Pure User Guide: Creating a dataset

In line with funders' guidelines and Lancaster University's [Data Policy](#) research data directly associated with publications (as a minimum) should be archived where possible. All research data to be archived must be associated with metadata. You can use Pure for the following options:

a. **You have published your data already in an external data repository:** You need to publish metadata about your dataset in Pure and add a persistent identifier (e.g. a DOI) to the Pure record.

b. **You want to publish your dataset in Pure** (with access restrictions/embargo if necessary): You create a metadata record in Pure and upload your data files. After the validation of your record by the Library you will receive a DOI which you can use to acknowledge your dataset in your publications (usually within two days).

c. **Your dataset is confidential and cannot be shared but you want to publish metadata about that dataset in Pure** and explain why the data cannot be made publicly available. After the validation of your record you will receive a DOI for that metadata record which you can use to acknowledge your (restricted) dataset in your publications.

Please note that each funding body will have different rules on when metadata and datasets have to be made available. The [Research Data Management Support Team](#) can advise staff further on their obligations, or see the links at the bottom of this guide.

Note on access options for datasets:
At the moment Pure does not support Controlled Access (data files are hidden, prospective downloader must email to be given access). This functionality is currently being developed and will be enabled in Pure later in 2015. See the [Appendix](#) in this document for an explanation of access and embargo options.

This user guide will help you to add a dataset into Pure.

1. **Where to add your Datasets in Pure**

On your Pure personal overview page, the "Datasets" option is at the bottom of the left hand menu. Clicking the "+" button will allow you to add a new dataset. Alternatively you can use the green "add new" button on the right hand side of the screen and then select dataset from the left hand list.
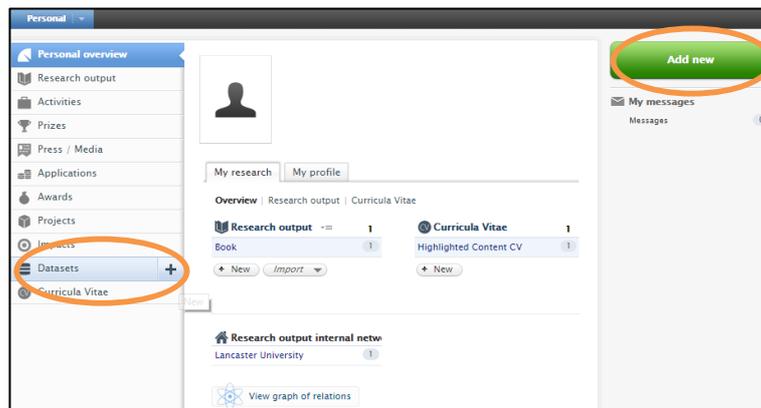
2. A new template will open up for you to add details of the dataset. Information about the field can be gained by hovering over the "i" icon. Any fields with a red * must be completed; these are **Title, People, Dataset managed by, Publisher** and **Date made available**.
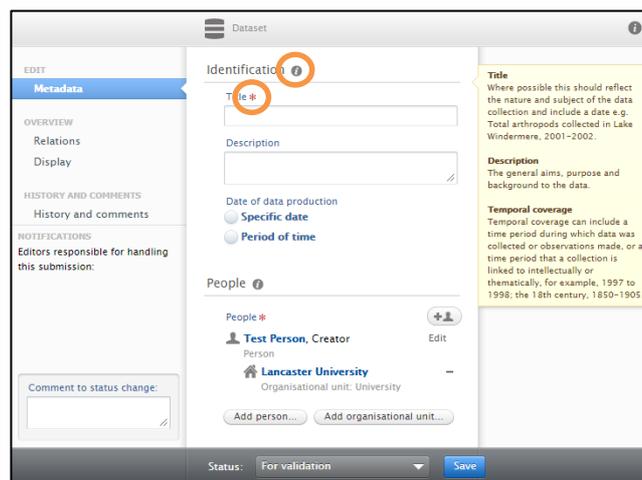

Temporal coverage refers to the period of time the data set covers for example that data covers animal records from 1850-1905, and the date of data production can also be completed. For some disciplines geographical coverages or geospatial points/polygons may be useful and this can also be specified.

3. **Adding persons and roles (mandatory field)**

Next you need to add the main researchers and contributors involved in producing the data. You will automatically be added to the dataset but colleagues can be added by clicking the 'add person' button and typing their name in the search box that appears. External colleagues can also be added using the 'create external person' function.

