

Data Catalogue – How to create a record for your data

This guide outlines step-by-step how you can create a record in the Data Catalogue to help others find your research data, and how to upload digital data files so that they can be archived on secure university storage.

Contents

What is the Data Catalogue for?	1
What is the Data Catalogue not for?	2
Logging in to the Data Catalogue to create a record	2
The three stages of creating a record and the three areas within the repository	2
Uploading Files	3
Uploading files with a total size below 4Gb.....	3
Uploading files with a total size above 4Gb	4
Licensing and availability	4
What is a README file?	5
Completing the details.....	6
Contributors.....	6
Project Details	7
Statement on legal, ethical and access issues	7
DOI (Digital Object Identifier)	7
Finishing a record.....	7
Further help	8

What is the Data Catalogue for?

The Data Catalogue serves two main purposes:

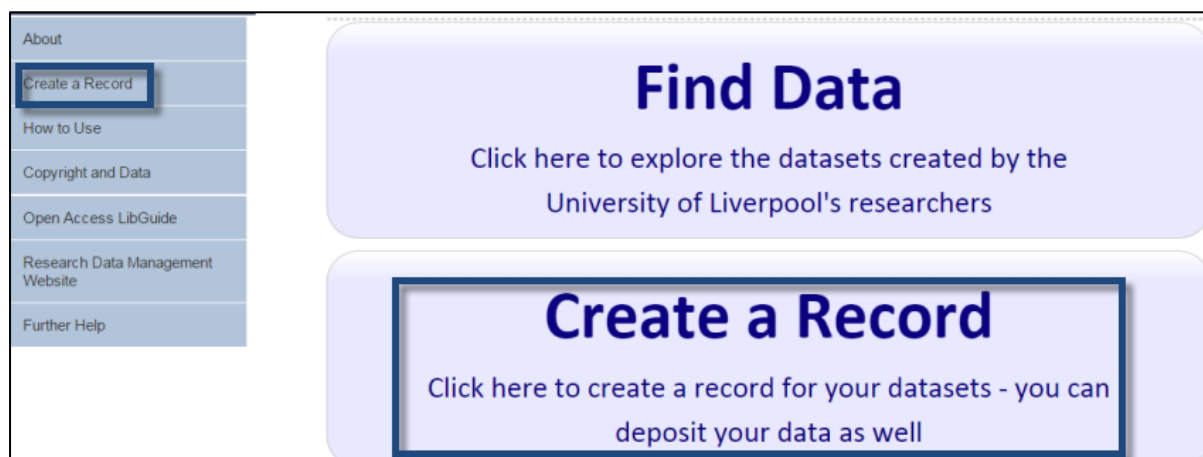
1. To provide a system in which you can create publicly-available records with discovery information about your finalised research data.
2. To provide a system in which you can upload your finalised research data, if you do not already have access to a discipline- or funder-specific database. Your data can be freely available to all, embargoed for a time period, or available only to specific categories of user as determined by you.

What is the Data Catalogue not for?

The Data Catalogue is not for storing data during the active part of your project, nor is it for the storing of data that is likely to change or be added to.

Logging in to the Data Catalogue to create a record

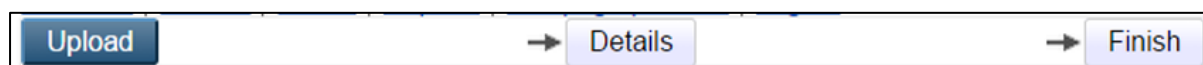
To do this, click “Create a Record” – you can do this either from the left hand menu available on all Data Catalogue pages, or from the large button on the Data Catalogue home page.



This will log you into the “Manage Deposits” screen of the Data Catalogue. Here you will see all the records you have created, and the areas of the Data Catalogue in which they currently reside (see the next section for more details). To create a new record, click “New Item”.

The three stages of creating a record and the three areas within the catalogue

There are three stages to creating a record in the Data Catalogue. These are outlined in the buttons across the top of the screen (see screenshot).



They are:

1. **Upload:** you will use this screen to upload your data files (see later notes *Uploading Data*). If you are creating a details-only record, which points to data you have stored in a discipline- or funder-specific database elsewhere, you can skip this step completely.
2. **Details:** you will use this screen to add details about your data. This will be the information that search engines will use to index your records and make them discoverable, and they will provide the information necessary for others to cite the record (see later notes *Completing the Details*).
3. **Finish:** once you have uploaded your data (if appropriate) and completed the Details, you will submit your record for review. A member of Data Catalogue staff will check the record to ensure it has re-use information included and that the Details have been completed correctly before making the record publicly available.

There are three areas within the Data Catalogue in which your records can be.

<u>Item Status</u>
Under Review
User Workarea
User Workarea
Live Archive
Live Archive

These are:

1. **User Workarea:** while records are in your work area you can continue to edit them as much as you wish. You can click “save and return” at any point to save the information you have entered so far into your work area. Items in your work area are only ever visible to yourself and Data Catalogue administrators.
2. **Under review:** once you have submitted a record for review, it will remain in the review area until it has been checked by a member of Data Catalogue staff. Records under review can no longer be edited.
3. **Live archive:** records in the live archive are publicly available and cannot be edited.

