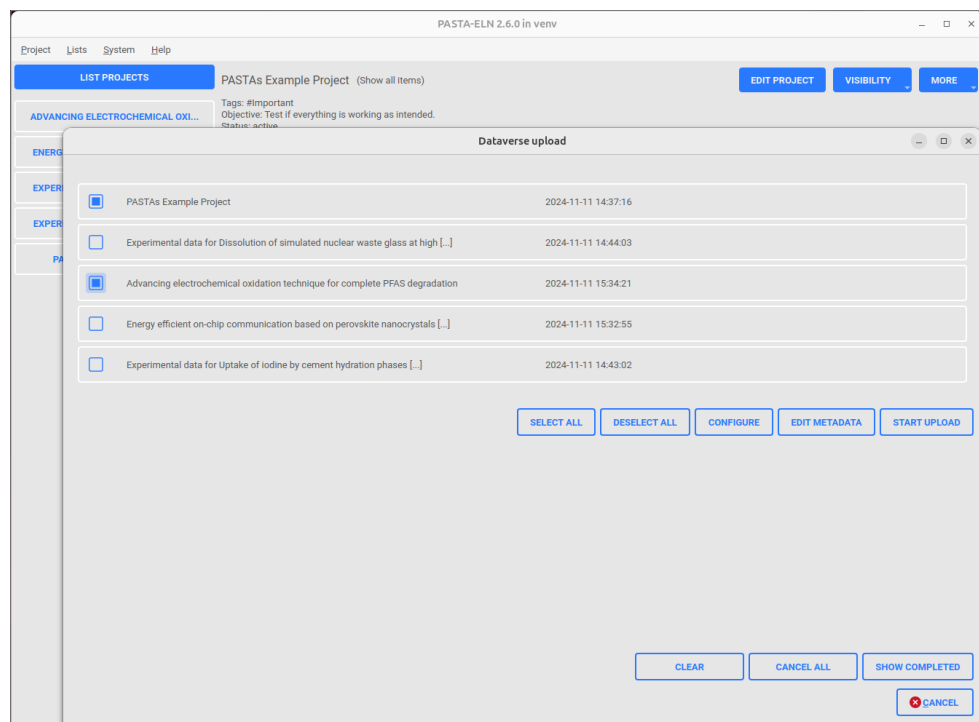


DATAVERSE INTEGRATION

26 NOVEMBER 2024 | JITHU MURUGAN

DATAVERSE INTEGRATION

PASTA project upload to dataverse



PASTA Application

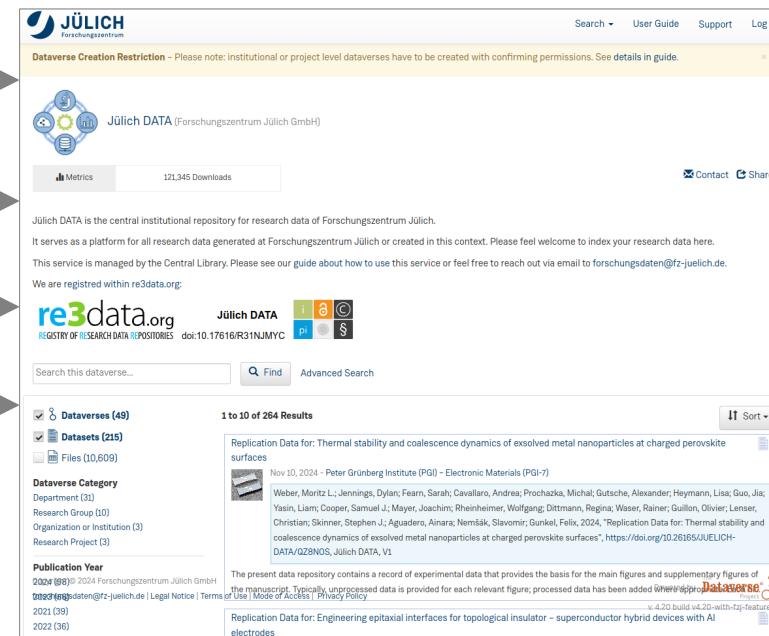
Data-verse configuration
(Auth Rest API Token)

