****



Adding datasets to Pure

**July 2023**

Contents

[Introduction 2](#_Toc140237923)

[Importing existing datasets 2](#_Toc140237924)

[How to import dataset metadata using Data Monitor 2](#_Toc140237925)

[Depositing new datasets 4](#_Toc140237926)

[When to deposit 4](#_Toc140237927)

[When you are about to submit your manuscript: 4](#_Toc140237928)

[After your manuscript has been accepted and is ready for publication: 5](#_Toc140237929)

[How to deposit 7](#_Toc140237930)

[Type 7](#_Toc140237931)

[Identification 7](#_Toc140237932)

[People 8](#_Toc140237933)

[Managing organisational unit 8](#_Toc140237934)

[Data availability 8](#_Toc140237935)

[Access option 9](#_Toc140237936)

[Relation to content 9](#_Toc140237937)

[Visibility (metadata) 9](#_Toc140237938)

[Status 9](#_Toc140237939)

[Save 9](#_Toc140237940)

[How to deposit (visual instructions) 10](#_Toc140237941)

[University Advice and Guidance 14](#_Toc140237942)

[Appendix 14](#_Toc140237943)

[Example of Description 14](#_Toc140237944)

[Example of ReadMe.txt file 14](#_Toc140237945)

# Introduction

There are two ways to add datasets to Pure. If your dataset has been deposited in an external repository and it has been assigned a DOI, you can import its metadata via the Data Monitor as described in the section “Importing existing datasets”. If you want to deposit your data files in Pure and receive a DOI from the Research Data Management team, you can create a new record as described in section “Depositing new datasets”.

# Importing existing datasets

## How to import dataset metadata using Data Monitor

1. Log in into [Pure](https://www.st-andrews.ac.uk/staff/research/pure/). From your personal profile, open a Datasets template by clicking on the green button labelled “+ **Add content”** which is found on the right-hand side of the screen (Figure 1). Then select “**Datasets/Software”** from the “**Choose submission”** window. A new pop-up window will open.

Screenshot of Pure showing a green clickable button labelled "+ Add content" and an orange arrow pointing at the button.



Figure 1 ‘+ Add content’ green button used to add content to Pure.

1. On the left-hand side menu of the new pop-up window, select “Import from online source” and click on “Data Monitor” (Figure 2).

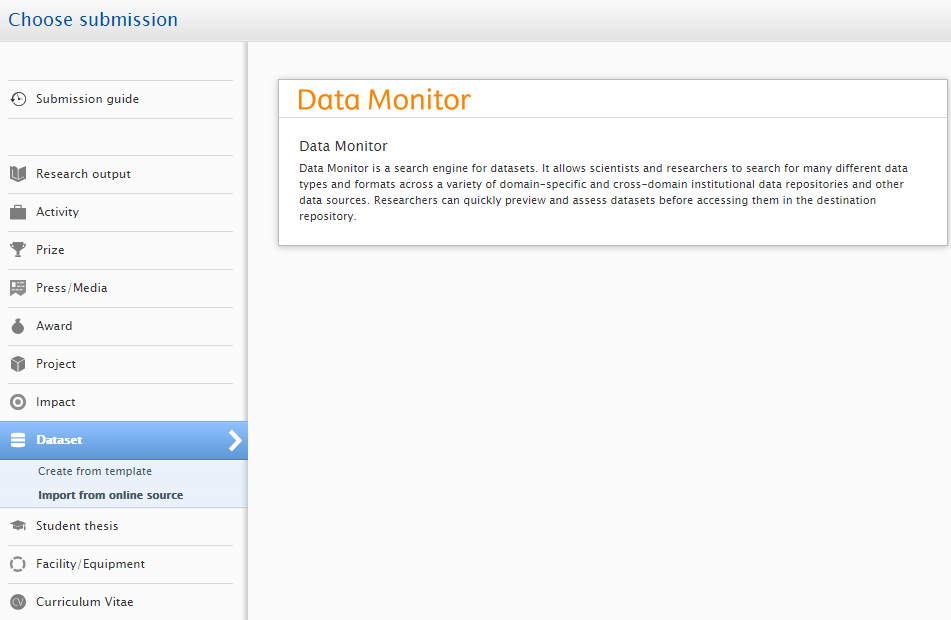


Figure 2 ‘Add content’ pop-up window and Data Monitor

1. On the Data Monitor search window (Figure 3), search for the dataset or software by title or DOI and click on it when found. Important: before importing, check that the dataset does not already exist in Pure. The import feature will not always highlight duplicates.

