hierarchy of checkboxes. The top level shows all the Organisational Units defined in the system and the second level shows the Projects that are defined for those Organisational Units. Selecting an Organisational Unit selects all of the Projects for that Organisational Unit. If only specific Projects are required, clicking on the **+** sign to the left of a Organisational Unit will expand the hierarchy and allow individual Projects to be selected or deselected.

Selecting more than one Organisational Unit or Project causes all files for any of those Organisational Unit or Projects to be listed.

8.2.2.13 Restricting by Data File Columns

The **Columns** interface allows you to search for TOA5 format Data Files that contain specified columns.
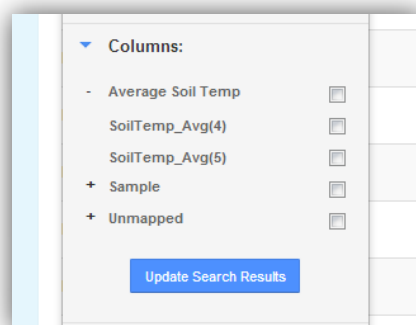
Like the Organisational Unit interface, this shows a two-level hierarchy of checkboxes. The top level contains all the Names that are defined in the Columns Mapping table. When these Names are expanded by clicking the **+** sign next to them, they show all of the TOA5 column headings which map to these Names. The last top level group is an extra group called **Unmapped** that contains all the TOA5 column headings that appear in any file in the entire

DC21 data store but are not mapped to a standard Name in the Column Mappings table. See *11.3 Managing Column Mappings* for more information.
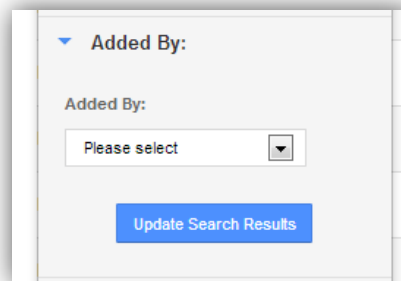
Selecting the checkbox for a standardised top-level Name will select all the TOA5 column headings that are mapped to it. Clicking the **+** sign to the left of the top-level Name will allow you to select them individually.

Selecting more than one checkbox will cause any file which has any 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.

8.2.2.14   Restricting by Person Who Added the File
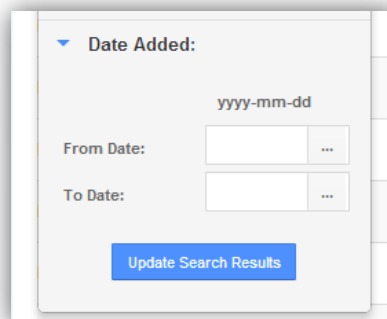
The **Added By** field allows you to search for files that were uploaded by a specific User.

The User must be selected from the list of a Users registered in the system.

8.2.2.15   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. See *4.2 Entering Dates and Times* for instructions on entering dates.

8.3   The Cart

The Cart operates like an e-Commerce shopping cart. DC21 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 10 *Downloading files* and Publishing in Chapter 9 *Publishing Your Data*.

Unlike most e-Commerce systems, 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 only one Cart per User account. If two people simultaneously use the same login account, the results can be unpredictable.

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 list.

The Cart status box, circled in red below, shows the number of files in your Cart and the total size of these files. It appears at the top right of many DC21 screens.
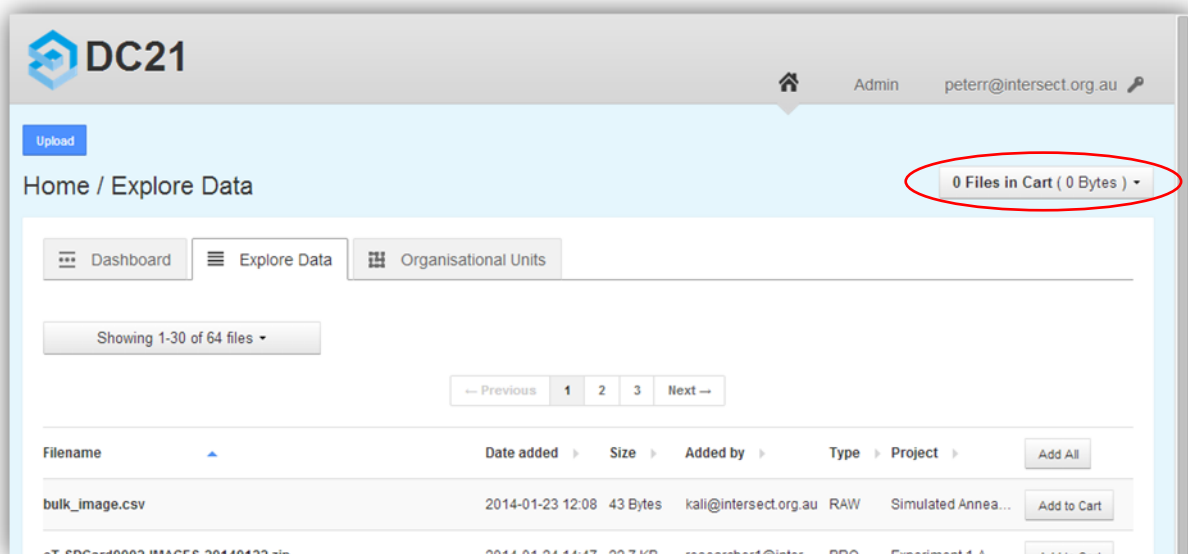
Click on the Cart status box to show a dropdown menu of operations which can be performed on the files in the Cart. These operations are:

Download   Click on this option to download Data Files to your local computer. See Chapter 10 *Downloading files* for instructions on using this feature.

Package   Click on this option to create a publishable Package containing all files in the Cart. See section *9.1 Creating a Package* 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 _8.3.1 Editing the Cart_ for more details.

