

# Guide: Depositing your work in eData – AI for Science collection

## 1. What is eData?

eData is an institutional repository where STFC staff can deposit data and software that support journal articles and other published research. It offers a service for the deposit, registration, preservation, discovery, sharing and citation of these digital objects. eData does not replace subject-based, national, or STFC Facility data repositories already in use. Researchers should check with their funder or facility for any specific requirements on where to deposit data or software.

### 1.2 AI for Science collection

We have created a bespoke subcollection, within eData, for datasets & software that relate to [AI for Science](#) initiatives within STFC. This collection has been developed to meet the needs of the AI for Science theme in Scientific Computing and permission to deposit into this collection is restricted to AI for Science Theme.

## 2. What should be deposited in eData?

eData is intended to store analysed data supporting published research, and any accompanying software created during the course of your research that is necessary to reproduce reported results. It is **not** intended to store:

- data still in active (live) use by research projects
- data that is confidential, sensitive or contains personal information
- software that has been Licenced by STFC for commercial use

Only STFC members of staff may deposit in eData, using their federal ID and password to log in.

## 3. Why use eData?

[The Research Councils' Common Principles on Data Policy](#) advocates the archiving and preservation of research data, and eData allows STFC staff to comply with this. You can also compliantly deposit data funded by the EU and other funders in eData. However, some funders have specific requirements about where data should be deposited. Although eData meets most funder requirements, there may be exceptions. If in doubt, please contact the eData team: [eDATA@stfc.ac.uk](mailto:eDATA@stfc.ac.uk).

## 4. Check your permissions and rights

Deposits in eData are made in accordance with our [Deposit Agreement](#). Before deposit, make sure you have checked with your project team/co-creators/third parties/Intellectual Property team that you have the correct permissions and rights to make your data or software available. Many organisations own the IP to software created by staff in the course of their work; always check the IP status before publishing software online.

## 5. Preparing your deposit

It is best to prepare your data and metadata (the information that describes your item) in advance. This makes the process easier, and ensures your deposit is correct at submission. For projects with a [Data Management Plan](#) (DMP) in place, your deposit should be managed and prepared in line with the DMP. **Please bear in mind that, once licensed, datasets and core metadata should not be changed.**

Software / code should be formatted in a consistent manner, clearly organised and carefully documented; it should be packaged with everything necessary to allow users to read and run the software, including licence information.

Please see the [Preparing your data](#) and [research data management](#) pages on The Source for further information about preparing your data/software for publication.

### 5.1 Deciding what data to publish

Before publishing, you should consider what data and/ or software would be required to understand, replicate or build upon the research and the long-term value of your data.

The following should be deposited into a repository, where possible:

- Data and software that underpins a paper or publication
- Data that cannot be reproduced, or would be too expensive to reproduce
- Data and software that could have significant long-term value and/or be reused by other researchers, both within STFC and the wider scientific community
- Software, code or script that has been produced specifically for a study. Particularly where the code has been used to create the dataset or would be required to replicate or validate the study. This includes both software and any plug-ins, extensions or applications written for existing software.

Depending on the nature of the study, AI for Science staff may be required to publish a combination of the of the following:

- Input / training data
- Software / code / script used to process, analyse, or create the data
- Output data

**Please note:** eData should only be used to publish data where STFC is the rights holder for the data. Additionally, it may not always be appropriate to publish all data associated with a particular study.

For this reason, **we request that you contact the eData team ([eData@stfc.ac.uk](mailto:eData@stfc.ac.uk)), before you begin a new deposit**, so that we can provide advice and guidance on what data you need to publish.

## 5.2 For studies producing both Data and Software

If your study has produced both data and software, they must be submitted to eData as separate deposits, one for the data file(s) and one for the software file(s). This is because different usage licences must be applied to data and software.

Please submit your software to eData following the procedure outlined in section 6. Starting a new submission in eData. However, please upload the software file(s) into eData, rather than providing the File URL.

## 5.3 Software plug-ins and applications

If you have created any form of custom plug-in, application, extension or code for existing software as part of your research, this will also need to be published, especially if it is required to replicate your study and/or reuse your data.

When publishing your plug-in, please follow the instructions outlined in section 6 (Starting a new submission in eData) and provide details of the software the plug-in was written for. Where possible, please include:

- **Software Title** – The name of the software the plug-in was written for
- **Publisher** – The organisation who published the original software
- **Version** – The specific version of the software the plug-in was created for
- **Details of how to access the software**
  - **For Open Source software** – please provide the DOI for the specific version of the software used
  - **For Licenced Software** – please provide details of who to contact to request a copy of the software or how to purchase the software, where possible.