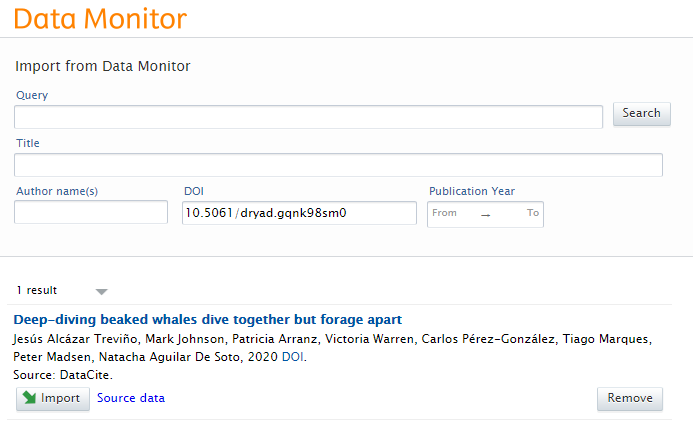


Figure 3 Data Monitor search window.

1. After clicking ‘Import’, a new pop-up window will open for the dataset or software. On this new screen, check that the authors suggested and their affiliations are correct.
2. You will be presented with two options:
   1. ‘Import and Review’, which allows you to have a final check of the dataset or software before saving it. This will allow you to create any link to other content in Pure such as research outputs, other datasets and projects;
   2. ‘Import and save’, which will save the record for the RDM team to check. The team might get in touch if they have any questions.

# Depositing new datasets

## When to deposit

Below is the sequence of steps that we advise you to follow in order to deposit your dataset in Pure and obtain a DOI. Figure 4 illustrates these steps using a flowchart. More details are provided in the following sections.

### When you are about to submit your manuscript:

1. Create a basic dataset record in Pure:
   * Provide at least Title, Authors with affiliations and Date made available (the year will be enough, the RDM team will fill in the rest of the date) (see the ‘How to deposit’ sections for more details).
   * Set the metadata visibility as Backend. Since the record is in preparation it should not be publicly visible yet (see section ‘Visibility (metadata)’ for more details).
   * Set the records’ status as “Entry in progress” (see ‘Status’ section for more details).
   * Let us know if you require your data to be made public straightaway, for example for access by the reviewers.
2. The RDM team will email you an **inactive** DOI
3. Add the DOI to the manuscript using:

“The research data (and/or materials) supporting this publication can be accessed at [DOI]”

You can find more information about data citation and access statements on the [RDM website](https://www.st-andrews.ac.uk/research/digitalresearch/researchdata/publisharchiveandpreserve/datacitationandaccessstatements/).

1. Submit your manuscript

### After your manuscript has been accepted and is ready for publication:

1. Upload your final data files in Pure, but set the file visibility to “**Backend**” (for more details on file visibility, see the ‘Data availability’ sections). Check that title, authors and year made available are correct, add a description and set the status as “For Validation” (see ‘Status’ section for more details). Important: Until we change the files’ visibility, the files will not be visible by the public.
2. Send us an email asking to activate the DOI and to let us know when and on what terms the data should be published. Please note that for this step the metadata (title, authors, description and publisher) will be made publicly available on the [University’s portal](https://risweb.st-andrews.ac.uk/portal/). It is important you notify us if, for some reason, this cannot happen (for example because of a Journal policy).
3. Once the article is published, we will make the datafiles publicly available. If you have set an embargo date on your datasets (see the ‘Data availability’ sections for more details), they will not appear on the portal until the embargo expires.