Dataset creation with
metadata population

ELN File generation

ELN File upload

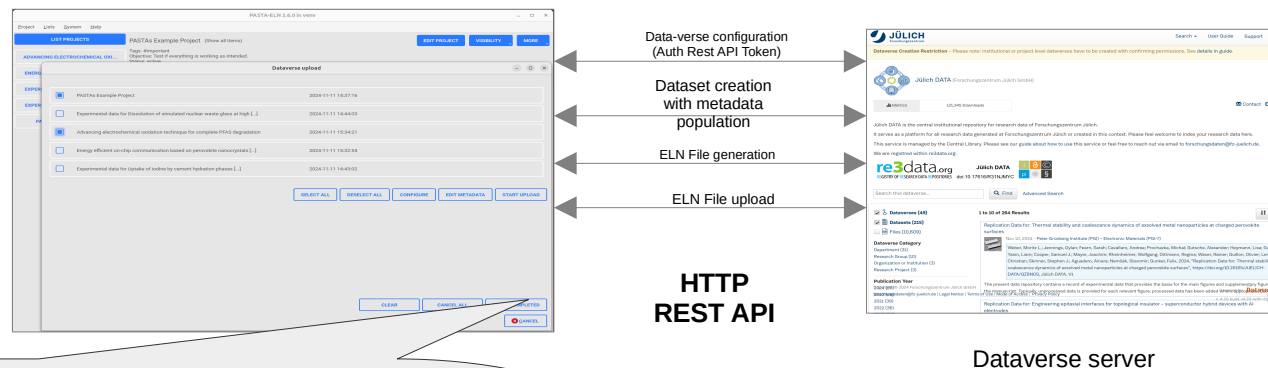
HTTP REST API



Dataverse server

DATAVERSE INTEGRATION

PASTA project upload to dataverse



Features

- Dataverse Login Configuration Dialog
- Metadata editor dialog
- Project upload configuration dialog
- Upload Queue: Parallel upload of multiple project
- Upload history dialog

DATAVERSE INTEGRATION

How to use dataverse upload feature?

- Open “Dataverse Configuration” dialog
- Enter server URL & API Token combination
- Click on “Verify and load” button to verify the given combination and load the available dataverse
- Select the needed dataverse ID
- Click Save button

1

- Open “Upload to dataverse” dialog
- Click “Configure” button to edit the upload configuration
- Save the configuration.
- Click on “Edit Metadata” button to enter the metadata needed for dataverse. Make sure to at the least enter the minimal set.
- Click save to see the summary and save the metadata successfully.

2

- Click “Show Completed” button to see the data upload history
- The upload history will be shown in a sorted order (descending: finished time)
- Enter search term to filter the upload by any field displayed in UI