8.3.1   Editing the Cart Contents

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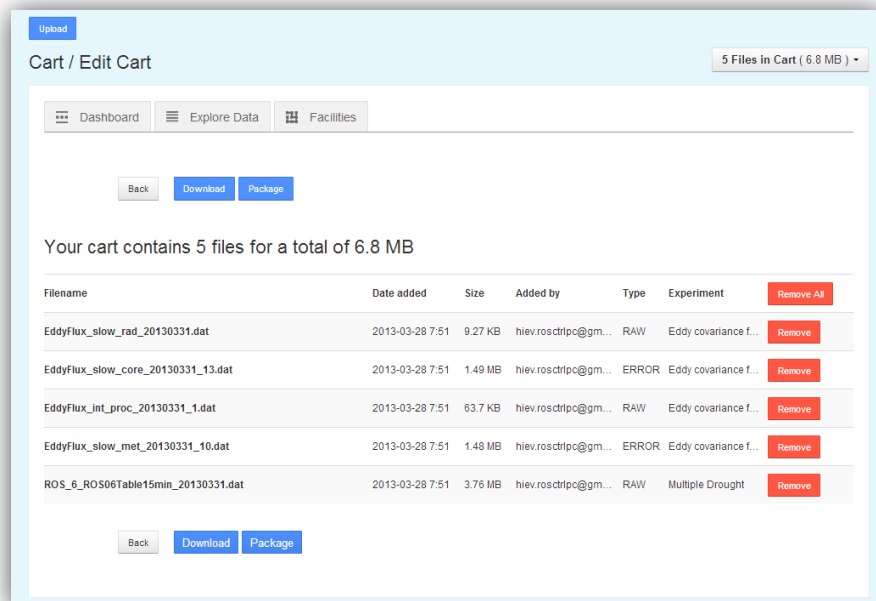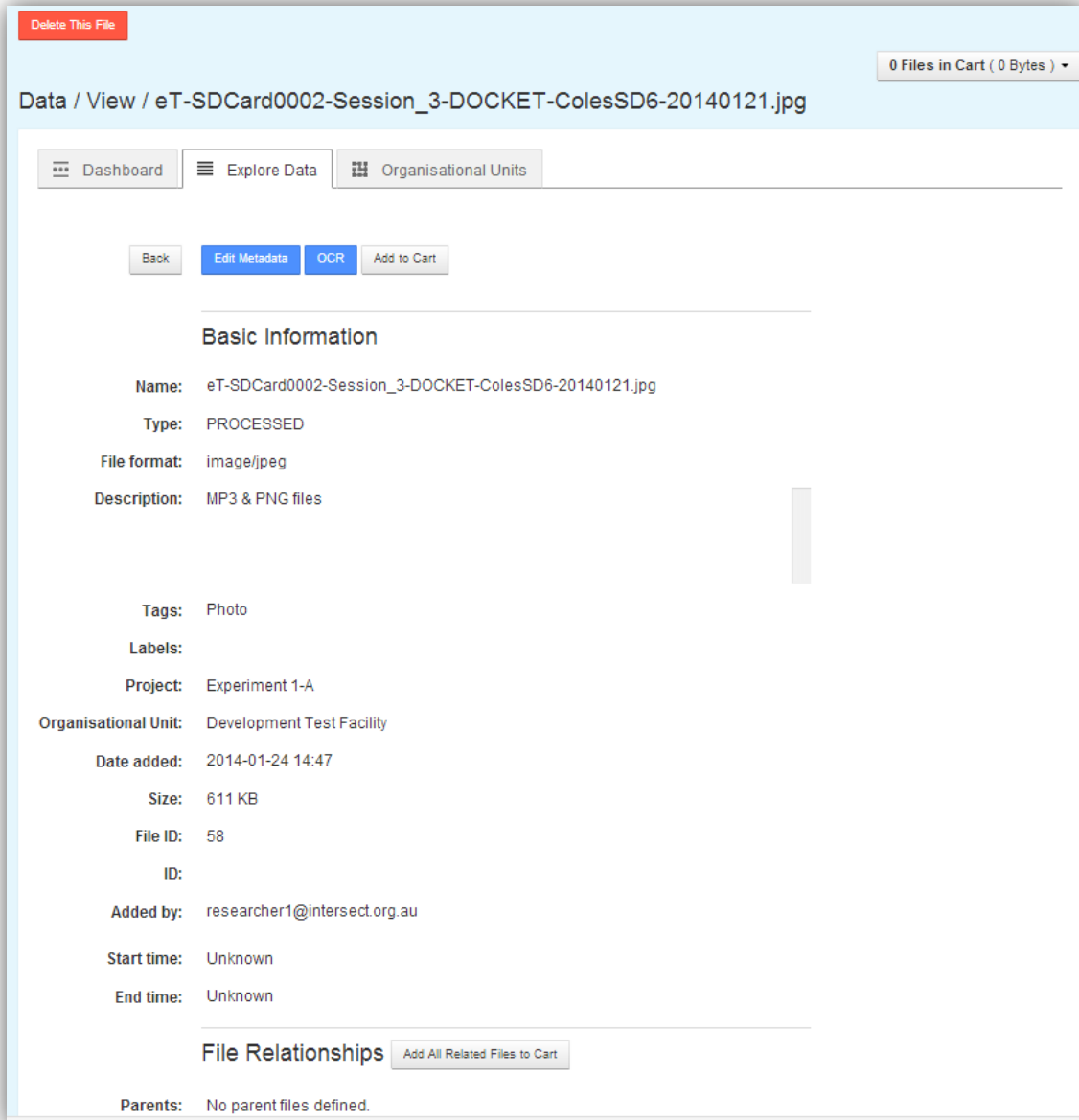
There is one line in the table on this screen for each file in your Cart. These files can be removed individually by clicking on Remove on the relevant line. If the Remove All button is clicked, the Cart will be emptied and further Cart operations cannot be performed until files are again added to the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

Download and Package buttons are also available on this screen. See section _9.1 Creating a Package_ and Chapter 10 _Downloading files_ for more information about the operations initiated by these buttons.

8.4 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, and the **File Relationships** section, although some fields are not shown for some file types. TOA5 files also show **Information From The File** and **Columns** sections.

There are a number of buttons which may appear on this screen, depending on the type of Data File and its status.

Back   Click this button to return to the list of Data Files.

Add to Cart   This button only appears if the Data File is not already in your Cart.

	Click it to add the file to your Cart.
Remove from Cart	This button only appears if the Data File is already in your Cart. Click it to remove the file from your Cart.
OCR	This button only appears if the Data File has one of the image file MIME Types as defined in your system's configuration parameters. (See section *11.5.2 OCR Processing parameters*.)
	Click this button to start a background Optical Character Recognition process. The output of this process will be written to a new text Data File. The name of the new Data File will be based on this Data File's name, but with a .TXT extension. The Data File will immediately appear in the file lists in the Explore Data tab. However, its size will be zero until it has been processed. Check the Creation Status to see when the background processing has completed.
	Users can click the OCR button even if they do not own this Data File. They will own the output .TXT file.
SR	This button only appears if the Data File is one of the video or audio file MIME Types as defined in your system's configuration parameters. (See section *11.5.3 Speech Recognition Processing parameters*.)
	Click this button to start a background Speech Recognition process. The output of this process will be written to a new text Data File. The name of the new Data File will be based on this Data File's name, but with a .TXT extension. The Data File will immediately appear in the file lists in the Explore Data tab. However, its size will be zero until it has been processed. Check the Creation Status to see when the background processing has completed.
	Users can click the SR button even if they do not own this Data File. They will own the output .TXT file.
Edit Metadata	This button appears if you have permission to edit this Data File's metadata. 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.
	Clicking this button will take you to a form that allows you to modify the file's Metadata.
Parents and Children	In addition, clicking on the name of a file in the Parents or Children related file lists will jump to the Metadata View screen for that file.

When you click on Edit Metadata , a screen similar to the one below is shown.
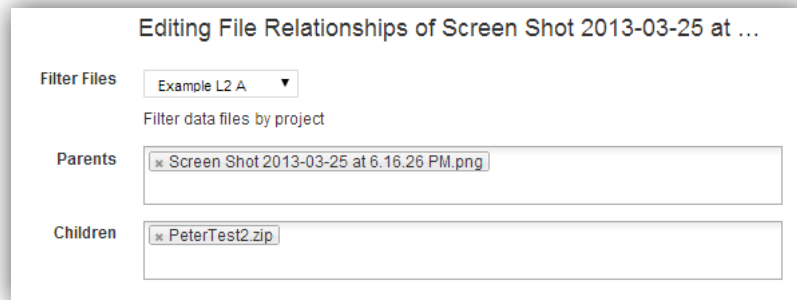
The parts shown on this screen will depend on what type of file you have selected.