Once you have located the person, please ensure you edit their role and organisation affiliations in the production of the dataset in question. If you have already added them and want to edit their role or organisation affiliations further, you can do this by clicking on the 'Edit' button next to their name.
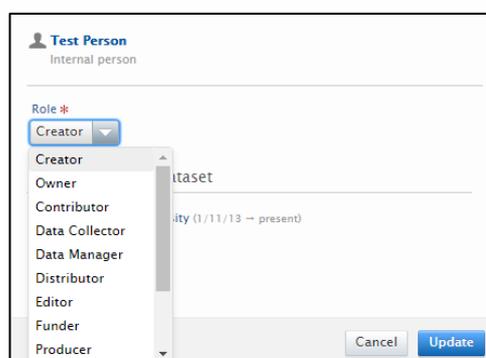
4. Setting the Managing Organisation (mandatory field)

The Managing Organisational Unit needs to be the **Department** which will hold or manage the data, and **not** a research centre, so please change this if needed.
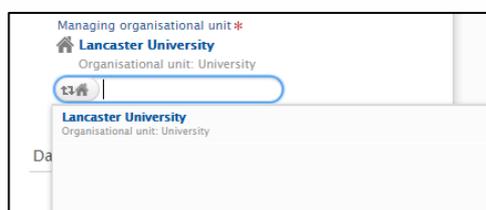
5. Add a Publisher (mandatory field)

If:

    a.   You have published your data already in an external data repository, the publisher will be the external repository owner, e.g. UK Data Archive.

     b.   You want to publish your dataset in Pure, then the publisher will be Lancaster University.

     c.   Your dataset is confidential and cannot be shared but you want to publish metadata about that dataset in Pure, then the publisher will be Lancaster University but you will need to place an embargo/restrictions on the dataset itself.

If you need to change the publisher once added, you can do so using the 'change publisher' function.

6. Adding a DOI

Add an existing DOI if one already exists (e.g. if the dataset is published elsewhere). However if you wish to publish your data in Pure, then a DOI will be created for you during validation and it will be emailed to you within two working days after submitting the record.
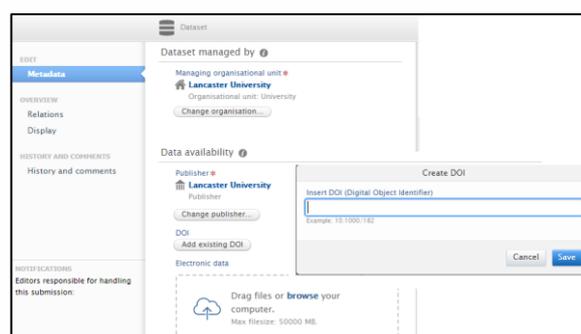

7. Adding the dataset files

You can add the dataset itself if you want to publish using Pure. If the dataset is published externally, there is no need to add it in Pure, but do add the DOI to the external repository. Details of physical data (e.g. sample specimens, works of art etc.) can also be listed. Select 'Upload Electronic Data' to add the

data files and drag the file to the box and add further relevant details for the dataset, such as title and type.

Multiple files can be dragged or uploaded at this point. Large files may take a moment to buffer and this uploading process is shown on the screen. Some browsers may struggle to upload files larger than 4GB on occasion.
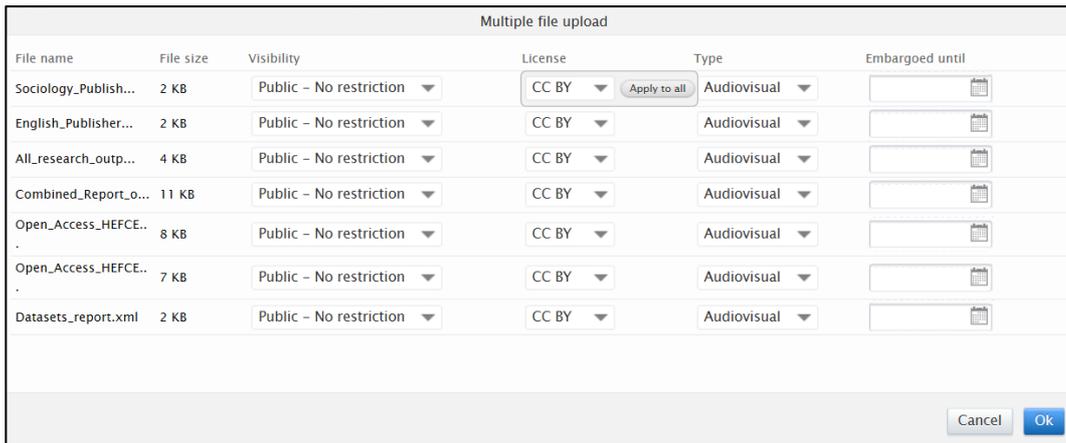
Multiple file upload						
File name	File size	Visibility	License		Type	Embargoed until
Sociology_Publish...	2 KB	Public – No restriction ▼	CC BY ▼	Apply to all	Audiovisual ▼	📅
English_Publisher...	2 KB	Public – No restriction ▼	CC BY ▼		Audiovisual ▼	📅
All_research_outp...	4 KB	Public – No restriction ▼	CC BY ▼		Audiovisual ▼	📅
Combined_Report_o...	11 KB	Public – No restriction ▼	CC BY ▼		Audiovisual ▼	📅
Open_Access_HEFCE..	8 KB	Public – No restriction ▼	CC BY ▼		Audiovisual ▼	📅
Open_Access_HEFCE..	7 KB	Public – No restriction ▼	CC BY ▼		Audiovisual ▼	📅
Datasets_report.xml	2 KB	Public – No restriction ▼	CC BY ▼		Audiovisual ▼	📅
					Cancel	Ok