4

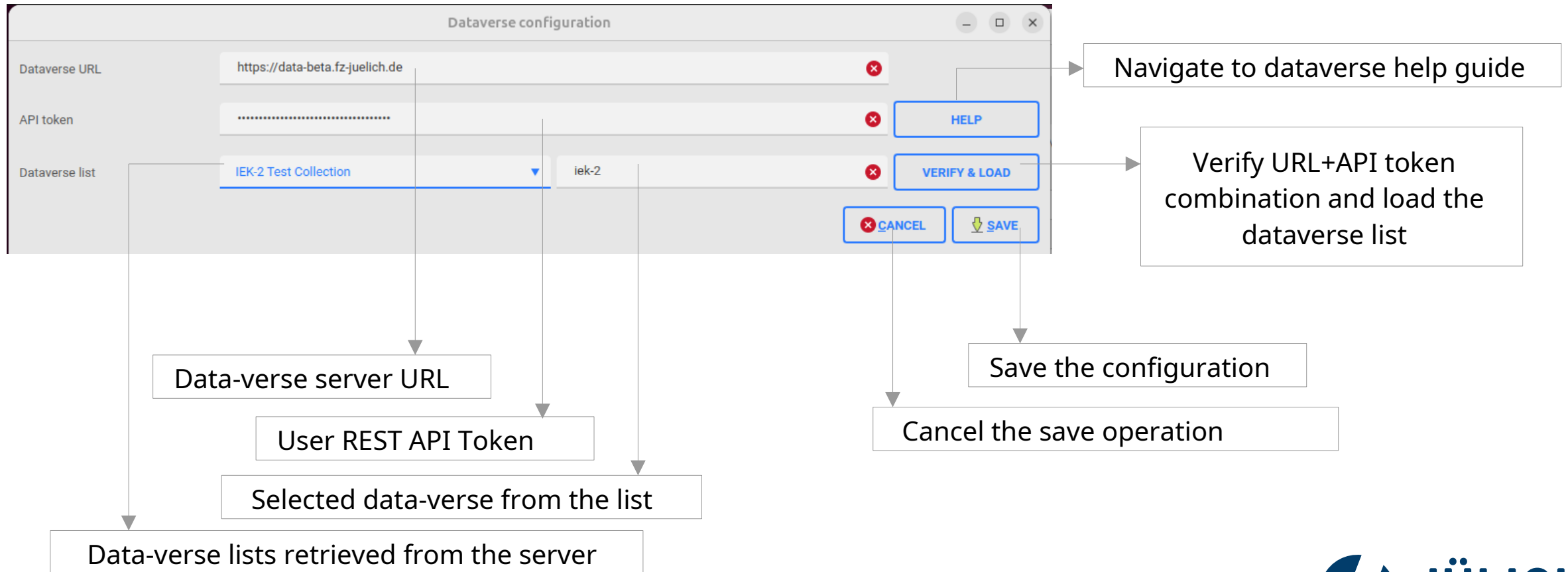
- Select the projects to be uploaded to the dataverse.
- Click “Start Upload” button to enqueue the projects to upload queue.
- Click on “Show log” for each project in the queue to see individual upload log
- Click “Cancel” of individual projects to cancel the ongoing upload
- Click “Cancel” all to cancel all the ongoing upload in the queue
- Click “Clear” button to clear all the finished uploads from the queue

