

Annotating Metabolic Processes: Instructions for Using Argo

Prerequisites

This task requires some proficiency in using Argo. Please make sure that you know how to:

- set up component parameters,
- run workflows,
- track the progress of processing workflows and open the Manual Annotation Editor,
- use the Manual Annotation Editor.

Most of these actions are demonstrated in the step-by-step instructions below; however, you are encouraged to visit the [tutorials page](#) before continuing.

Workflows

For the purpose of evaluating the usability of Argo, there are three workflows prepared for the curators:

1. **Manual annotation:** This workflow reads PubMed abstracts and opens a Manual Annotation Editor for a curator to tag the entities of interest without any support from automatic processing available in Argo.
 - Input: PubMed abstract identifiers.
 - Output: Annotation files in XML (interchangeable) format.
2. **Automatic annotation:** This workflow reads PubMed abstracts and performs automatic recognition of the entities of interest. This workflow is purely automatic and does not involve any manual intervention from the curators. The objective of this workflow is to automatically “pre-annotate” the input abstracts for later manual inspection.
 - Input: PubMed abstract identifiers.
 - Output: Annotation files in XML format.
3. **Manual correction:** This workflow reads files that already contain annotations (coming from the second, automatic workflow) and opens a Manual Annotation Editor for a curator to correct (remove, add, modify) the automatically recognised annotations.
 - Input: Annotation files in XML format.
 - Output: Annotation files in XML format.

Although the second and third workflow could be combined into a single workflow, for the purpose of evaluation they are defined separately. The three workflows are publicly available in Argo.

Step-by-step Instructions

For convenience, you are advised to open Argo’s main application in a separate window and keep these instructions and the annotation guidelines handy.

Launch the Argo application in a separate window now!

A. Initial setup

The initial setup involves signing in to or registering an account with Argo, as well as creating folders that will be used to output annotations to. You will also make copies of publicly available, read-only workflows so that you can make some adjustments on them.

[Hide initial setup step-by-step instructions](#)

Step 1. Sign in

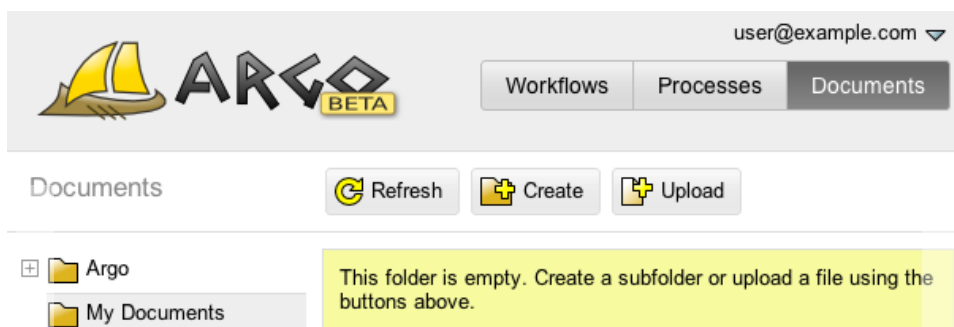
If you have not done so yet, create an account in Argo and sign in to it. Although it is possible to use Argo without registering, any user-created data (workflows and documents) will be automatically deleted at the end of a visit.

Tip: Registering an account in Argo requires a valid email address for verification that is sent immediately after creating an account. Please check your email spam/junk folders if you do not receive the verification email within minutes.

Step 2. Set up output folders

Each of the three workflows results in creating files containing annotations. Create the following three folders in your account's Documents view:

- Manual annotation
- Automatic annotation
- Manual correction



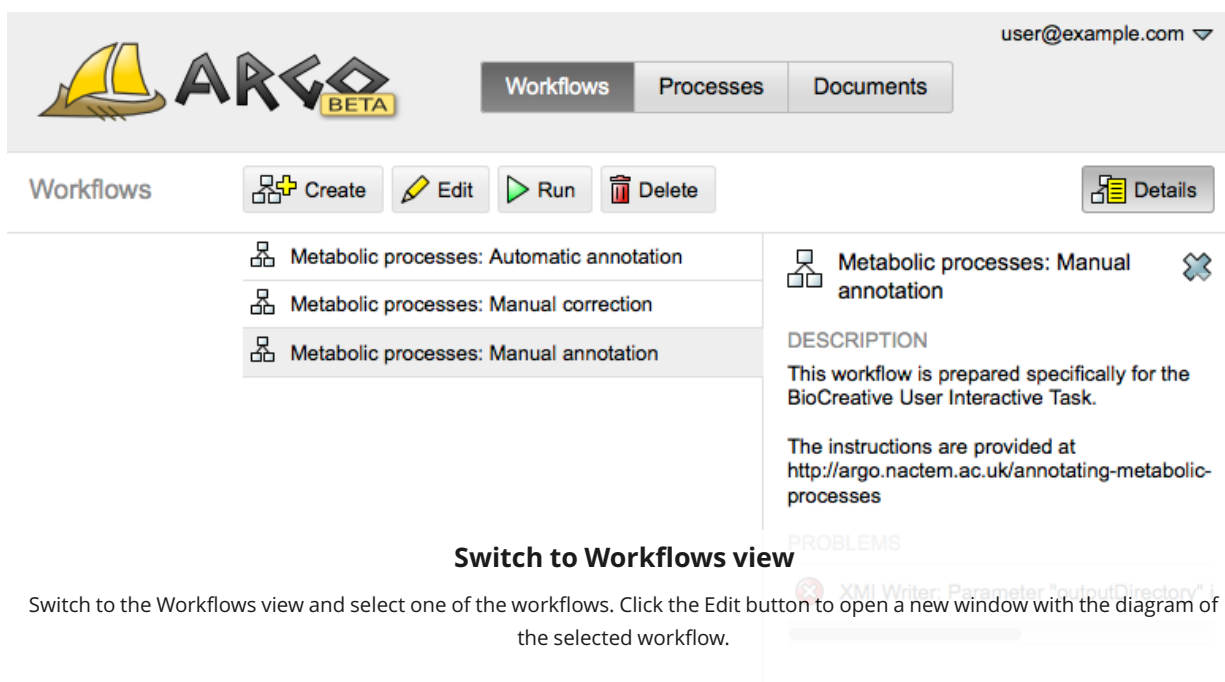
Switch to Documents view

Switch to Documents view and make sure the "My Documents" folder is selected in the left-hand-side panel. To create a new folder click the "Create" button.

Step 3. Create copies of the public workflows

The three workflows are publicly available for reading only. In order to be able to change their settings (for example, selecting an output folder for generated files) you have to make copies of these workflows.

Edit each of the workflows and save its copy. Although it is possible to have multiple workflows with the same name in Argo, we advise you to change the names of copies by, for example, preceding each of them with the "Copy of" prefix.



Switch to Workflows view

Switch to the Workflows view and select one of the workflows. Click the Edit button to open a new window with the diagram of the selected workflow.

B. Manual annotation

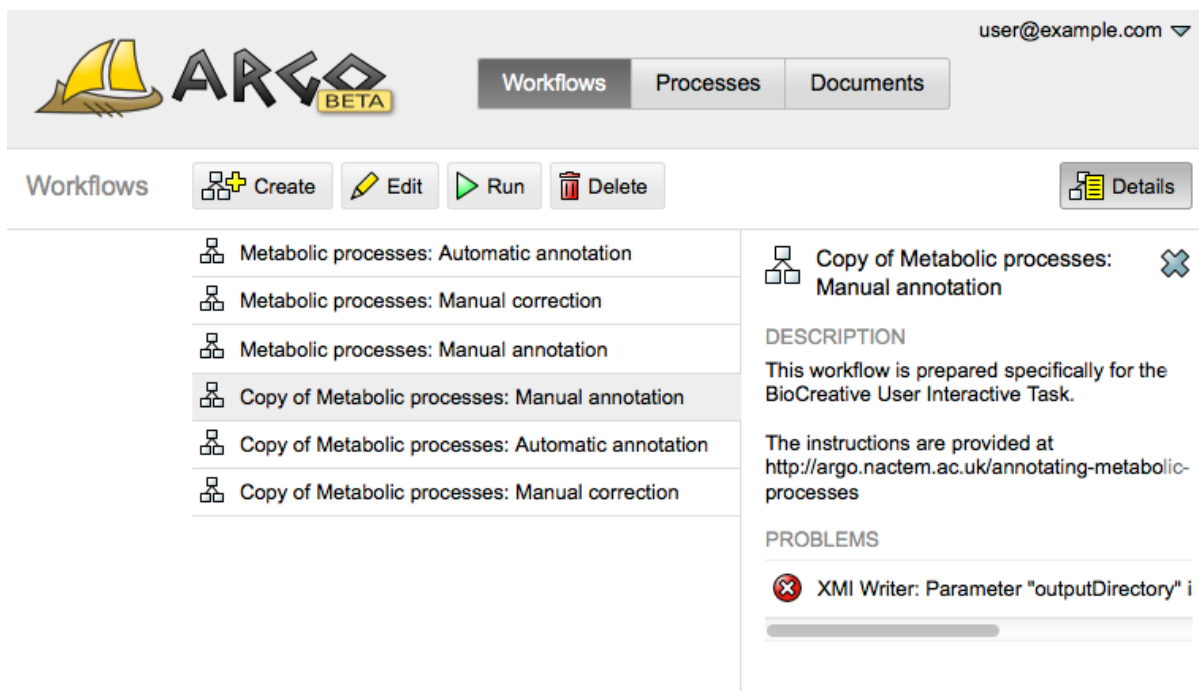
This phase involves configuring the manual annotation workflow by setting up input PubMed IDs and folders for output annotations as well as running the workflows and performing the manual annotation.