Those fields shown on this screen without boxes cannot be modified. They are fields which are managed by DC21. Those fields marked with * must not be left blank.

The Labels assigned to this file can be modified on this screen. The method of modifying is the same as that for entering labels. See section *4.3 Entering Labels* for instructions.

Dates or Times can be modified using the method described in section *4.2 Entering Dates and Times*.

The Parent-Child Relationships in which this file participates can also be changed on this screen.
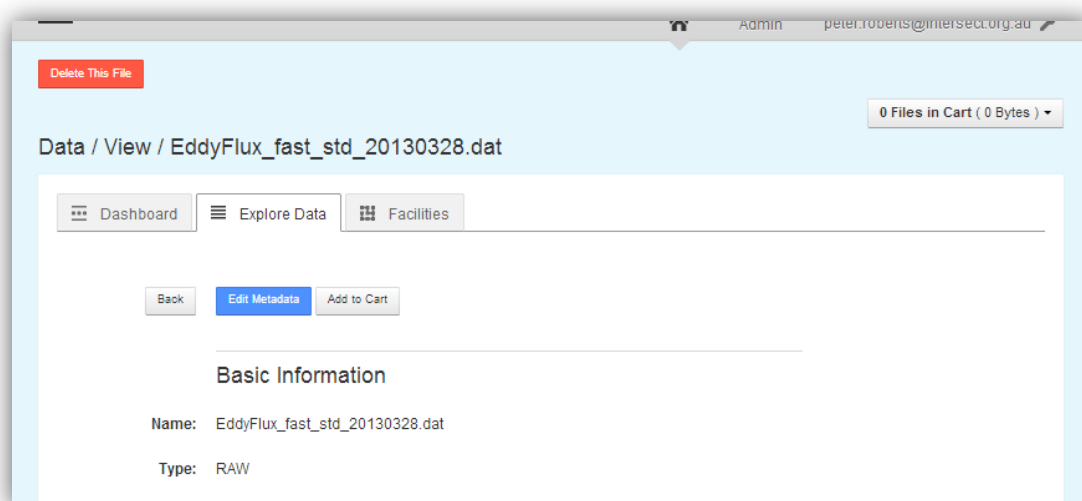
Remove a Relationship by clicking on the X within the relevant filename box. Add a relationship by clicking within the Parents or Children boxes with the mouse and starting to type the related file name. Once you have typed three characters, all data files within the system which match those characters will be shown in a dropdown list. Select the one you require with the mouse or the arrow and tab keys. Adding the name of a file into one of these relationship boxes will also cause the inverse relationship to be recorded in that file's Metadata.

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

8.5   Deleting a Data File

Note   Deleting a file removes it completely from DC21. *It is irreversible.*

In order to delete a file from the DC21 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 Delete This File button will appear at the top of the Metadata View screen.

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

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 created, or you have Administrator permission.
Note	In versions of DC21 before Version 2.0, it was not possible to delete a file with a non-null ID without first clearing the ID field. This restriction has been eliminated.

9   Publishing Your Data

When enough data has been produced from a facility to warrant publishing or a researcher wants to publish their research paper and site their research data package, the data collected can be Published to an institutional repository and/or the Australian Research Data Commons.

Before Publishing, the Data Files to be published must be combined into a single Package file. Package files are ZIP files which use the Bagit format, which is described in *Appendix A - The Bagit format*.

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 via the OAI-PMH protocol. After this has occurred, it will become discoverable in Research Data Australia.

The process of Publishing involves a few steps:

- The User creates a Package which contains the Data Files to be Published using the DC21 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 DC21's Publish function. This copies the RIF-CS file into a specific 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 DC21 Metadata when or if this has occurred.

Note	Files in a Packaged ZIP file are copies of the files in DC21. Changing the files in DC21 will not affect the files in the Packaged ZIP file.
Note	The exact details of the harvesting operation may vary from site to site. At some sites, the Packaged ZIP file may not be copied during harvesting and may be accessed directly by external data users.

> Note   The RIF-CS file is created when the Package is created, and the RIF-CS file is the file that is Published. Therefore, changes to the Package Metadata after creation do not affect what is Published.

9.1   Creating a Package

Creating a Package creates two related components:

Bagit ZIP 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. This Readme.HTML file is both human readable and machine readable. See *Appendix A - The Bagit format* for more information.

Matching RIF-CS File   This file contains a copy of the Package Metadata which is entered at the time the Package is created. Only selected parts of the Metadata from the associated Experiments and Facilities are copied into the RIF-CS file. See *Appendix B - RIF-CS* for more information.

To create a Package containing one or more files:

- Determine the external Package ID that will be assigned to this Package. Ask your site's Application Manager for information on how this ID is determined on your site.

- Add the required file or files to your Cart, ensuring that the Cart contains only those files you wish to include in your Package. See section *8.3 The Cart* for instructions on using the DC21 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 Package on that screen. If this method is used, the Cart can be reviewed prior to Packaging.

- The New Package screen will be displayed as follows.

- Enter the Metadata to be associated with your Package. See section *6.2.1 Basic Information* for details of this Metadata. Note that you cannot enter a value for the ID field – it will be automatically generated when the package is created.

- Choose whether you want to run the package creation as a background task by checking the **Run in background?** checkbox (this is the default).

  You should run the package creation as a background task if you think it might take a long time to create the package. You will then be able to continue other DC21 operations without waiting for the processing to complete. You can monitor the progress of the creation by refreshing the browser window and looking at the "Creation Status" and "Package progress" Metadata fields. You will receive an email when the background processing is completed.

You can uncheck the checkbox but this will mean you will have to wait until the package is created before you can continue to use the DC21 application.

- Click on `Create Package` to cause your Package file to be created and saved. The ID field will be automatically generated as part of the package creation process. The Package file can now be viewed in the Explore Data tab. Note that if you click on `Back`, you will be returned to the Explore Data tab and the Package will not be created.

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

Note	If you do not run a large packaging task in background, the DC21 server will allocate the maximum permitted processing capability to this task until it is completed. This could result in all server processor cores becoming dedicated and stalling the DC21 server for all other users, especially if a number of users simultaneously do this.
Note	It is very important to check the Package Metadata before clicking on `Create Package`. This Metadata, including selected Metadata from the included Data Files and the Facilities and Experiments they reference, 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.
Note	If a background task fails to complete, there is an administration function available to see more information and to kill it if necessary. See section *11.4 Managing Background Tasks* for details.

9.2   Publishing a Package

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

Publishing can be done by Users with Administrator permissions only.