3

DATAVERSE INTEGRATION

Dataverse Configuration

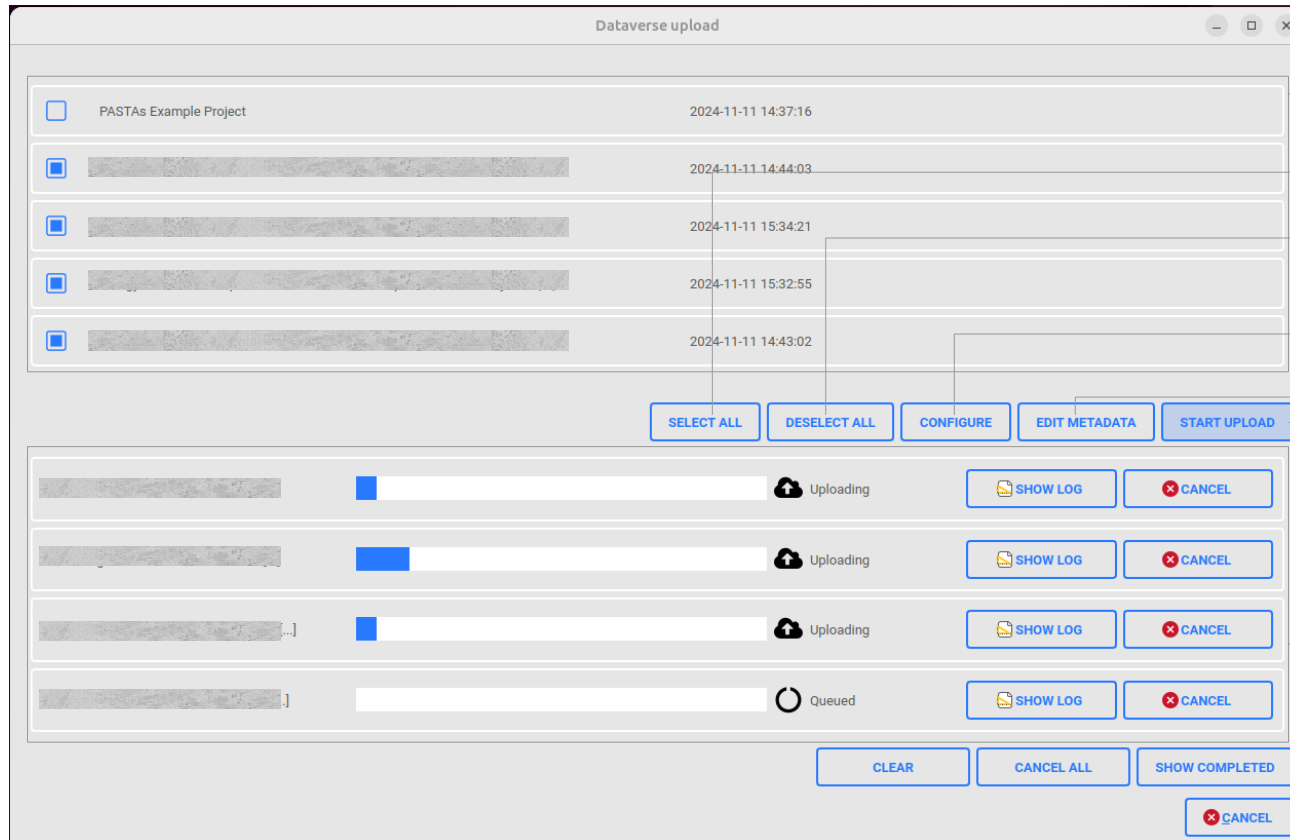
- Run Pasta Application
- Go to “Project group | Dataverse Configuration” or press F10



DATAVERSE INTEGRATION

Dataverse Upload Page: UI Elements

- Run Pasta Application
- Go to “Project group | Upload to Dataverse” or press F11
- Note: Make sure that the Configuration exists, otherwise will be taken to configuration page



Projects Queue: List of PASTA Projects

Select all projects in projects queue for upload

De-select all projects in projects queue for upload

Opens dialog to configure the project upload properties

Opens dialog to edit the metadata for the dataset.

Start upload of the selected projects in project queue

Upload Queue: List of PASTA Projects being uploaded

DATAVERSE INTEGRATION

Dataverse Upload Page: UI Elements

- Run Pasta Application
- Go to “Project group | Upload to Dataverse” or press F11
- Select the projects from the “project queue” and click “Start Upload” to begin the upload.

The screenshot shows a web application window titled "Dataverse upload". At the top, there is a table with columns for "Project Name" and "Project Created time". Below the table, there are several project entries, each with a checkbox, a progress bar, and a status icon. A "START UPLOAD" button is highlighted with a red box. Below the project entries, there are buttons for "SELECT ALL", "DESELECT ALL", "CONFIGURE", "EDIT METADATA", and "START UPLOAD". At the bottom, there are buttons for "CLEAR", "CANCEL ALL", "SHOW COMPLETED", and "CANCEL".

Callouts on the right side of the image point to various UI elements:

- Project Name
- Project Created time
- Upload Project Name
- Upload Project progress
- Upload Project status (uploading/finished/error)
- Click to expand/collapse upload project log information
- Cancel project upload
- Clear all finished items in the upload queue
- Cancel all running items in the upload queue
- Cancel all and close window

DATAVERSE INTEGRATION

Dataverse Upload Page: UI Elements

- Run Pasta Application
- Go to “Project group | Upload to Dataverse” or press F11