[Hide manual annotation step-by-step instructions](#)

Step 4. Manual annotation: Change the workflow's settings

The first workflow that you will work with is the manual annotation workflow. This workflow takes a number of PubMed abstract identifiers as input and generates annotation XMI files on its output.

- Set up the list of PubMed IDs of the PubMed Abstract Reader component (the first one in this workflow) to the first 5 IDs shown below.
- Set up the output folder of the XMI Writer component (the last one in this workflow) to the “Manual annotation” folder you created in Step 2.



The screenshot shows the ARGO BETA web interface. The header includes the ARGO logo, a user dropdown (user@example.com), and navigation tabs for Workflows, Processes, and Documents. Below the header, there are buttons for Create, Edit, Run, Delete, and a Details button. A list of workflows is displayed, with 'Copy of Metabolic processes: Manual annotation' selected. To the right, the details for this workflow are shown, including a description and a list of problems.

Workflows

- Metabolic processes: Automatic annotation
- Metabolic processes: Manual correction
- Metabolic processes: Manual annotation
- Copy of Metabolic processes: Manual annotation**
- Copy of Metabolic processes: Automatic annotation
- Copy of Metabolic processes: Manual correction

Copy of Metabolic processes: Manual annotation

DESCRIPTION

This workflow is prepared specifically for the BioCreative User Interactive Task.

The instructions are provided at <http://argo.nactem.ac.uk/annotating-metabolic-processes>

PROBLEMS

- XMI Writer: Parameter "outputDirectory" i

Select workflow

Select the manual annotation workflow and edit it by clicking on the Edit button.

PubMed abstract identifiers for the manual annotation workflow:


10502406	11284709	11393269	11590700	12547826
12821135	12972033	14709628	15126328	15547048
16081671	16216874	16293890	16479044	16601807
16980304	17520698	17698513	17977830	18664505
18759268	19161989	19442645	19786063	20155626
20467561	20615392	20876576	21468576	21804305

Note: Although it is possible to supply the entire PubMed ID list into the PubMed Abstract Reader component, it is not advisable. Argo will not save any of the annotated documents to the output folder before they are *all* complete. This comes from the fact that Argo allows you to revisit annotated documents and keeps the changes in an internal temporary location until you decide that all documents are complete (that is, there is no more changes to be done). Since the document set is fairly large, it may take days to annotate all of them which carries the risk of losing your work if Argo encounters technical problems. It is therefore advisable to annotate this set in smaller chunks.

Step 5. Manual annotation: Run the workflow and perform manual annotation

Run the manual annotation workflow and launch the Manual Annotation Editor. The editor will allow you to mark mentions of the entities of interest. The entity types present in the editor to choose from are Chemical, GeneOrGeneProduct, and Process. All of them have exactly the same features (or attributes). The *begin* and *end* features will be set automatically once you have marked a span of text. The *id* feature, however, has to be set manually. This feature is an external resource identifier of the marked span of text as defined in the [guidelines](#).

Once the manual annotation is done, all output XMI files should be visible in the “Manual annotation” folder.


user@example.com

Workflows
Processes
Documents

Workflows

Create
Edit
Run
Delete
Share

Details

<div> Metabolic processes: Automatic annotation </div> <div> Metabolic processes: Manual correction </div> <div> Metabolic processes: Manual annotation </div> <div> Copy of Metabolic processes: Manual annotation </div> <div> Copy of Metabolic processes: Automatic annotation </div> <div> Copy to Metabolic processes: Manual correction </div>	<div> Copy of Metabolic processes: Manual annotation </div> <div> DESCRIPTION This workflow is prepared specifically for the BioCreative User Interactive Task. The instructions are provided at http://argo.nactem.ac.uk/annotating-metabolic-processes </div>
---	--

Run workflow

Tip: The Manual Annotation Editor window can be closed and reopened multiple times without losing any previous annotations. This is useful if you decide to take a longer break and close your browser. Next time, simply go back to the running workflow (in the Processes view) and click on the Launch Manual Annotation Editor button (as shown in the slides above).