## 6. Starting a new submission in eData

You will first need to sign in to eData using your federal ID and password. If you are signing into eData for the first time, the eData team will need to activate your account and assign your user permissions before you can begin depositing data. You will receive an email from [edata@stfc.ac.uk](mailto:edata@stfc.ac.uk) once your account has been activated.

Note: user permissions are manually assigned by eData administrators and may take 1-2 working days to be actioned.

**Before beginning a new submission: Please contact [eData@stfc.ac.uk](mailto:edata@stfc.ac.uk) to discuss what data/software needs to be deposited.**

To begin a new submission, hover your mouse over the menu bar on the left of the screen and select new item. Then select the collection you wish to submit to.

### 6.1 File upload

Note: You only need to upload files, when depositing Software, code or script. If your eData submission relates to Data files you do not need to upload the files and will instead need to add the location of the file(s) to the File URL field – [See section 6.12 File URL](#).

eData will accept deposits of up to 10GB. Please contact the eData team for larger deposits. Either drag and drop the files you wish to deposit, or click browse.

Try to give files meaningful names that reflect the file content, avoiding spaces and special characters.

### 6.2 Title

Make the title of your work meaningful and distinctive, especially if you have a number of similar items to deposit. Where possible, the title should summarise the content of the dataset and should include key words to help make your dataset discoverable.

**Please be aware:** that the title of your dataset will form part of the citation, should your data be reused or referenced in another study.

### 6.3 Creators

Please enter the name(s) of the creator(s). At minimum, please enter the creator(s) last name and first initial (eg: Smith, A). Creators with an ORCID can search for their ORCID account; enter the creators name, search and then select the correct record from the drop-down list. Alternatively, it is possible to search using an ORCID ID only: enter the ORCID into the last name field and select search.

We recommend that you record the **ORCIDs** of all the creators (if applicable) and have them to hand when you start your deposit. ORCIDs provide a persistent digital

identifier that distinguishes each individual researcher. [Register here for an ORCID](#) if you don't already have one.

## 6.4 Description

Describe and document your work thoroughly so that it can easily be found, understood and used by others. Imagine what someone without your specific expertise would need to know about the data or software, if they were viewing it for the first time. Using the abstract of an accompanying article is not the best way to describe your data. The description should include some explanatory information about your research methods, describing how, why and when your data were created, what your data are, and whether you used any specific software. Spell out any acronyms used, and explain the labels of any variables and values.

Where possible, the description entered into this field of the submission form should be between 50 and 250 words in length. Detailed descriptions can be contained within a 'README' file – preferably in plain text format. Further guidance on [writing a README file for research data](#) can be found on The Source or contact [eData@stfc.ac.uk](mailto:eData@stfc.ac.uk) for an example template.

## 6.5 Keywords

Think carefully about which keywords to use: imagine what you would look for if you were performing a search for your work. Strike a balance between general and very specific/specialist search terms.

## 6.6 Date of Issue

Is the date that your dataset was first published or publicly distributed. For most datasets this will be the date you submitted it to eData. For datasets that are under embargo, please enter the date the embargo will end and the data will become publicly available. If you are entering legacy data, please enter the date the dataset first became publicly available.

This is a mandatory field and will form part of the data citation. You must enter the year at minimum. However, the day and/or month can be left blank if they are not applicable. **Please Note:** entering a future date into this field will not apply an embargo to your dataset. Please see the Access conditions section of the form to apply an embargo.

## 6.7 Type

Using the drop down list, please select the content type of the item(s) uploaded to eData. If more than one type applies, please select collection.

Type definitions:

Item Type	Definition
<b>Audiovisual:</b>	A series of visual representations imparting an impression of motion when shown in succession. May or may not include sound.
<b>Collection:</b>	An aggregation of resources, which may encompass collections of one resource Type as well as those of mixed types. A collection is described as a group; its parts may also be separately described.
<b>Dataset:</b>	Data encoded in a defined structure.
<b>Image:</b>	A visual representation other than text. Digitised or born digital images, drawings or photographs.
<b>Model:</b>	An abstract, conceptual, graphical, mathematical or visualization model that represents empirical objects, phenomena, or physical processes.
<b>Software:</b>	A computer program in source code (text) or compiled form. Use this type for all software components supporting scholarly research.
<b>Sound:</b>	A resource primarily intended to be heard.
<b>Workflow:</b>	A structured series of steps which can be executed to produce a final outcome, allowing users a means to specify and enact their work in a more reproducible manner. Examples include: Computational workflows involving sequential operations made on data by wrapped software and may be specified in a format belonging to a workflow management system, such as Taverna

\*Metadata Working Group. (2019). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.3. DataCite e.V. <https://doi.org/10.14454/7xq3-zf69>. Pages 38 – 40

## 6.8 Language

This is the primary language of the resource. **Note:** this relates to the written language e.g. English, French etc and not the computing language. This field is optional and can be left blank.