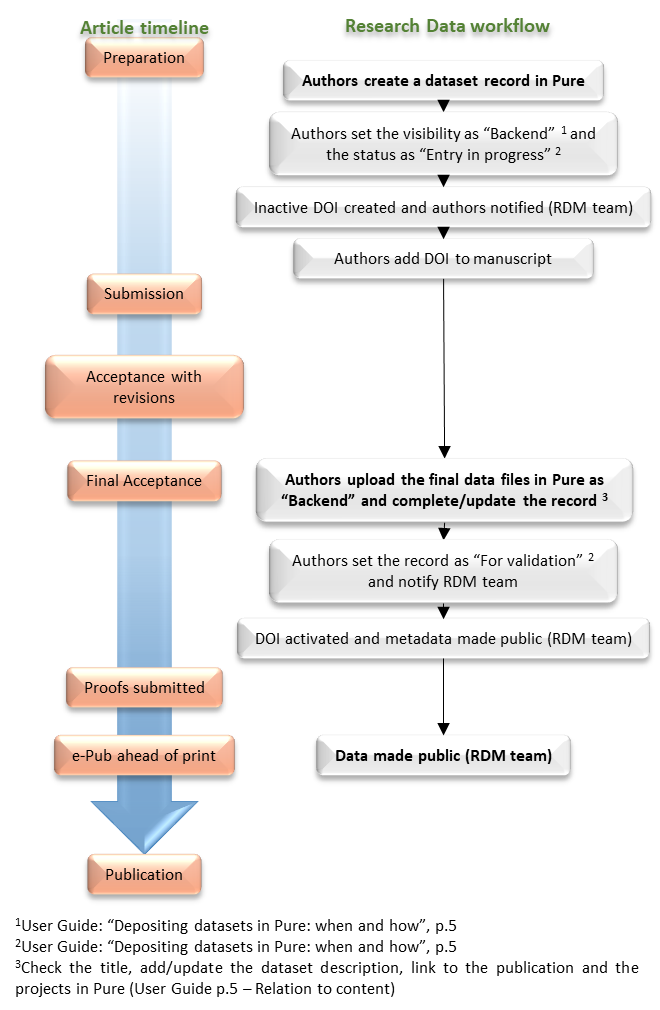


Figure 4 Dataset or software deposit workflow against the article publication process. 1Visibility (metadata) section; 2Status section, 3Check the title, add/update the dataset description, link to the publication and projects in Pure (‘ Relation to content’ section).

## How to deposit

Below is a detailed description of how to deposit, a simplified visual representation of these instructions with key highlights is available in the next section.

Log in into [Pure](https://www.st-andrews.ac.uk/staff/research/pure/). From your personal profile, open a Datasets template by clicking on the green button labelled “+ **Add content”** which is found on the right-hand side of the screen (Figure 5). Then select “**Datasets/Software”** from the “**Choose submission”** window. A new pop-up window will open.

Screenshot of Pure showing a green clickable button labelled "+ Add content" and an orange arrow pointing at the button.



Figure 5 ‘+ Add content’ green button used to add content to Pure.

### Type

From the drop-down menu at the top of the record you can select the type of record you are about to create. Three options are available: dataset, thesis dataset and software.

### Identification

**Title:** We advise to use the form: “*Title of the article (dataset)”.*

**Description:** you should briefly describe the dataset, what type of files are attached and in which formats, if proprietary software is required (provide details) and any other relevant information that can help the user understand the datasets. More detailed information could also be provided as a ReadMe.txt file alongside the data files. See Appendix for more details and examples.

**Date of data production:** this can be either a specific date or a period of time over which the data have been collected/produced. Please note that this field becomes mandatory only if one of the two options is selected (Figures 6 and 7).