The screenshot shows the 'Dataverse upload' window. At the top, there is a table with columns for checkboxes, project names (some redacted), and timestamps. The second row is selected. Below the table are buttons for 'SELECT ALL', 'DESELECT ALL', 'CONFIGURE', 'EDIT METADATA', and 'START UPLOAD'. A log window is open for the selected project, showing a 'Finished' status and a 'SHOW LOG' button (highlighted with a red box). The log text includes: 'Upload initiated for project...', 'Step 1: Generating ELN file for project...', 'Successfully generated ELN file: /tmp/tmpj9eandrd/ExperimentaldataforDissolutionofsimulatednuclearwasteglassathighsurfaceareatosolutionvolume.eln', 'Step 2: Creating dataset...', 'Dataset creation succeeded with PID: doi:10.5072/FK2/OIUZWR', 'Step 3: Checking if dataset is unlocked after the publish operation.', 'Dataset with PID: doi:10.5072/FK2/OIUZWR is locked. Lock type: finalizePublication, Message: Publishing the dataset; Registering PIDs for Datafiles; Validating Datafiles Asynchronously', 'Dataset with PID: doi:10.5072/FK2/OIUZWR is locked. Lock type: finalizePublication, Message: Publishing the dataset; Registering PIDs for Datafiles; Validating Datafiles Asynchronously', 'Dataset with PID: doi:10.5072/FK2/OIUZWR is unlocked already!', 'Step 4: Uploading file /tmp/tmpj9eandrd/E...olutionvolume.eln to dataset with PID: doi:10.5072/FK2/OIUZWR....', and 'ELN File uploaded successfully, generated file PID: doi:10.5072/FK2/OIUZWR'. At the bottom of the log window are buttons for 'CLEAR', 'CANCEL ALL', and 'SHOW COMPLETED'. A 'CANCEL' button is also present at the bottom right of the main interface.

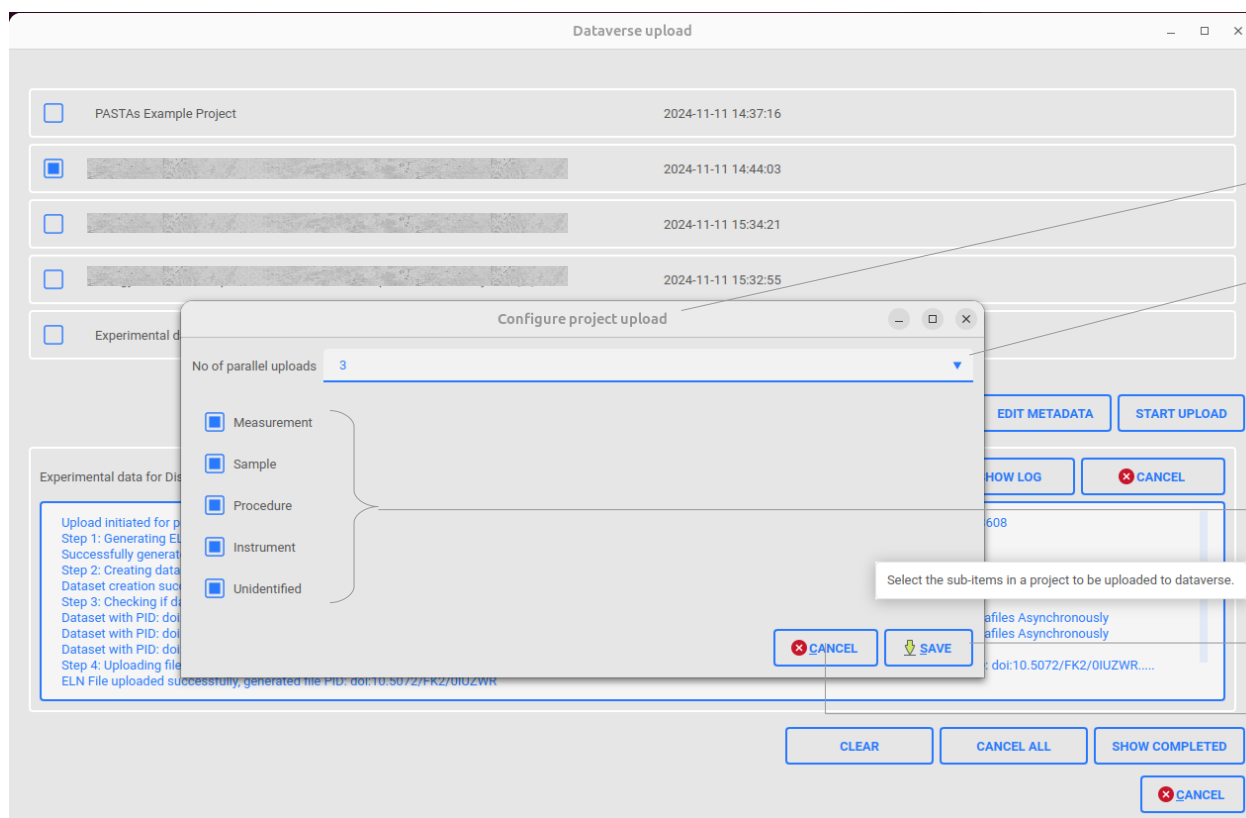
Click on SHOW LOG button to expand or collapse the log window

