

# ReCollect

RESEARCH DATA PLUGIN FOR EPRINTS

## Eprints ReCollect User Documentation

*Adapted for the Essex pilot repository implementation*

*March 2013*

1. Why should I use Essex Research Data?
2. Active research and data collection: ensuring readiness
3. Completed research - what to do with your data now
4. Depositing data
  - 4.1 Uploading your data
  - 4.2 Licences
  - 4.3 Describing your data collection
  - 4.4 Terms and conditions, approval and updates
5. Contact information

### 1. Why should I use Essex Research Data?

Essex Research Data was developed as a pilot data repository to provide a means to facilitate the sharing of research data, both within the university and with the wider community. Sharing data is an opportunity to achieve greater impact, value and acknowledgement for such non-traditional research outputs, and to demonstrate to a wider audience your and the university's research achievements.

As a researcher you might want to share research data via a data repository for the following reasons:

- Data underpinning a publication, where reviewers or readers may request access to the data
- Funder or publisher requirement to preserve and share research data
- Legal or other requirement to retain research data after a project ends
- Potential use in future research
- Teaching purposes
- Support decision-making and policy formation

There is also scope to submit high quality data outputs towards academic assessment exercises such as the Research Excellence Framework, so looking after your research data now could be of significant benefit in the future. In more general terms, data sharing is a case of best practise, particularly if the research is publicly funded whereby data can be seen as a public good, contributing to wider societal benefits.

Although we have made the process of depositing data with Essex Research Data as intuitive as possible, we have created this document to provide a step by step guide with additional detail on certain aspects of the process.

## **2. Active research and data collection: ensuring readiness**

To ensure readiness for depositing data in a repository, it is crucial that the careful management of your data begins as early as possible in the data collection process. There are several steps you can take which will ensure that later depositing of your data in a repository is as simple as possible.

The following generic recommendations provide a good start:

- Capture and create metadata during the data collection process. Metadata is essentially data about data. Think about the information required to enable a complete understanding of your dataset., such as variable descriptions or sources of data. This may well be specific to your discipline.
- Check, validate and clean your data as you progress. This is likely to be part of a process you carry out anyway, but a little foresight can go a long way in alleviating workload later on.
- Describe data and the method resulting in its collection in as much detail as possible. Capture this information while it is fresh in your memory, as it may not be as easy to remember that crucial detail when wrapping up the project further down the line.
- Label files appropriately. Come up with a transparent, standardised way of naming your files.
- If your data contains sensitive information, keep a copy that does not contain this information for later sharing. This will reduce time spent on the sometimes meticulous nature of anonymisation.

Detailed recommendations vary from research case to research case. We recommend that you contact the university's research office or use suitable discipline specific resources for guidance. For example, the UK Data Archive's [Create and Manage web pages](#) provide extensive and up-to-date guidance on best practise for looking after your research data throughout a projects lifespan.

## **3. Completed research - what to do with your data now**

Before you begin you may wish to refer back to any data management planning you have carried out prior to or during the project, in order to ascertain any obligations/requirements there may be regarding your data.

Generally, large data collections that require significant investment in storage will not be appropriate for the institutional data repository, Essex Research Data (ERD). ERD is ideally suited for discrete collections that contain explanatory materials and will be of interest to the wider academic world. If you are concerned about the size or complexity of your data collection, please consult the Repository Administrator for guidance.

## **4. Depositing data**

Before you start the deposit process, ensure that you have gathered all the necessary data files and accompanying material. Remember that the intention is for the dataset to be accompanied by information that allows it's re-use by another researcher, so please do provide all necessary materials and descriptive detail, or links to relevant publications.

The information you enter is automatically saved as you move through the deposit process by clicking the 'Next' button. It is not necessary to press the 'Save' button before doing this. If you press 'Cancel' before doing either of these however, you are at risk of losing the information submitted.

Not all the fields are mandatory, so it is important to think carefully about whether any optional fields might be useful in describing your data. See section three for further information on filling in some of the more specialist fields such as those pertaining to geography.

#### 4.1 Uploading data

The first step of the process is to upload the data files and accompanying files. Choose the best option for your kinds of data files and file structure.

- Small number of files (less than 10): upload as individual files
- Large number of files or multiple versions of the same files: group files in zip files in a logical way; grouping options are:
  - Content type (e.g. data / documentation/ metadata)
  - Format (e.g. images / videos / transcripts)
  - Format type (e.g. proprietary format / open format)
  - Dependency (all related files in one zip, maintaining file/folder structure). Sometimes the usability of your data relies on maintaining dependencies between files, possibly including a folder structure.
- In some cases upload all files for deposit in a single zip; this is not encouraged unless your data is very structurally simple, as important information is lost if file level descriptions are not included where possible.

Uploading zip files save you having to tag individual files with file-level metadata, a task which could be time consuming for large numbers of files. Once files have been uploaded, you need to add a small amount of metadata to each file. The most important field is 'Description'. Paired with your choice of 'Content' type, you provide context for each file, and enable understanding the meaning of the files.