Uploading Files

In this stage of creating a record you can upload finalised data, ideally along with a README file (see later notes *What is a README file?*).

The way in which you upload your files will depend on your answer to the following question: Are you uploading files with a total size that is above or below 4Gb?

Uploading files with a total size below 4Gb

This is a straightforward process, with two options available to you – to upload files directly, or to upload them via a URL if your data are available at a specific online location (note this method does not upload any discovery details about your data, only the data files themselves).



To upload files directly

1. Select “From file(s): <4Gb”.
2. Click “Choose file”.
3. Navigate to the file(s) you wish to upload and select it.
4. The file has now been uploaded.
5. Click “Show Options” to enter important licencing and availability information (see later notes on *Licensing and availability*).

You can now move onto the Details screen by clicking “Next”.

To upload files from a URL

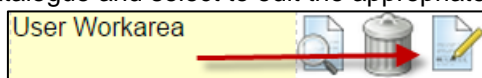
1. Select “Other methods: no file size limit”.
2. Enter the URL into the “Capture from URL” box.
3. Click “Upload”.

You can now move onto the Details screen by clicking “Next”.

Uploading files with a total size above 4Gb

This is a more involved process, and may require you to start the process of creating the Data Catalogue record, wait for a data staging area to be completed for you, then finish the process of creating the Data Catalogue record at a later date.

1. Select “Other method(s): no file size limit”.
2. You now have a choice: if you have a URL from which your data can be downloaded, follow the instructions in the section above on *To upload files from a URL*; if you do not have a URL from which your data can be downloaded, follow the instructions below.
3. Under the “Staging” heading, click the link as indicated to send an email requesting that a data staging area be created for you. There is no need to add any text to the message – simply make a note of the number in the subject line of the email. You can now either click “Save and return” and return to finish the record once the data staging area is ready (takes up to 48 hours), or you can click “Next” and fill in the Details screen before saving the record.
4. It takes around 48 hours for a data staging area to be set up. You will be emailed with instructions once this is ready.
5. Once you have been sent an email saying the data staging area is ready, create a folder on the University network (e.g. within your collaborative project filestore) with a name corresponding to the number for this record, which was also the number in the email subject line; e.g., if the number in the email subject line was 1234, create a folder on the University network called 1234.
6. Copy your data files (including a README file – see *What is a README file* below) into this folder.
7. Log back into the Data Catalogue and select to edit the appropriate record



8. Now in the Upload screen select “Other methods: no file size limits”, and you will see the list of your data files.
9. Click the button “Click here to add all the above files to this item” to upload the data into the Data Catalogue.
10. Click “Show Options” to enter important licencing and availability information (see *Licensing and availability* below).

You can now move onto the Details screen by clicking “Next”.

Licensing and availability

The screenshot shows a form with the following fields:

- Visible to:** A dropdown menu set to "Anyone".
- Content type:** A dropdown menu set to "UNSPECIFIED".
- Description:** An empty text input field.
- Embargo date:** A date selection interface with "Year:" (empty), "Month:" (dropdown set to "Unspecified"), and "Day:" (dropdown set to "?").
- License:** A dropdown menu set to "Creative Commons: Attribution 4.0".

Each field has a question mark icon to its right, and a "Hide options" button is visible in the top right corner.

In this screen you can set the following important information:

- **Visible to:** for data that can be made openly available to all, leave this at the default “Anyone”; if users of data will only be able to access data under certain conditions, change this to “Data Catalogue staff only”, and be sure you complete the information on the Details screen under the heading “Statement on legal, ethics and access issues”.
- **Embargo date:** if there is an embargo period during which no-one should be able to access the data themselves, set the end date of the embargo here – after this date, the choice you made in the “Visible to” menu will be enforced. Bear in mind the details about the data will still be visible to all.
- **Licence:** the default licence is a Creative Commons – Attribution licence. The Data Catalogue is a tool primarily for the sharing of research data. As such, the default recommended licence under which data deposited in the Data Catalogue are held is a [Creative Commons – Attribution licence](#). This clearly lets potential users know that they are free to re-use your data as part of their own research, provided they give you proper attribution. This licence supports the wide sharing of data for the greater good of the research community.
 - You will be able to select a more restrictive licence if you wish, and the Data Catalogue provides links to further information on those licences. High level guidance on copyright and licensing issues, along with links to more in-depth information, are available on the Data Management website’s [Copyright page](#).
 - It may be that you are required by your funder to use a specific type of licence that is not among those listed on the Data Catalogue. In these cases please send an email to mooread@liv.ac.uk with a link to a description of the licence and we will add it to the list.

What is a README file?

A README file is a crucial piece of information necessary for others to understand and make use of data. It is important that this is as clear as possible, as it may be possible that other researchers will have uses in mind for your data that you had not considered yourself. It is useful to ask yourself the question “if I was someone else who didn’t know how these data were created and what the variables meant, what would I need in order to understand them?”.

Examples of things to include in a README file include:

- Data creation or observation conditions – what software or survey instruments were used to collect the data? What were the environmental conditions of observation data?
- Code label and field descriptors – to what do data variables refer? Can someone with knowledge in the same field actually make use of the data?