To Publish a Package:

- View the Package's Metadata by clicking on its filename on either the Dashboard tab or Explore Data tab to see the following.
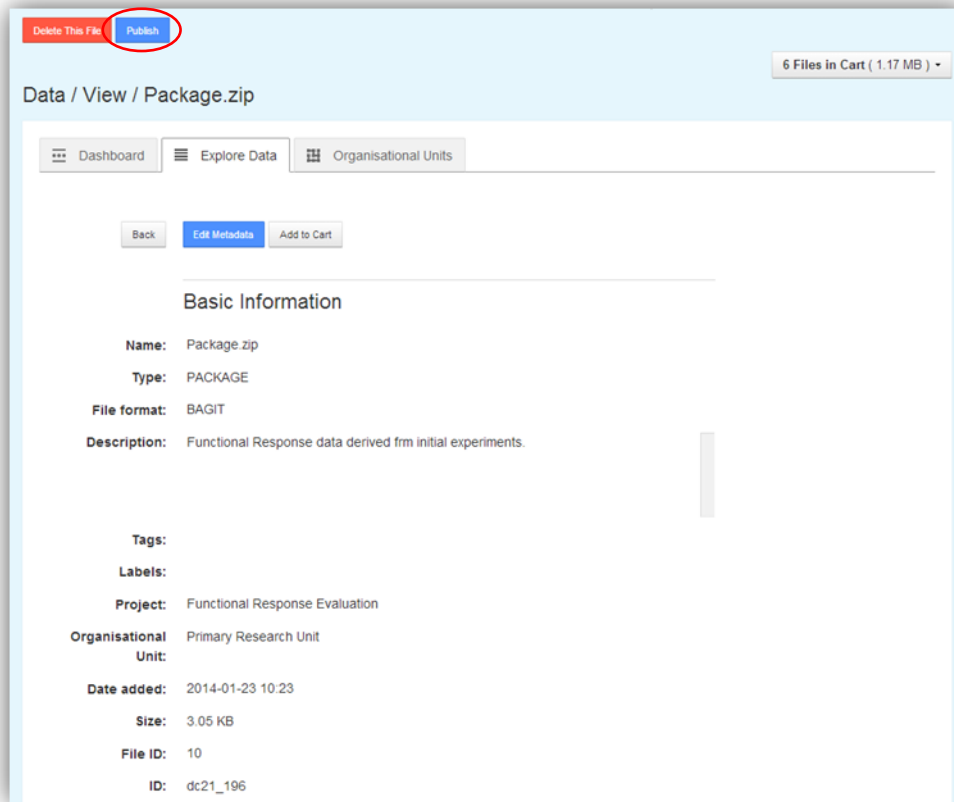
- Review its Metadata to ensure you have selected the correct Package file and that it is ready to Publish.

- 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 **Cancel**, you will be returned to the Package Metadata screen.

9.3 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 described in this section.

Once Packages are Published using the DC21 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 DC21 if the Package has already been harvested or not.

9.3.1   Publishing a second time

DC21 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 original version has already been harvested. The exact results will depend on the way your system is configured.

9.3.2   Deleting Published Packages

It is possible to use the Delete function on any Package file that you have created. (If you have administration privileges, you can also delete the Package files created by others.) However, the actual package file and associated RIF-CS, while no longer visible through the DC21 user interface, are not physically deleted from the system. Instead, they are moved to an archive directory on the DC21 server. If necessary, a system administrator can recover these files for you.

If you delete a Package before it is harvested, the Package file and corresponding RIF-CS will be archived and no longer available for harvesting. This effectively undoes the Publish function and the Package will not be harvested.

If you delete a Package after it has been harvested, the Package file and corresponding RIF-CS will be archived, but it will not affect any already harvested version of the RIF-CS and Package. If you subsequently don't want the deleted package to be available externally, you will need to have it manually removed from the application which harvested it, e.g.: a local metadata store and/or the RDA website.

Note   On sites where the harvesting process copies only the RIF-CS file and external users reference the Bagit ZIP directly on DC21, you must be wary of deleting a Published Package as it will invalidate the links in the harvested RIF-CS file.

9.3.3   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 and re-creation depends on whether harvesting has already occurred (see previous section).

9.3.4　Correcting Published Packages

If a Package is Published incorrectly and you know it has not been harvested, deleting it is sufficient to avoid the Publishing of the incorrect data.
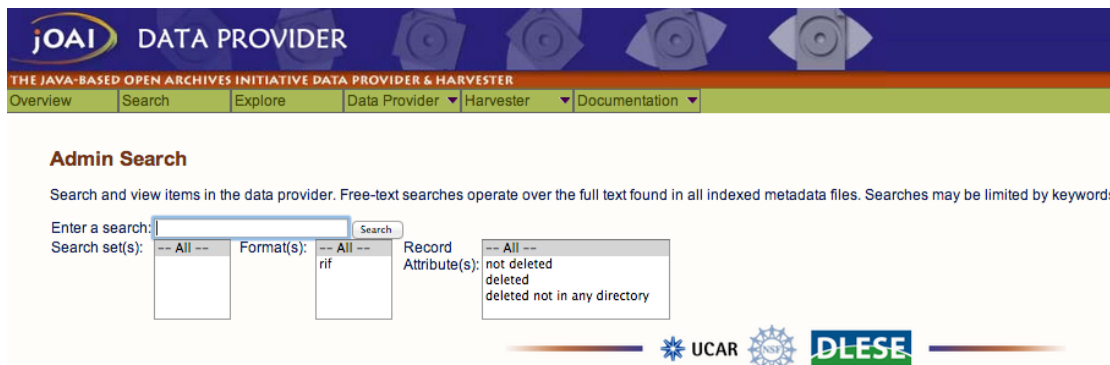
If it has been harvested, deleting it will have no downstream effect. It will be necessary for you to contact the administrator of your Published data store and ask for its removal.

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

9.4　Viewing Published data

DC21 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. See you Application Administrator for these details and also the URL of your DC21 server.

The descriptions of published Packages can be viewed by going to the jOAI web interface at **http://<**your.DC21.server>**/oai/admin/query.do** and performing a search.

Clicking the **Search** button with the search field blank will show all published Packages.

rif-cs-5

ID: rif-cs-5
Native file format: rif
File location: /data/dc21-data/published_rif_cs/rif-cs-5.xml
File last modified: Apr 30, 2012 1:26:12 PM EDT
OAI datestamp: Apr 30, 2012 1:37:13 PM EDT (2012-04-30T17:37:13Z)
OAI identifier: oai:jp-dc21-staging.intersect.org.au:rif-cs-5
Status: Ready for harvest.
Available formats:
   rif: [view | validate | GetRecord]

rif-cs-4

ID: rif-cs-4
Native file format: rif
File location: /data/dc21-data/published_rif_cs/rif-cs-4.xml
File last modified: Apr 30, 2012 1:24:30 PM EDT
OAI datestamp: Apr 30, 2012 1:37:13 PM EDT (2012-04-30T17:37:13Z)
OAI identifier: oai:jp-dc21-staging.intersect.org.au:rif-cs-4
Status: Ready for harvest.
Available formats:
   rif: [view | validate | GetRecord]

10   Downloading files

DC21 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 *8.3 The Cart* for instructions on how to add Data Files to your Cart. Take care not to select too much data for one download. Depending on the speed of your data link, it could take a very long time to download.

- Click on the Cart status box to open the Cart dropdown menu.
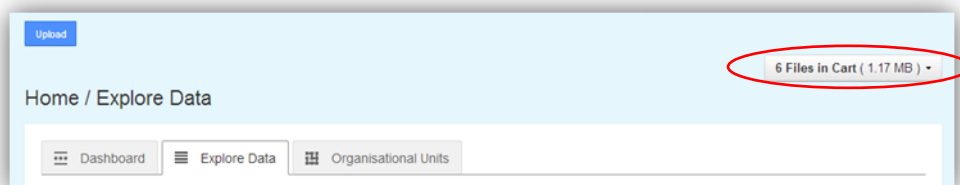
- Select **Download** from the dropdown menu. Alternatively, select **Edit Cart** from the dropdown menu and then click on the Dowload button on that screen, as shown below. If this method is used, the Cart can be reviewed prior to Downloading.