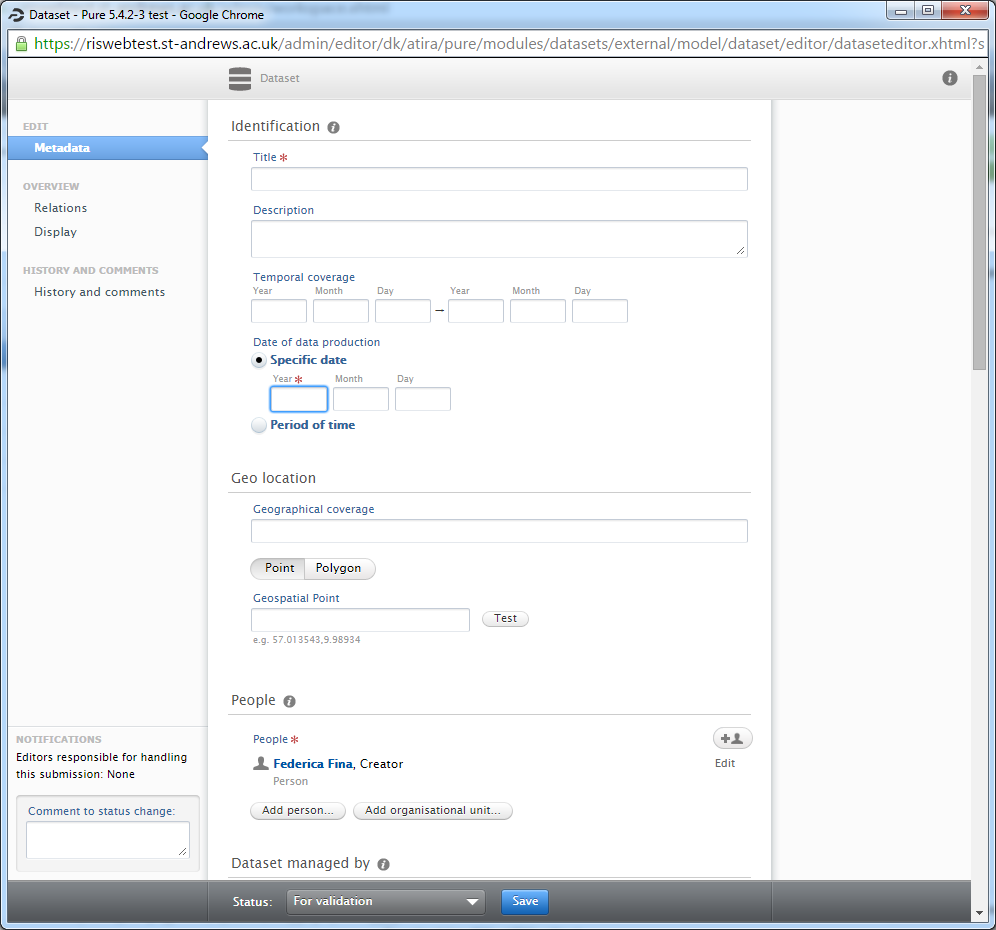


Figure 6 Screenshot of the date of data production/specific date.