## 6.9 Linking related publications

Please link your eData deposit to any related published works, e.g. journal articles, which reference your data. You can add the related article DOI to the eData record in the 'Referenced by' field. Please enter DOI(s) as a URL e.g. <https://doi.org/10.xxxx/xxxx>.

**If you don't know the DOI of an accompanying article until after you have submitted your dataset, please contact the [eData team](#) and we will add it to the metadata.**

For data that supports a journal article or publication, it is good practice to include a [data access statement](#) in your manuscript. The purpose of the statement is to tell the reader where the supporting data can be found and under what conditions it can be accessed.

## 6.10 Funders

It is important to credit the funder of the work. Keep a record of funder names and grant numbers prior to deposit as you will be asked to provide this information.

## 6.11 Access Conditions

This field can be used to apply an embargo to your dataset. The embargo hides the files within your submission for a defined period of time (although the metadata will still be openly available). Consider whether you need to embargo your work. Various funders have different policies on this and you should [consult the policy of your funder](#) directly. It may not always be possible to make some data open for legitimate reasons; any restrictions should be justified and justifiable.

## 6.12 File URL

Please use this field to provide details of where the data file(s) are located within echo. Please enter the location of the data file(s) within echo, as a URL, in the following format: <https://s3.echo.stfc.ac.uk/<bucket-name>/<object-key>>.

**Note:** you must provide the location for each individual file and not just the bucket. For datasets that contain multiple files, you will need to create a zipped folder and provide the location of the zipped folder.

## 6.13 Licenses

A license must be applied to all datasets deposited in eData. The license you select will determine how your data can be used by others. Creative Commons [CC-BY](#) licence is the standard licence approved by most funders for data.

[Creative Commons Licenses](#) **are not** suitable for software and an open source software license will need to be applied. If you wish to use an alternative licence for your deposit, please select "Other". **Any special licences or conditions of use must be documented in a plain text file uploaded alongside the dataset.** You will require line management approval before applying an open source license and depositing your software in eData.

Software with commercial potential or that has been licensed by STFC for commercial use, must not be deposited in eData. If you are unsure about commercial potential or appropriate software licences, please contact [eData@stfc.ac.uk](mailto:eData@stfc.ac.uk) or use the [STFC Inventor portal](#) to request a licence from the Intellectual Property team.

**Note:** The work can only be licensed by the copyright holder or under legal authority from the copyright holder. Co-creators must be consulted before a license is assigned. Once assigned, a license is permanent and cannot be revoked, nor can the content be changed.

Before depositing your dataset, you should refer back to your project documentation, collaborative agreement or Data Management Plan. For some projects, the license under which the dataset can be shared, may have been agreed or outlined at the beginning of the project.

## 7. Submission review and approval

**Note: All submissions to the AI for Science collection will be reviewed by an AI for Science Collection Administrator following submission.** The review process may take 1-2 working days to be completed. During this time, your submission will not be visible within eData.

Once reviewed the editor will either:

- Approve the submission  
Once approved, the deposit record will go live in eData and the DOI will be created.
- Reject (and return the submission back to the depositor) with comments  
The editor will return the submission back to the depositor and provide details of any changes / edits that need to be made before the submission can be approved. This may include; providing additional metadata (contextual information), checking the file location has been provided and/or a file(s) have been uploaded etc..
- Reject (stating that the submission is unsuitable for eData)  
If a deposit has been reviewed and been found to be unsuitable for inclusion in the collection, it may be rejected, stating the reason for rejection and alternative options (such as submitting to an alternative departmental collection) where appropriate.

Following the review, you will receive an automated email stating that your deposit has either been approved or rejected and any changes that need to be made (where applicable).

Once a deposit has been approved, the deposit will become visible within eData ("go live") and the DOI will be minted. **Once approved, you will be unable to make any changes to either the data / software files or the mandatory citation metadata.**

## 8. Citing datasets and measuring impact

Your deposit will be assigned a DataCite **DOI** after submission, usually within 4 hours. This will appear in the **URI** field of the record. Always send your publisher a '[Data access statement](#)' referencing the DOI for your underlying data so that people know how to access it.

To cite your data, you may wish to use the following format: *Creators. Publication Year. Title. [format and/or medium]. Publisher/Repository. DOI.*

For further advice, see the [Digital Curation Centre](#) website. If you wish to search for a particular DOI in eData, enclose it in speech marks "10.5286/edata/527". You can see if other people have cited your data using the **Data Citation Index**, via [Web of Science](#) (requires an [OpenAthens](#) account).