- A file dialog box will open. Navigate to the sub-directory on your local computer 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.

Note   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 as much of the Metadata for each of the files as is included in that Packaged ZIP file. See section *9.1 Creating a Package* for instructions on creating a Packaged ZIP file and *Appendix A - The Bagit format* for details of the Bagit format, which is used for Packaged ZIP files.

11 DC21 Administration
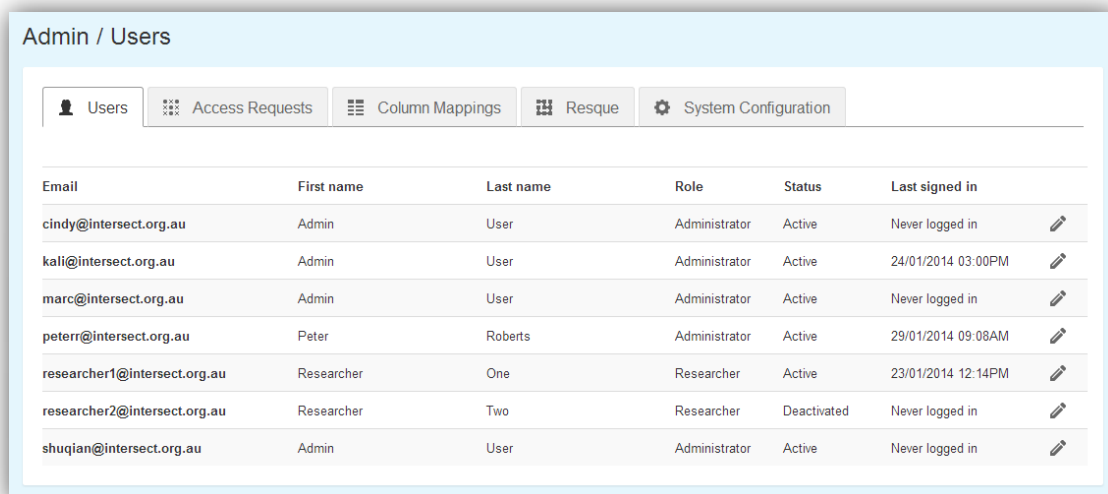
When a User account is created it is associated with a role within DC21. This role dictates what permissions the User has within the system. The most powerful role a User can be given within the system is that of the Administrator.

Only Administrators have access to the **Admin** link at the top right of the main screen. It accesses the Admin section, which has five tabs:

Users	This tab is used for managing the details of the Users who can access the system.
Access Requests	This tab is used for processing requests to access the system made by new Users.
Column Mappings	This tab is one method of managing the list of defined Column Mappings.
Resque	Resque is the background process manager for Ruby on Rails. This tab shows a list of background tasks and allows them to be managed.
System Configuration	DC21 can be configured for your organisation's needs. This tab gives access to those configuration parameters, except for changing the organisation icon displayed on each screen. That icon is set during installation. See the Installation Notes and Deployment Guide for instructions on setting this icon.
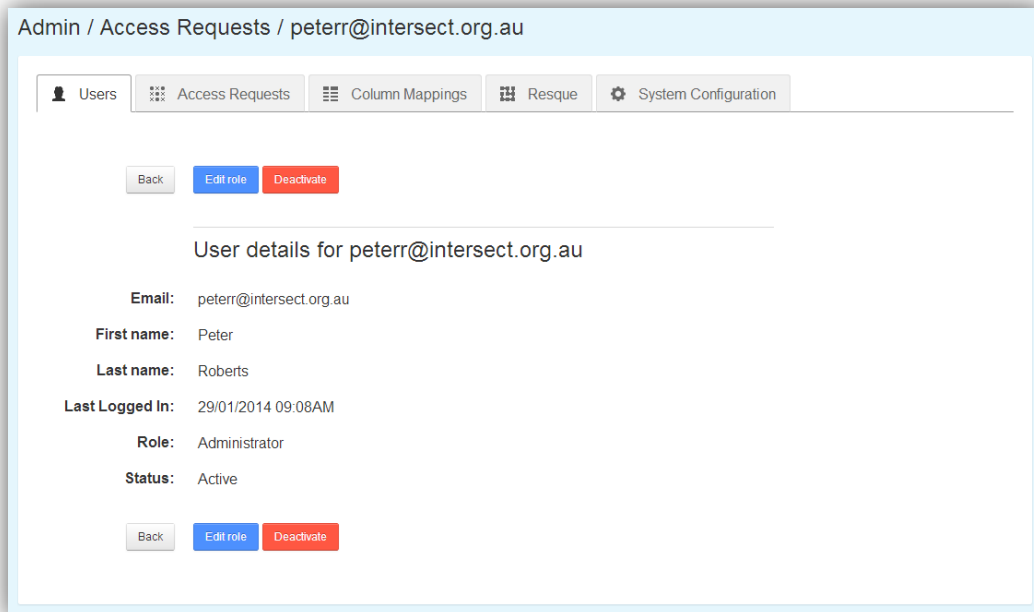
11.1 Managing Users' Details

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.
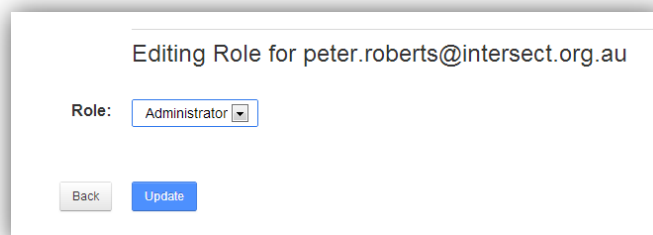
Two functions are provided:

Deactivate Disable the account from being used to login to the system. No data uploaded by the User will be deleted. This button is not displayed for deactivated Users.

Activate Re-activate a User's account which has been previously deactivated. This button is not displayed for active Users.

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 *3.1 Classes of Users* for information on roles.

You can click on the pen ✎ at the right of each User's entry on the Users list to jump directly to this dialog.

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.

11.2 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.
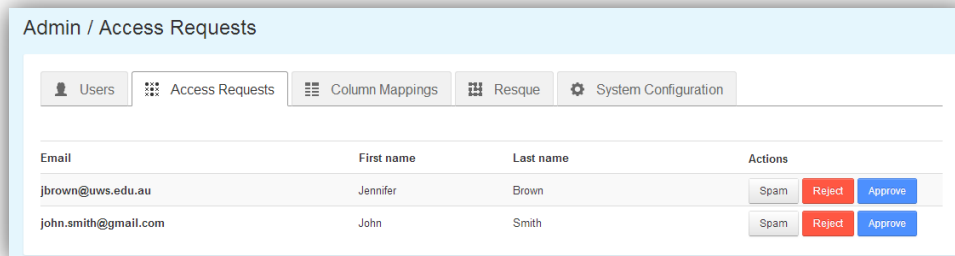
Each access request line in this table has three buttons:

Spam	Click on this button to ignore the account request and remove it totally from the system. All future requests from this email address will be blocked.
Reject	Click on this button to reject the access request. An email informing the User that his or her request for an account has been rejected will be sent to the email address listed.
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 *3.1 Classes of Users* 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.

11.3 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.)

(TOA5 files are the output from Campbell Scientific Data Loggers. Not all DC21 sites will use this device, so this section may not be relevant to you.)