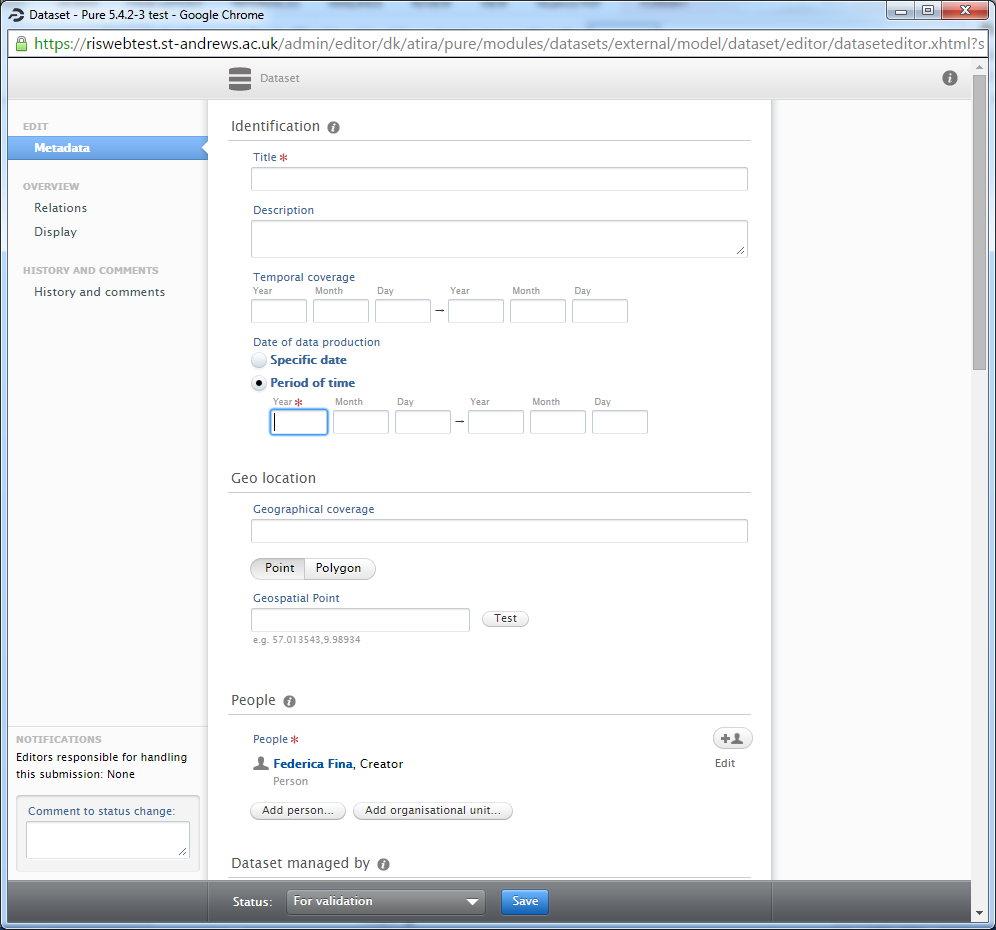


Figure 7 Screenshot of the date of data production/period of time.

### People

This will be pre-populated with your details. You will need to add the remaining authors and their affiliations based on their contribution to the creation of the data files. You can add either internal or external authors by clicking on “Add person…”. By selecting ‘Search’ you will be able to add internal people, while by selecting ‘Create external person’ you will be able to add external contributors (Figure 8). For each person you will be able to specify their role in the creation of the datafiles.