Some re-use information, such as field descriptors, can be included within the data files themselves. Other re-use or contextual documentation may take the form of supporting documents, such as lab books or interview schedules – include these with your files wherever appropriate.

Completing the details

The Details screen is where you will enter discovery and citation information about your data – what is the name of the data set, what keywords describe its content, who created it, who funded the underlying research, and so on.

Most of these details should be clear from their names, and certain key fields on this screen have had their help notes opened by default (if you need help notes for other fields, click the ? icon next to them).

The screenshot shows a web interface for data upload. At the top, there are three tabs: 'Upload', 'Details' (which is active), and 'Finish'. Below the tabs are four buttons: '< Previous', 'Save and Return', 'Cancel', and 'Next >'. The main content area is divided into three sections, each with a star icon and a title:

- Title of dataset**: A text input field with a help icon (?) on the right.
- Collection description**: A text input field with a help icon (?) on the right. Below the title, there is a paragraph of help text: "Provide a description of the data that will help others find it when searching the web. The description should be clear and concise. Include the kinds of phrases you yourself would use to find this type of dataset. If the data collection has an abstract then this could be entered here."
- Keywords**: A text input field with a help icon (?) on the right. Below the title, there is a paragraph of help text: "You must enter at least one keyword to describe your data. Choose keywords you would use yourself to find this data. Separate multiple keywords with commas"

Some of the fields require a bit more explanation, as given below.

Contributors

First, to answer the question you may be asking – no, you can't import an author list from elsewhere. Contributor names are entered manually, but entering details of Liverpool-based contributors is made easier by an auto-complete function:

- If you type in your staff ID, you will see your details – just click on your name (not the email link) and your details will be added.

- If you type in your Liverpool co-authors' names, a list of suggestions will display – again, click on the appropriate name to add it.

After you have done this, you may if you wish edit the way your name displays (in this way the Data Catalogue currently differs from the University Repository for publications).

A word on staff ID

It is very important to have your staff ID in the record. If you do not, it will be extremely difficult to integrate the Data Catalogue record with other University systems.

Project Details

If your data have arisen from an externally funded project with details in IRIS, you can enter the IRIS Project ID for the appropriate project, and the rest of these fields will be automatically completed.

Statement on legal, ethical and access issues

It is vitally important that if there are any legal or ethical considerations with regards to uploaded data, these are entered clearly here.

It is especially important that you clearly outline under what circumstances potential users can access data if you have chosen not to make them available to anyone, and instead only want them available to those who meet specific criteria. Data will or will not be made available according to the instructions here, so it is vital your instructions are clear.

Unclear instructions here may see Data Catalogue staff return the record to your work area with a request to make the instructions clearer.

Be aware that this statement will be publicly viewable to anyone.

DOI (Digital Object Identifier)

A Digital Object Identifier is a unique identifier for a piece of information on the web. If your data are held in an external discipline- or funder-specific database (e.g. the UK Data Archive), they will likely already have been given a DOI. If that is the case, insert the DOI here. Don't include anything before the first 10, i.e. do not enter information starting <http://dx.doi.org>.

If you are uploading data into the Data Catalogue itself, leave this field blank. Once you have finished creating the record a DOI will automatically be generated for your data. You will be able to use this to promote your data, and others can use it both to link to and cite your data.

Finishing a record

Once you have uploaded your data (if appropriate) and completed the Details screen, click "Finish". Read the information on the screen, and then once you are happy click "Submit for review". Only do this once you are sure you don't want to edit the record further, as once it is under review no further editing is possible.

Deposit item: Third test dataset

University home > DataCat > Deposit item: Third test dataset

Logged in as Martin Wolf | [Manage deposits](#) | [New Deposit](#) | [Manage records](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Reports](#) | [Edit page phrases](#) | [Logout](#)

Upload → Details → Finish

For work being deposited by its own author: In self-archiving this collection of files and associated bibliographic metadata, I grant DataCat: Test the right to store them and to make them permanently available publicly for free on-line. I declare that this material is my own intellectual property and I understand that DataCat: Test does not assume any responsibility if there is any breach of copyright in distributing these files or metadata. (All authors are urged to prominently assert their copyright on the title page of their work.)

For work being deposited by someone other than its author: I hereby declare that the collection of files and associated bibliographic metadata that I am archiving at DataCat: Test) is in the public domain. If this is not the case, I accept full responsibility for any breach of copyright that distributing these files or metadata may entail.

Clicking on the deposit button indicates your agreement to these terms.

After the record has been reviewed, one of two things will happen. Either:

- a) Data Catalogue staff will identify a problem with the record and contact you for clarification; or (more frequently)
- b) Data Catalogue staff will approve the record and move it into the live archive. At that point the details will become publicly available, and the access restrictions (if any) you have applied to the data themselves will be enforced.

Further help

If you require further technical assistance, please send an email to data@liverpool.ac.uk, including "Data Catalogue" in the subject line.

You can access further information about Research Data Management at <http://www.liv.ac.uk/csd/research-data-management/>.