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 DC21 at installation. In addition, further mappings can be added as they are needed.

These Columns Mappings are used by DC21 for two purposes:
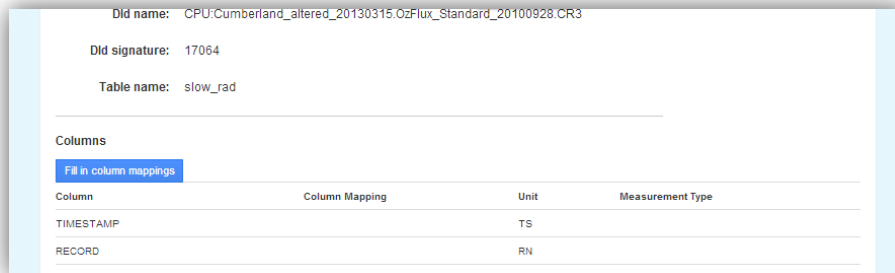
- When the Metadata for a TOA5 file is displayed, the Column Mappings table is checked against the column names in the TOA5 file. If any match is found, the Name from any matching entry in the Columns Mapping table is shown in the Metadata display in the Name column of the Columns information for that file.

- When searching for Data Files by Columns, the options in the search parameters are the column headings from all TOA5 Data Files stored in DC21. 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.

11.3.1   The Column Mappings tab

This tab allows Users with Administrator permission to add and delete column mappings.

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

To **add** more mappings click on Add Mapping 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 on Submit Column Mappings .
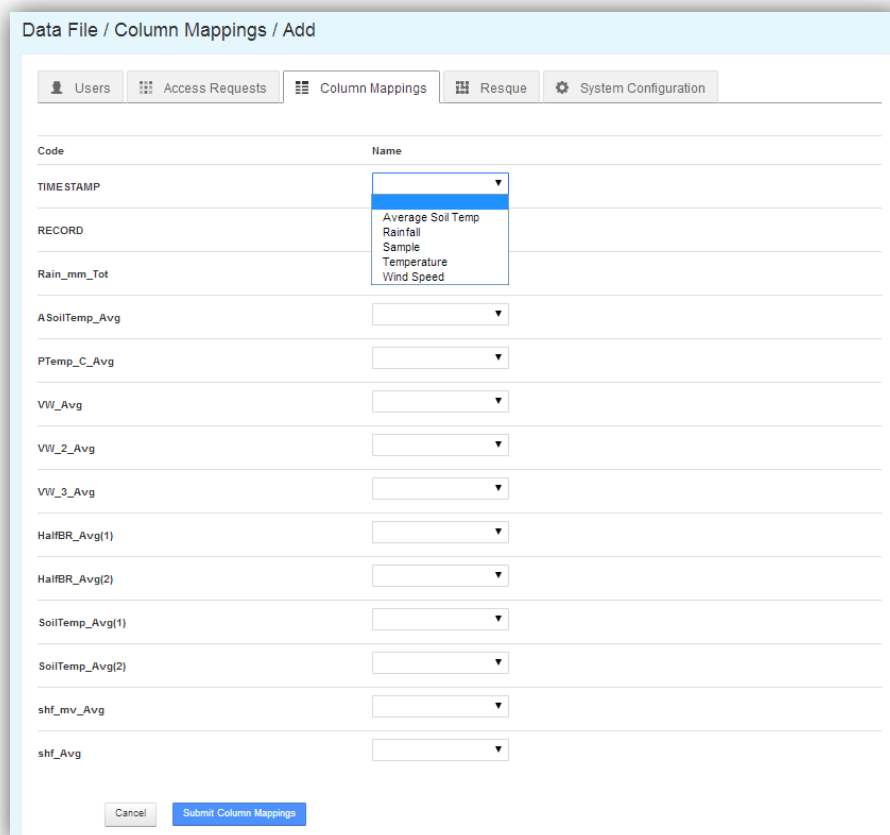
11.3.2   Updating from the Explore Data tab

Column mappings may also be defined using Fill In Column Mappings button on the **View Metadata** page for a TOA5 file.

This is the preferred method for updating Column Mappings, as it avoids the need to manually type the column headers into the Code fields. However, it permits setting mappings relevant to the columns in this TOA5 Data File only.

All Users can perform this function.

The list of unmapped column headings in this TOA5 Data File is displayed in the 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.

11.4 Managing Background Tasks - Resque

DC21 incorporates a Resque server. Resque is the Ruby background task administration system and normally runs on the same server as DC21.

If you are an administration user, you can access the Resque administration screen by selecting **Admin** at the top of the DC21 main screen and then selecting the **Resque** tab.
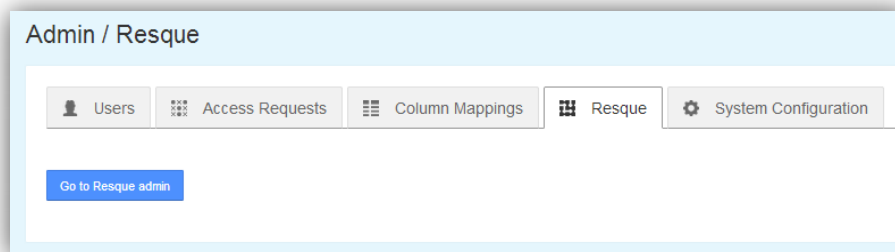
Click on Go to Resque admin to display the following Resque screen in a new browser tab.
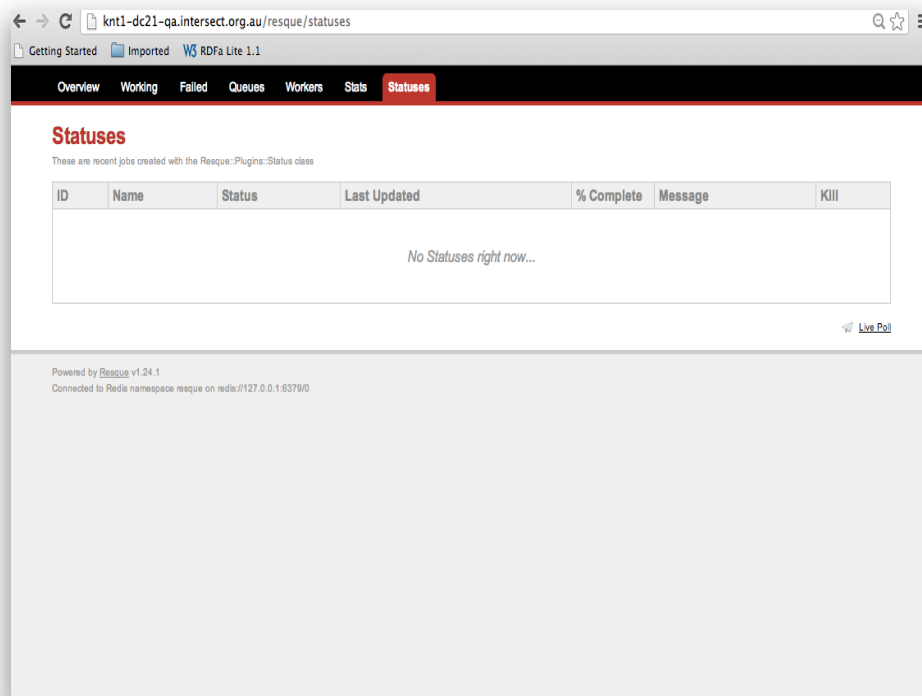
Operation of this screen is beyond the scope of this document. However, you can use this screen to view the status of currently running background tasks, and to delete tasks (via the **Kill** function) if they aren't working properly. Killing a background task will not delete the partially created DC21 Package file. The Package file should be manually deleted after the background task has been killed. See section *8.5 Deleting a Data File*.