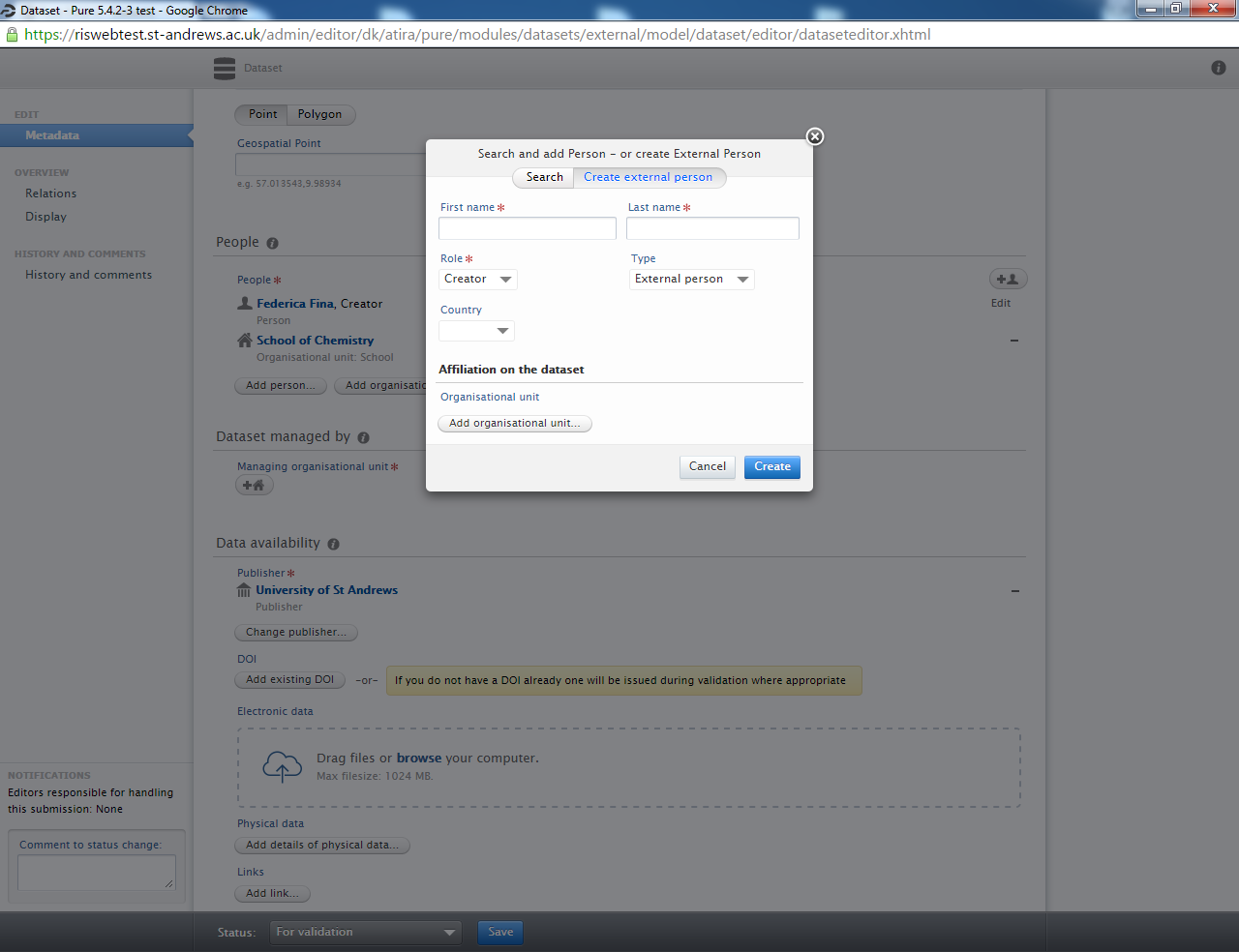


Figure 8 Screenshot of the ‘Add person’ pop-up window.

### Managing organisational unit

This section will be pre-populated with your details. It can be changed to a co-author’s internal organisational unit (School) if deemed more appropriate.

### Data availability

In this section you will be able to add files and specify other metadata such as publisher, files visibility, licence and date made available, including an embargo.

**Publisher.** This is set by default to “University of St Andrews”. You should leave it as default unless you are creating a record for data deposited in another repository (e.g., the data were deposited by a co-author in another university’s repository).

**DOI.** If a DOI does not exist, the RDM team will provide you with one and fill in this field. If the data are published in another repository or database and a DOI already exists, you can “Add existing DOI”.

**Electronic data.** this is where the files can be uploaded. After having selected the files for upload, the ‘Multiple file upload’ pop-up window will open (Figure 9). Note that the maximum file size is currently set at 20GB, however you can upload multiple files. For larger files, please, contact the RDM team at [research-data@st-andrews.ac.uk](mailto:research-data@st-andrews.ac.uk).

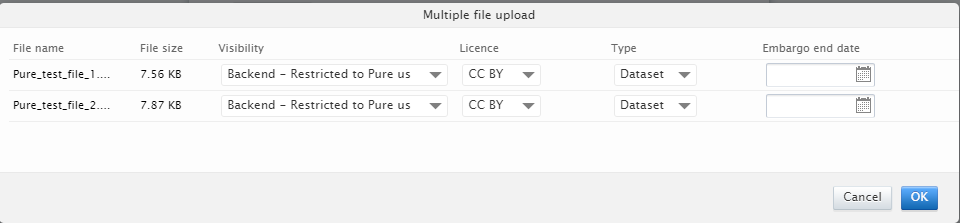


Figure 9 Screenshot of the “Multiple file upload” pop-up window.