Click to see the upload history

DATAVERSE INTEGRATION

Dataverse: Configure Project Upload Dialog

- Run Pasta Application
- Go to “Project group | Upload to Dataverse” or press F11
- Configure project upload dialog helps in saving the upload properties



Click on “Configure” button to open this window

Select the number of parallel uploads from the list.
2 to 5 parallel uploads possible and default value is 3.

Select the items to be zipped during the ELN generation.
When a new type is added via data hierarchy tool, the item will
be added here too, but will be de-selected by default.

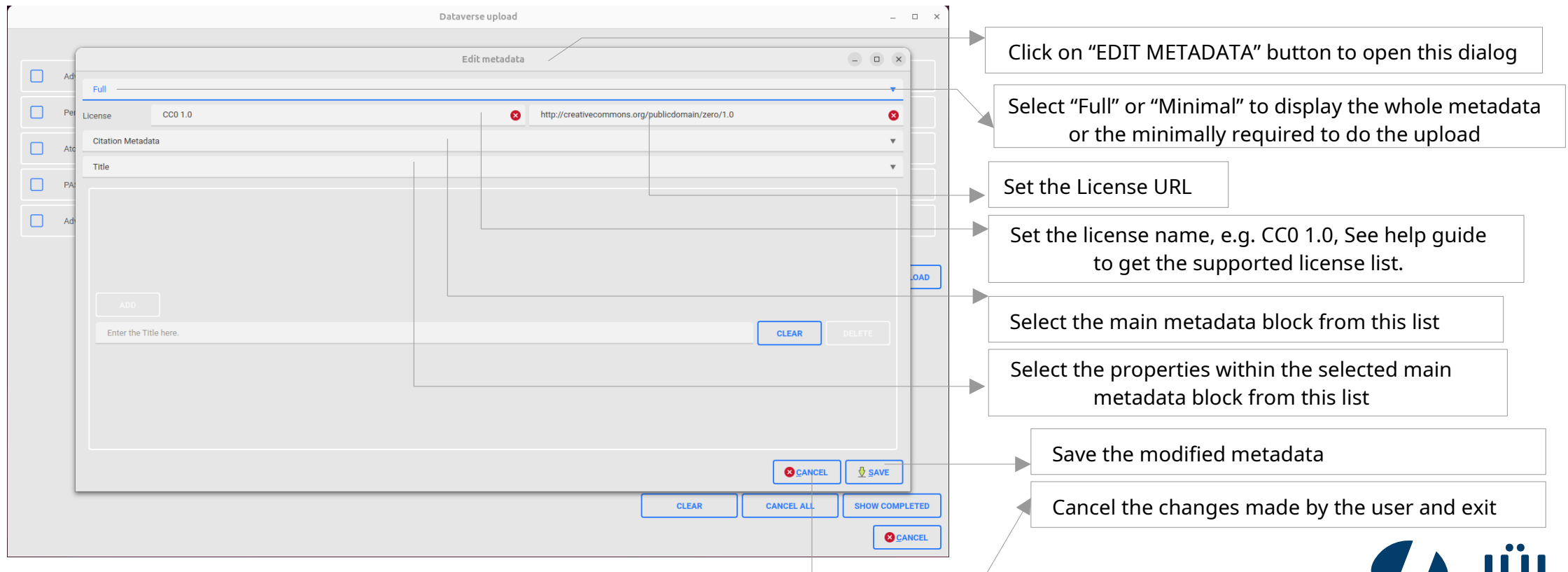
Save the upload configuration

Cancel the changes made by the user and exit

DATAVERSE INTEGRATION

Dataverse: Edit Metadata Dialog

- Run Pasta Application
- Go to “Project group | Upload to Dataverse” or press F11
- Edit metadata dialog helps in the modification of metadata needed to create the datasets.



The screenshot shows the 'Edit metadata' dialog box within the 'Dataverse upload' application. The dialog is titled 'Edit metadata' and contains several fields and buttons. A dropdown menu is set to 'Full', and the license is set to 'CC0 1.0' with a URL 'http://creativecommons.org/publicdomain/zero/1.0'. There are sections for 'Citation Metadata' and 'Title'. At the bottom, there are buttons for 'CANCEL', 'SAVE', 'CLEAR', 'CANCEL ALL', and 'SHOW COMPLETED'. Callout boxes provide instructions for each part of the dialog.

- Click on “EDIT METADATA” button to open this dialog
- Select “Full” or “Minimal” to display the whole metadata or the minimally required to do the upload
- Set the License URL
- Set the license name, e.g. CC0 1.0, See help guide to get the supported license list.
- Select the main metadata block from this list
- Select the properties within the selected main metadata block from this list
- Save the modified metadata
- Cancel the changes made by the user and exit

DATAVERSE INTEGRATION

Dataverse: Edit Metadata Dialog

- Run Pasta Application
- Go to “Project group | Upload to Dataverse” or press F11
- The first combo-box allows the user to see only the minimal or the full metadata list during edits.