C. Automatic Annotation

This phase involves configuring the input and output annotation folders of the automatic annotation workflow and running it.

[Hide automatic annotation step-by-step instructions](#)

Step 6. Automatic annotation: Change the workflow's settings

The second workflow you will run is the automatic annotation workflow. Similarly to the previous workflow, this workflow takes a number of PubMed abstract identifiers as input and generates annotation XMI files on its output.

- Set up the list of PubMed IDs of the PubMed Abstract Reader component (the first one in this workflow) to the first 5 IDs shown below (note that they are different from those in Step 4).
- Set up the output folder of the XMI Writer component (the last one in this workflow) to the "Automatic annotation" folder you created in Step 2.

PubMed abstract identifiers for the automatic annotation workflow:

11097864 11375903 11453994 11723127 12635099

12960407 1435739 15115889 15135306 15955068

16156861 16236141 16426572 16563290 1666626

17483544 17619801 17848950 18515279 18729198

19053182 19221000 19464575 19805354 20187624

20493829 20705923 20937724 21743969 22046279

Step 7. Automatic annotation: Run the workflow

Run the automatic annotation workflow. The XMI files containing the automatically generated annotations will be placed in the "Automatic annotation" folder.

D. Manual correction

This phase is similar to the manual annotation phase. The difference lies in the source of input annotations which comes from the output of the previous, automatic annotation workflow.

[Hide manual correction step-by-step instructions](#)

Step 8. Manual correction: Change the workflow's settings

The third workflow you will work with is the manual correction workflow. This workflow takes a number of annotation files in XMI format as input (in this case those produced by automatic annotation in Step 7) and generates annotation XMI files on its output.

- Set up the input folder of the XMI Reader component (the first one in this workflow) to the "Automatic annotation" folder.
- Set up the output folder of the XMI Writer component (the last one in this workflow) to the "Manual correction" folder you created in Step 2.

Step 9. Manual correction: Run the workflow and perform manual correction

Run the manual correction workflow and launch the Manual Annotation Editor. The editor will allow you to correct the annotations generated by the automatic annotation workflow. Following the annotation [guidelines](#), you can add missed annotations, remove spuriously recognised ones, change labels (e.g., from Chemical to GeneOrGeneProduct), change spans and change values of the *id* feature as necessary.

Once the manual correction is done, all output XMI files should be visible in the "Manual correction" folder.

Frequently Asked Questions

• Can I try the workflows with other PubMed IDs for practice?

Yes, you may enter any (preferably pertaining to metabolic processes) PubMed IDs as input to the workflows. Alternatively you can select some from the list below:

21223739	22064374	22585829	3037304	7906730
----------	----------	----------	---------	---------

8076368	8445979	8558417	8915012	9972481
---------	---------	---------	---------	---------

• What should I do if a recognised entity cannot be identified in ChEBI/UniProt/CTD?

Nothing. It is expected that some entities will be new to the external dictionaries. They will be automatically pulled from the annotations in post-processing (not part of this curation task) and will constitute a set of new entries for the external dictionaries.

• How can I view or change annotations at a later visit?

As long as the workflow with a Manual Annotation Editor is running, you can view or make changes to any of the documents that are part of this workflow by switching to the Processes view, identifying the right process, and clicking on the Launch Manual Annotation Editor button.

If the process has finished, you can still view or change annotations; however, you have to create or edit another workflow. Notice that the "Metabolic processes: Manual correction" workflow is exactly what is needed to accomplish it. Simply edit this workflow, create a copy, rename it to something memorable, and change the input of the XMI Reader to the files/documents you want to view/edit. For the XMI Writer, if you wish to overwrite the input files, select the same folder and tick the "overwrite" checkbox.

If you only want to view the files, remove the XMI Writer component from the workflow altogether.

• Can I see if XMI files contain annotations in the Documents view?

XMI files contain annotations even if you have not created any explicitly; they include document metadata. Currently, the Documents view does not visually differentiate between XMI files with "default" annotations and those with added annotations.

• Can I download the annotations?

Yes. Each of the workflows includes the XMI Writer component that writes the original text and annotations to your document space. Simply switch to the Documents view, locate and select the documents you wish to download and click on the Download button in the toolbar.

Note: Currently you may download only one document at a time.