Visibility – this should be always set as “Backend” (otherwise files will show on the portal before the intended time). We will wait for the publication of the article before changing the visibility of the data files to “Public” and publishing the data. Please contact us if you are not sure what to set.

License – Please, choose a licence as this will inform the user on the terms and conditions of use. You can use either Creative Commons or Open Data Commons licences. For more information contact us or visit <http://creativecommons.org/> and <http://opendatacommons.org/>.

Type – make sure that the appropriate type is selected. If unsure, select ‘dataset’.

Embargoed until – you can decide when the files will become visible on the portal. Usually the RDM team manually releases them once the article is published.

**Physical data:** here you can describe any physical item that can be made available upon request (e.g. samples, tissues, notebooks, etc.).

**Date made available:** you can simply add the year and we will fill in the rest when the files will become publicly available.

Access option - For use by the RDM team.

### Relation to content

In this section, you will be able to make links to other items in Pure. As a minimum. you should link to the relative article and the grants associated with it. You can create a link by clicking on the “+” icon and searching by Title, Pure ID or Grant number. If you need any assistance, please get in touch.

### Visibility (metadata)

This will dictate if the metadata (general information: Title, description, authors, publisher, etc.) will be publicly visible on the University’s portal. **Important**: This will be set as public for the activation of the DOI.

### Status

By default this is set as “For validation”, you should change it to “Entry in progress” for as long as the record is not complete. Once the final files have been uploaded and you require the activation of the DOI, please, set it as “For validation”.

### Save

Important: remember to save when done or all the changes will be lost

## How to deposit (visual instructions)

Once you have logged in you will be taken to your Personal page within Pure. To create a new dataset/software record, start by pressing the green ‘Add content’ button at the bottom right.