Modified the selected property of the metadata block here

Add new entry for the Author

Delete the full entry to the left

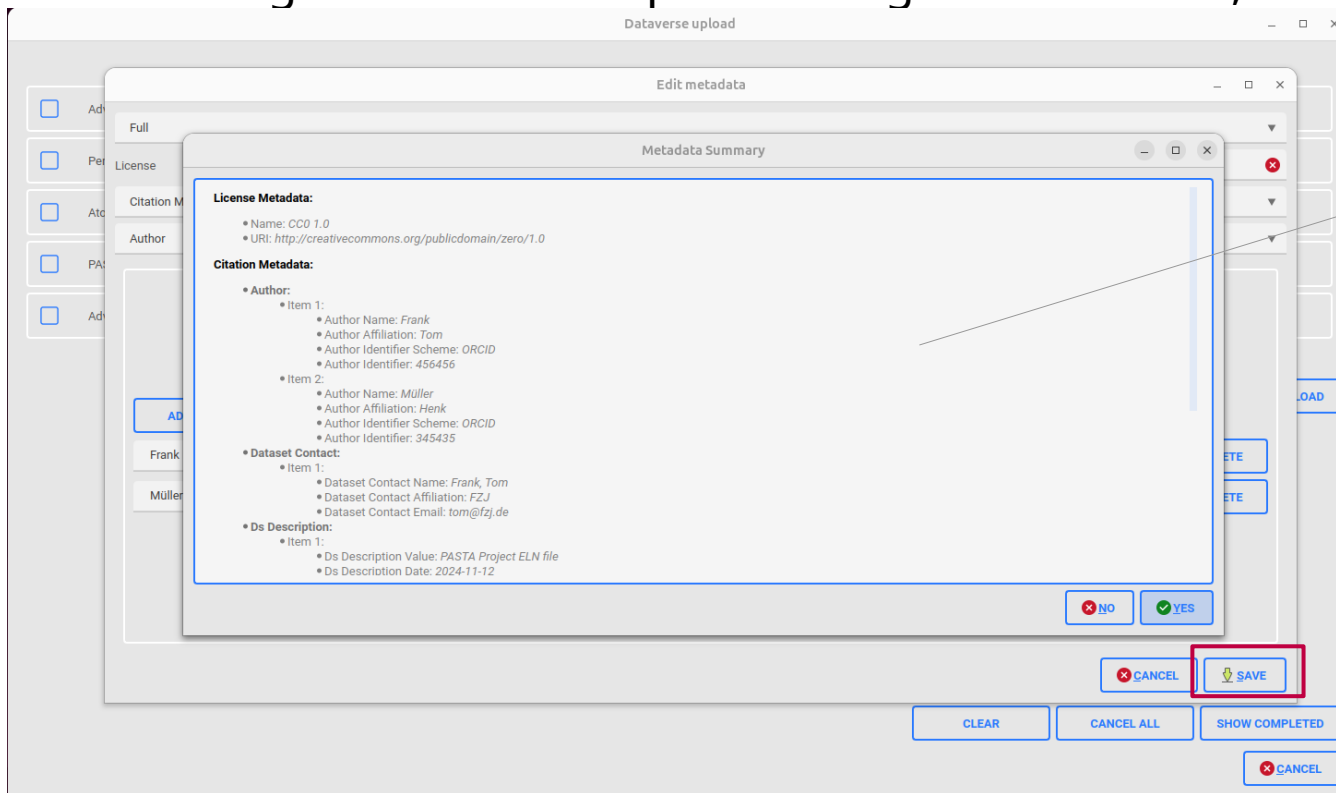
Delete the full entry to the left

Fields for the metadata property

DATAVERSE INTEGRATION

Dataverse: Edit Metadata Dialog

- Run Pasta Application
- Go to “Project group | Upload to Dataverse” or press F11
- After making the edits and upon clicking “Save” button, the summarized list is displayed to the user.



After the edits, once when the user clicks “SAVE” button the summarized metadata is displayed. User can accept the changes by clicking “YES” or go back to the edit page by clicking “NO”

DATAVERSE INTEGRATION

Dataverse: Upload History

- Run Pasta Application
- Go to “Project group | Upload to Dataverse” or press F11
- Click “SHOW COMPLETED” button to open Dataverse Upload History Dialog

The screenshot shows a 'Dataverse upload history' dialog box overlaid on a 'Dataverse upload' window. The dialog contains a table with the following data:

Project Name	Dataverse URL	Start Time	Finished Time	Status
[Redacted]	doi:10.5072/FK2/PTHYWD	2024-11-14 15:32:26	2024-11-14 15:32:30	Finished