Note	If the background task detects an error during processing it will attempt to set the Creation Status to FAILED and write the error message into the output Data File's Description field. In this case, the Resque job entry for this task should also provide an indication as to why the failure occurred.

11.5 Tailoring DC21 for Your Organisation's Needs

DC21 is designed to be installed at different sites and to be used for different types of research data files. Therefore, DC21 offers facilities to change a number of System Configuration parameters to tailor it for each site.

Note	To minimise confusion, configuration options should be set up prior to any Users accessing the newly installed system.

After installation of DC21, refer to the Post Upgrade Instructions in the **Deployment Guide – First Time Server Build** for detailed instructions on tailoring DC21 for operation on your site. The System Administrators installing DC21 must negotiate with the Users of the DC21 system to determine many of the parameters that are set in these steps.

- Register your server with (Australian Access Federation) AAF to permit AAF authentication to be used when logging in to DC21.

- Set a commercial SSL certificate to prevent browser warning messages from appearing each time a User accesses DC21.

- Configure RIF-CS harvesting.

- Install and configure the RESTful API Uploader. Follow the links to this tool for your version of DC21 in the ReadMe and Release Notes at https://github.com/IntersectAustralia/dc21-doc/blob/master/README.md.

- If required, install and configure the DC21-eyetracker-packager. eyeTracker is a stand-alone commercial product published by Tobii. (See http://www.tobii.com/en/eye-tracking-research/global/products/.) DC21-eyetracker-packager integrates eyeTracker into DC21. Follow the links to this tool for your version of DC21 in the ReadMe and Release Notes at https://github.com/IntersectAustralia/dc21-doc/blob/master/README.md

- If required, create an ABBYY user account to be used for Optical Character Recognition (OCR) services. See section *11.5.2 OCR Processing parameters* below.

- If required, create a KOEMEI user account to be used for Speech Recognition (SR) services. See section *11.5.3 Speech Recognition Processing parameters* below.

- Set the local Logo image

After the above are set up, the System Administrator must view and modify the DC21 System Configuration parameters. You must be a DC21 User with Administrator permissions to do this. There is an initial DC21 Administrator user account set up as part of DC21 installation. Its email address and password are selected by the installer.

To view the System Configuration parameters update screen, click on **Admin** on the top of the DC21 main screen. Then click on the System Configuration tab to see the following view.

To modify these parameters, click on Edit System Configuration. All parameters can be modified. Those parameters flagged with * in the Edit System Configuration screen are mandatory.

There are three categories of parameters shown in this view, which are explained in the following three sections.

11.5.1   System Configuration parameters

These parameters allow you to set your organisation's name and contact details, and set the jargon you wish to use.

Most of this information is provided for populating the various fields of the Readme.html file during research data Packaging and Publishing.

Local System Name:	This name appears at the top of each screen, replacing the string "DC21" in all screen shots shown in this document.
	Enter the short name by which your system is commonly known.
Research Centre Name:	Enter the name of the organisation which will own the data uploaded to your system.
Overarching Entity:	Enter the name of the organisation to which your Research Centre belongs. This will typically be the name of an academic institution.
Address:	Enter the address of the organisation to which your Research Centre belongs. A street address is most appropriate.
Telephone Numbers:	Enter a telephone number that can be used to contact your Research Centre. An administration telephone number may be most appropriate.
Email:	Enter an email address that can be used to contact your Research Centre. An administration email address may be most appropriate.
Description:	Enter a brief description of your Research Centre.
URLs:	Enter the public URL for your Research Centre.
Type of Org Unit (Singular): Type of Org Unit (Plural):	Enter the singular and plural terms by which you refer to your Organisational Units. These may be "Research Group" and "Research Groups" or perhaps "Facility" and "Facilities". Refer to Chapter 5 *Organisational Units and Projects* for more information on how these strings are used.
Type of Project (Singular): Type of Project (Plural):	Enter the singular and plural terms by which you refer to your Projects. A Project is generally defined by funding. These may actually be "Project" and "Projects" or perhaps you use a different term, such as "Experiment" and "Experiments". Refer to Chapter 5 *Organisational Units and Projects* for more information on how these terms are used.
Project Parameters:	This option enables support of Project Parameters. The use of Project Parameters are is generally appropriate for systems which will hold environmental research data.
	If this parameter is Enabled, the Add New Parameter button on the Project details screen will be displayed. See section *5.5 Setting Up Project Parameters*.

11.5.2   OCR Processing parameters

DC21 supports processing of image files to extract text and write it to a .TXT file which can then be viewed, edited or processed as required.

DC21 supports the use of two OCR engines: Tesseract (see https://code.google.com/p/tesseract-ocr/) and ABBYY (see http://abbyy.com and http://ocrsdk.com/). Tesseract is a free, open source OCR engine which has been directly integrated into DC21. As an alternative to Tesseract, users can subscribe to the ABBYY cloud service. Intersect recommends the use of the ABBYY OCR engine.

OCR processing is always queued and executed as a background task under control of Resque (see section *11.4 Managing Background Tasks - Resque*). At the time the task is queued, an empty output file is created and its Metadata is updated to reflect the processing status as processing proceeds. Use the Metadata View screen to review this Metadata (see *8.4 Viewing and Editing a File's Metadata*). When processing completes successfully, an email is sent to the initiating User. If it fails, no email is sent, but an error message appears in the output Data File's Metadata Description field.

These parameters control whether this function is supported and sets parameters for its operation.

If the **OCR Supported MIME Types** field is empty, all other fields in this group are ignored.

Auto OCR on Upload:	If set to **Enabled**, OCR processing will automatically be queued for background processing on all uploaded Data Files, provided that Data File has one of the MIME Types listed under **OCR Supported MIME Types**, and its name matches the **Auto OCR Regular Expression** given below.
	Even if this flag is set to **Disabled**, OCR processing can still be queued manually from the Metadata View screen.
Auto OCR Regular Expression:	Automatic OCR on uploaded Data Files will not be launched if the Data File's name does not match this Regular Expression. (The OCR button still appears on the Metadata View screen even if the file name does not match this regular expression.)
	If this field is empty, all Data Files with a MIME Type supported for OCR are processed. Take care not to leave just spaces or trailing spaces in this field, as they will be considered part of the Regular Expression and restrict the matched Data File names.
	See section *8.2.2.1 Regular Expressions* for information on regular expressions and how they are matched.
OCR Supported MIME Types:	List here the MIME Types which will be supported for OCR processing. Only those files with one of these MIME Types can be queued for OCR processing, either automatically following upload or manually.
	To turn off all OCR processing, leave this field blank. The OCR button in the Metadata view screen will also be suppressed.
	MIME Types must be selected from the dropdown list which appears when you click in this control. Type at least three characters to reduce the list to

those MIME Types which include those characters.

Including MIME Types which do not represent images, or which are not supported by your selected OCR engine, will result in processing errors.

Tesseract as integrated into DC21 supports image/jpeg, image/png and image/tiff files. Refer to the ABBYY web site for more information on its supported MIME Types.