Note that you can also set an embargo for each file of up to one year. Should a longer embargo be needed, this can be requested by contacting the Repository Administrator.

#### 4.2 Licences

The key differentiation in licencing research data is between open licences and those reserving certain rights for the creator. We encourage the use of open licences, as these allow the most freedom for reuse, and therefore maximise the potential of data for further impact. There are a variety of options. Below is some information on commonly used licences which are available through the Essex Data Repository.

Thought should be given to which is the most appropriate for your data. Carefully assess the suitability of each to your data – and remember that when dealing with non-sensitive data, granting greater freedom is almost always going to be of benefit to both yourself and others. Licencing options are described in detail in the Digital Curation Centre guidelines.

1. Creative Commons
  - a. Attribution (CC BY)
  - b. Attribution Share Alike (CC BY-SA)
  - c. Attribution No Derivatives (CC BY-ND)

- d. Attribution Non-Commercial (CC BY-NC)
  - e. Attribution Non-Commercial Share Alike (CC BY-NC-SA)
  - f. Attribution Non-Commercial No Derivatives (CC BY-NC-ND)
2. Restrictive Licence

As general guidance, you should aim to release your data under as relaxed a licence as possible, e.g. Creative Commons. This maximises the reuse value of the data.

### 4.3 Describing your data collection

The next stage of the deposit process is to record metadata for the entire data collection you uploaded; descriptive information that applies to all the files. This can for example mean describing your research project from which the data results, or a publication that the data underpin.

There are both mandatory and optional fields here, designed to capture as much information as possible. Help information is available via the white on blue question mark icon alongside each field. Additional guidance is provided here on certain fields that may not be as intuitive to users.

#### Type of data

Here you are asked to choose one or more data ‘types’ relevant to your deposit, from a short list. This provides a useful way to categorising collections and ultimately helps people find your data. While we have tried to account for every possible broad data type, this is challenging due to the breadth of research undertaken at the University of Essex. If you feel there is something important missing, please send your suggestion to the Repository Administrator.

#### Geographic coverage

This field allows you to enter free text, so it is important to try and keep your entry focused. For example, “England; Wales; Scotland” or “Boroughs of Greater London” would be suitable entries. On the other hand, “county” is too vague (which county?), and “Portland” ambiguous due to a lack of context (Portland, Oregon; Portland, Maine; or Portland, Dorset?). You may also include details of spatial resolution in this field. If doing so, be sure to include the unit of measure.

#### Geographic location (WENS bounding box)

These are the four coordinates (west, east, north, south) that ‘bound’ or geographically enclose the geographical area of your data collection. For those familiar with geospatial data this field should be fairly intuitive. For others, there is unfortunately no built-in utility for generating this information, so instead you are presented with several ways of acquiring it manually. The most straightforward, and the one we shall recommend here, is using OpenStreetMap, a free web map service, created by a community of volunteers. Following the steps below, you can generate the four coordinate values required:

1. Navigate to the [OpenStreetMap main page](#). You should see the default map view.
2. At the top of the interface are four text buttons, the last of which is Export. Click on this, and you should see four values arranged in a diamond like shape next to the map. These are your longitude and latitude values, representing the current view in the map window.

3. Adjust your view so that it covers the full geographical extent of your data. For example, if your data were collected in London, you could search for Greater London in the search box and the map window will reorient itself.
4. You can then copy the values representing this area back into the deposit window, ensuring that the WENS labels match.

For those interested in a more in-depth, technical insight, the full metadata profile is in presented in Appendix 1.

### **Legal and ethical issues**

Where applicable, it can be useful to note any legal and/or ethical issues that affect this particular dataset, and how they have been mitigated to enable its availability through the repository.

For example, that the names of research participants have been anonymised via the use of pseudonyms.

### **Data processing/preparation activities**

It can be useful for a potential user to know how the data has been transformed after collection and prior to upload here. For example: the conversion of all files to an alternative file format, the removal of particular identifying variables, and the unpacking of a database into flat files.

### **Original data publication details**

If you have already published this collection through a data journal, subject specific archive or other means, you must provide details of this here. If this is not the case, all you need to do is select 'Unpublished' under the 'Status' field.

### **Editorial note**

In this field you can enter a text note that will only be visible to the repository administrator. This would be an ideal place to include explanations for any of your other actions elsewhere in the deposit process. For example: why you have specified a particular level of access or embargo period for a file.

## **4.4 Terms and conditions, approval and updates**

For the final step of the deposit process, you must agree to the terms and conditions of Essex Research Data before you submit your data. These are displayed during this last stage of the deposit process, and are also available to view in the Repository Policies section of the site. After agreeing and clicking 'Accept', your deposit will be reviewed by an administrator before it appears in the live repository. You may be contacted by the administrator if there are any problems with your deposit. Otherwise, you will receive a notification when your item goes live shortly.

You have the option to update your deposit at any time. Any changes you make will have to be approved by the administrator again. You may either make changes to the existing collection record 'in-situ', or create a new version. The latter is useful if you have made significant changes and the previous version retains its value alongside this new version.

## **5. Contact information**

For further guidance, please contact the Repository Administrator.