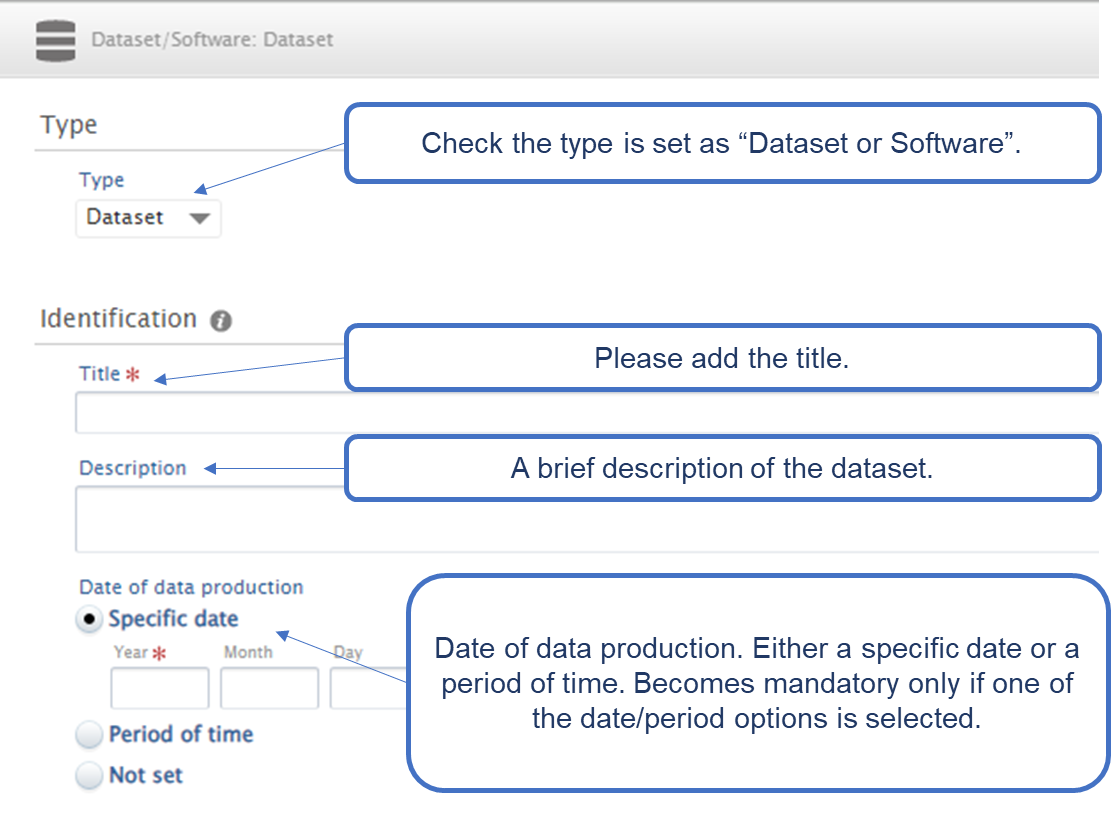
A screenshot of a computer

Description automatically generated

The ‘Choose submission’ screen will pop up, select ‘Dataset/Software’ from the left-hand menu, then choose the ‘Dataset’ option (under Dataset Types).

A screenshot of a computer

Description automatically generated



A screenshot of a chat

Description automatically generated

A screenshot of a chat

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

# University Advice and Guidance

| **Resource** | **URL** |
| --- | --- |
| University RDM website | <https://www.st-andrews.ac.uk/research/support/open-research/research-data-management/> |
| Contact us | Email - [research-data@st-andrews.ac.uk](mailto:research-data@st-andrews.ac.uk) ; Phone – 01334 46**2343** or 01334 46**2322** |
| Licences | <http://creativecommons.org/> ; <http://opendatacommons.org/> |
| Pure | <https://www.st-andrews.ac.uk/staff/research/pure/> |
| Research Portal | <https://risweb.st-andrews.ac.uk/portal/> |
| General Pure enquiries | Pure Live System - [purelive@st-andrews.ac.uk](mailto:purelive@st-andrews.ac.uk) |

# Appendix

With your data, you should provide enough information to give some context. This can take the form of a simple description (in the “Description” field), a more comprehensive ReadMe.txt file or both. You should include basic information on the methodology employed to acquire the data (online surveys, paper surveys, instrumental analyses, etc.) together with type of files, their formats and what software is required to open, visualise and analyse them. If you have any questions or would like some help, please do not hesitate to get in touch.

## Example of Description

The attached data files underpin the publication “*The title of your article*”. The following file types and formats are included:

- X-ray photoelectron spectroscopy files: .vms (can be opened with a text editor)

- X-ray diffraction files: .xrdml (proprietary format) and .udf (can be opened with a text editor)

- UV-Vis spectroscopy files: .jws (proprietary format) and .txt (can be opened with a text editor)

- Transmission electron microscopy: .dm3 (proprietary format) and .jpg

More details on the software required are provided in the ReadMe.txt file.

## Example of ReadMe.txt file

The X-ray Photoelectron Spectroscopy (XPS) data have been acquired using a Kratos AXIS Ultra DLD, with a monochromatic Al Ka (1486.6 eV) X-ray source. The files are in .vms format and can be opened with any text file handling software (Notepad, Word etc.). The proprietary software that would allow for the visualisation of the data and their analysis is CasaXPS (http://www.casaxps.com/).

The powder X-ray diffraction data were collected using an Empyrean PANalytical series 2 diffractometer with a Cu Ka radiation source (l = 1.5406 Å). The files are offered in two different formats:

- .xrdml, an xml coding that can be opened with the Notepad application and with the PANalytical software;

- .udf, that can be opened with any text handling software. It contains details of the analysis and the intensity counts.

To perform analysis on this data files software such as PANalytical may be required.

The UV-Vis spectra were acquired using JASCO-V-650 spectrophotometer. The files are in both proprietary file (.jws) and text file (.txt). The proprietary format requires the software SpectraManager by JASCO to be opened this software will also allow to perform further analysis of the data.

Transmission electron microscopy images were acquired using a Jeol JEM-2011 operated at 200 KV using a Gatan 794 CCD camera to record digital images. The original images are in .dm3 format, which can be visualised with the free software ImageJ but require the Gatan software to be analysed. Images in .jpeg format have also been included.