Please note: If your file size is exceptionally large or your encounter problems uploading your data then please email the Research Data Management Support Team for further help and details on how to add the dataset.

At this point you can set the visibility, license, type and embargo to all the files. If you wish to apply the same setting to each file select the "apply to all" feature next to the box. The license will default to CC BY unless changed.

You can also set the dataset's visibility as required, along with adding an embargo to the dataset. This embargo affects the file itself, and not the metadata, so if you set an embargo for 6 months for example, the metadata will be visible on the Research Portal immediately after validation, but the dataset file will only become visible in 6 months' time.

It is possible to set the visibility, license, type and embargo to be different for each file.
Further guidance and help on the various visibility settings can be found in the appendix at the end of this guide.

Details of the physical data can be added below.

Physical data
Add details of physical data...

8. Date made available (mandatory field)

The Date made available is essentially the Publication date of the data. If an embargo period has been used on the data file, then the date made available should be the day after the embargo period ends (so the day your data files become visible).

9. Contact Information

The person added here must be someone added in the people section. A generic contact email will be displayed and RDM Enquiries will field all emails regarding datasets.

10. Note any Legal/Ethical constraints

You also need to provide details of any legal/ethical constraints upon the dataset. If you cannot share your dataset due to legal/ethical constraints, please provide further details of the reasons in the relevant text box that appears once a legal/ethical constraint box has been selected. Funders like EPSRC require this information when data is not openly shared.

Further guidance and help on the various visibility settings can be found in the [appendix](#) at the end of this guide.

11. **Relating other content to the dataset**

Any other content associated with the dataset, including projects or equipment, can also be linked by clicking the + button in the relevant section and searching for the record.

Please ensure you link your dataset to any research output records that use the dataset that are already added to Pure as well as any Research Projects.
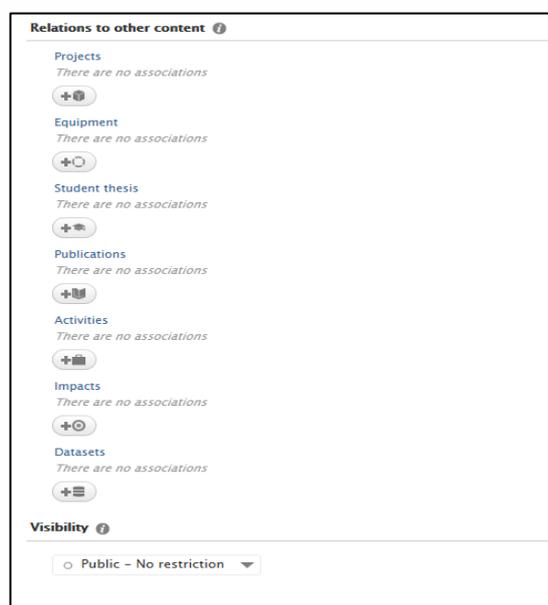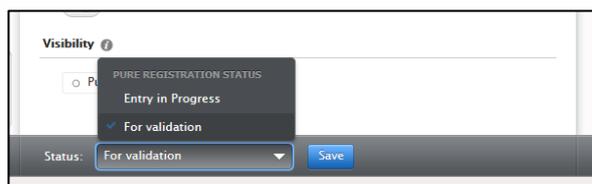

12. **Metadata Visibility**

Lastly, you can chose the visibility required for the dataset, and also save the dataset as an 'entry in progress' if you need to come back to it later. We **strongly recommend** using the default "Public – No restriction" option and restrict access on the file level if necessary.