ABBYY Host:	Enter the URL of the ABBYY host. The standard ABBYY subscription service is at cloud.ocrsdk.com. However, ABBYY allows users to buy a licence to create their own OCR server. If you are using your own server, that service's URL must go here.
	To use the Tesseract OCR engine which is integrated into DC21, leave this field empty.
ABBYY App Name:	See your System Administrator to find out your organisation's ABBYY Application Name. This name authorises access to your ABBYY account.
	This field is ignored if **ABBYY Host** is empty.
ABBYY Password:	See you System Administrator to find out your organisation's ABBYY password to match the ABBYY account name given in the question above.
	This field is ignored if **ABBYY Host** is empty.

11.5.3   Speech Recognition Processing parameters

DC21 supports processing of audio and video files to extract spoken words as text and write them to a .TXT file which can then be viewed, edited or processed as required.

DC21 supports use of the Koemei Speech Recognition (SR) engine, which is a subscription cloud service. Refer to http://www.koemei.com/ for more information about this service. If you do not have an account for the Koemei service, you will not be able to perform Speech Recognition. DC21 does not provide any fall-back integrated SR engine.

SR processing is always queued and executed in background processing under control of Resque (see _11.4 Managing Background Tasks - Resque_). At the time the task is queued, an empty output file is created and its Metadata is updated to reflect the processing status as processing proceeds. Use the Metadata View screen to review this Metadata (see _8.4 Viewing and Editing a File's Metadata_). When processing completes successfully, an email is sent to the initiating user. If it fails, no email is sent, but an error message appears in the output Data File's Metadata Description field. Use the Metadata View screen to see this Description (see _8.4 Viewing and Editing a File's Metadata_). The output file will always exist but may be empty, even if processing failed.

These parameters control whether this function is supported and sets parameters for its operation.

If the **SR Supported MIME Types** field is empty, all other fields in this group are ignored.

Auto SR on Upload:	If set to **Enabled**, SR processing will automatically be queued for background processing on any Data File which has one of the MIME Types listed under **SR Supported MIME Types**, and its name matches the

Auto SR Regular Expression.

Even if this is set to **Disabled**, OCR processing can still be queued manually from the Metadata View screen.

Auto SR Regular Expression:	Automatic SR on uploaded Data Files will not be launched if the Data File's name does not match this Regular Expression. (The SR button still appears on the Metadata View screen even if the file name does not match this regular expression.)
	If this field is empty, all Data Files with a MIME Type supported for SR are processed. Take care not to leave just spaces or trailing spaces in this field, as they will be considered part of the Regular Expression and restrict the matched Data File names.
	See section *8.2.2.1 Regular Expressions* for information on regular expressions and how they are matched.
SR Supported MIME Types:	List here the MIME Types which will be supported for SR processing. Only those files with one of these MIME Types can be queued for SR processing, either automatically following upload or manually.
	To turn off all SR processing, leave this field blank. The SR button in the Metadata view screen will also be suppressed.
	MIME Types must be selected from the dropdown list which appears when you click in this control. Type at least three characters to reduce the list to those MIME Types which include those characters.
	Including MIME Types which do not represent audio and video files, or which are not supported by the Koemei engine, will result in processing errors.
	Typically, selecting audio/mpeg and audio/x-wav may be appropriate.
Koemei Host:	Enter here the URL of the Koemei host. The standard Koemei subscription service is at www.koemei.com.
Koemei Login:	See your System Administrator to find out your organisation's Koemei account name. This name authorises access to your Koemei account.
Koemei Password:	See your System Administrator to find out your organisation's Koemei password to match the Koemei account name given in the question above.

12   Configuring Tags, Column Mappings and Experiment Parameters

When DC21 is first installed, the available Tags, Column Mappings and Experiment Parameters are populated in the database from the configuration file dc21app_config.yml. 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 on which DC21 is running on and 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 (e.g. <root>/etc/httpd/conf.d/rails_dc21app.conf). 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 your system, navigate to the location of the application installation (e.g. <root>/home/devel/dc21app/current) and enter the directory "dc21app/current". From here you can start the Rails Console using this 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 this 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 this 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">
```

13 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 these commands.

```
$ 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>
```

14   Revision History

Version No.	Revision Date	Summary of Changes	Revised by
V1.8	4 Apr 2013	Initial DC21 version. (Copied and substantially updated from User Manual Version 1.1 for DC21 V1.6b.)	Peter Roberts
V1.9	19 Jun 2013	Updates for version release 1.9.	Peter Bugeia
V2.0	30 Jan 2014	Updated for version DC21 V2.0. New functionality introduced for MU DIVER system. Document is now generic and not based on the HIEv system.	Peter Roberts

Appendix A - The Bagit format

BagIt is currently defined in an Internet Engineering Task Force (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] 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

README.HTML file

Within the data subdirectory of the Bagit ZIP file there is a README.HTML file, which is intended for both human and machine reading. Viewing this file in a web browser will summarise the Package contents, including its Metadata, list of its Data Files, and Metadata for the included Data Files.

The machine readable parts of this file conform to the RDFa Lite (Resource Description Framework in attributes Lite) specification described at http://www.w3.org/TR/rdfa-lite/. This semantic information can be parsed by a program which can then build discoverable facets of information for a search engine.

Appendix B - 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://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.

Example RIF-CS file

```xml
<?xml version="1.0" encoding="UTF-8"?>
<registryObjects xmlns="http://ands.org.au/standards/rif-cs/registryObjects"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://ands.org.au/standards/rif-cs/registryObjects
    http://services.ands.org.au/documentation/rifcs/1.3/schema/registryObjects.xsd">
  <registryObject group="University of Western Sydney">
    <key>/data_files/2734</key>
    <originatingSource>https://jp-dc21-staging.intersect.org.au/</originatingSource>
    <collection type="dataset">
      <name type="primary">
        <namePart>PB-ROS.zip</namePart>
      </name>
      <location>
        <address>
          <electronic type="url">
            <value>https://jp-dc21-staging.intersect.org.au/data_files/2734/download</value>
          </electronic>
        </address>
      </location>
      <subject type="local" xml:lang="en">Meteorological data</subject>
      <subject type="anzsrc-for">05</subject>
      <subject type="anzsrc-for">0502</subject>
      <description type="brief">Rain Out Shelter data for a couple of years. Rest of the
      description.</description>
      <description type="rights">http://creativecommons.org/licenses/by-nc-
      nd/3.0/au</description>
      <coverage>
        <temporal>
          <date type="dateFrom" date_format="W3CDTF">2011-06-19T21:05:00+10:00</date>
          <date type="dateTo" date_format="W3CDTF">2013-03-28T11:00:00+11:00</date>
        </temporal>
      </coverage>
      <coverage>
        <spatial type="gmlKmlPolyCoords">150.73946,-33.61006</spatial>
      </coverage>
      <relatedInfo>
        <notes>Published by Peter Bugeia (peter.bugeia@intersect.org.au)</notes>
```

```
        </relatedInfo>
        <relatedInfo>
          <notes>Unique ID: HIE-ID-0001</notes>
        </relatedInfo>
        <relatedInfo>
          <notes>Primary contact for ROS Weather Station is Craig Barton
        (c.barton@uws.edu.au)</notes>
        </relatedInfo>
      </collection>
    </registryObject>
</registryObjects>
```