To come back to a dataset that you saved as an 'Entry in progress' you need to go to your personal overview and select datasets in the left hand menu, then click on 'My datasets' and find it within the list provided.

13. Sending for Validation

Once you have completed a dataset, you can submit it to the library for validation by selecting 'For Validation' and clicking 'Save'. Please make sure you press Save to ensure all the data you have entered is stored in Pure. Please note, even though you may have clicked 'Create' earlier on in the process e.g. when adding the dataset file, you still need to click 'Save' at the bottom of the record for the information to be stored within Pure.

Additional information

Managing your Research Data guidance:
http://www.lancaster.ac.uk/library/rdm/

Funders' expectations:
http://www.lancaster.ac.uk/library/rdm/plan/funder-expectations/

This user guide is in addition to the videos and detailed help sections on Pure Support.

If you find any wrong or incomplete information in this guide do alert us so that we can update it. Please contact rso-systems@lancs.ac.uk.

Appendix

Dataset visibility

Currently there are multiple ways to control the visibility of a dataset:

1. Access to the electronic document level
2. Access on the dataset level (please note this option is not currently in use. Any changes made to the access here will not affect the dataset visibility).
3. Access on a record level

Please see the following instructions for advice on selecting the best option for your dataset:

    i.   Please leave "access to the dataset" section as not set, and use options on the full record and the electronic documents themselves to control the visibility.

    ii.   Placing an embargo on the electronic document file level (See Section 7 of the user guide) will instantly make the metadata available following validation but the file itself will only be available, (both to view and download) once the embargo has passed.

    iii.   If a document is embargoed none of the document information (including its metadata) will be seen on the web. However the metadata for the dataset record will still be seen.

    iv.   Altering the record visibility to prevent it from being visible on the web is not recommended for meeting open access requirements. It is better to restrict the dataset visibility at the electronic document file.

What do I need to if:

I want to publish my data but only after an embargo period of 12 months

1. Do not change default options to "Access to the dataset" (leave as "Not set") and "Visibility" (leave as "Public – no restrictions")
2. Add embargo period when you "Upload electronic data" and pick date under "Embargoed until".
3. This has the effect that your metadata will become visible on the Research Portal after validation but the data files will only be available to the public after the end of the embargo period.
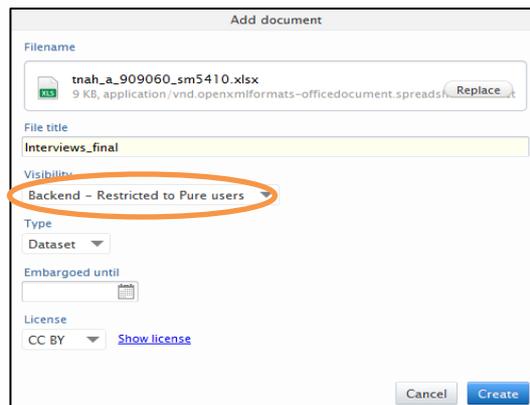
Please note that the Publication date (the mandatory field "Date made available") should be the date after the embargo period ends and your data files become visible.
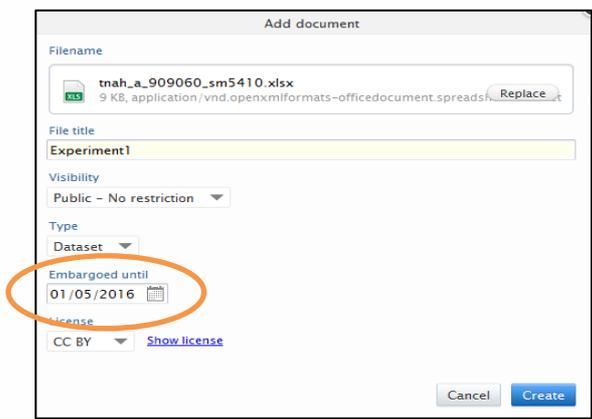
I want to deposit my data but only make the data available on request because it has a level of confidentiality

1. Do not change default options to "Access to the dataset" (leave as "Not set") and "Visibility" (leave as "Public – no restrictions")
2. When you "Upload electronic data" and upload a file choose "Backend – Restricted to Pure Users". Do not choose an Embargo date.
3. In your "Description" explain that data can be accessed on request and express what conditions must be met. **Also** further down on the form tick the appropriate box(es) in "Legal/ethical".

Your metadata will be (after validation) visible on Pure Research Portal but the data will not be available. The contact email address rdm@lancaster.ac.uk can be used by interested individuals to request access to the data. You will be notified when that happens.