[Redacted]	doi:10.5072/FK2/ZAQPB1	2024-11-14 15:32:26	2024-11-14 15:32:30	Finished
[Redacted]	doi:10.5072/FK2/O4W0R6	2024-11-14 15:32:24	2024-11-14 15:32:27	Finished
[Redacted]	doi:10.5072/FK2/TOVCDO	2024-11-14 15:32:23	2024-11-14 15:32:26	Finished
[Redacted]	doi:10.5072/FK2/916HZZ	2024-11-14 15:32:22	2024-11-14 15:32:25	Finished
PASTAs Example Project	doi:10.5072/FK2/BSNXQB	2024-11-14 15:32:20	2024-11-14 15:32:23	Finished
[Redacted]	doi:10.5072/FK2/XNEC4G	2024-11-14 15:32:19	2024-11-14 15:32:23	Finished
[Redacted]	doi:10.5072/FK2/PKBJTJ	2024-11-14 15:32:19	2024-11-14 15:32:22	Finished
[Redacted]	doi:10.5072/FK2/B2HAHW	2024-11-14 15:32:16	2024-11-14 15:32:20	Finished

Buttons at the bottom of the dialog include 'CLEAR', 'CANCEL ALL', and 'SHOW COMPLETED' (highlighted with a red box). The background window shows a list of projects with 'START UPLOAD' buttons.

After the uploads, click “SHOW COMPLETED” button to open “Dataverse upload history” dialog which lists all the upload tasks done in the past in a descending order for the finished time.

Search bar for the finished uploads, filtering will be done instantaneously and the search term will be applied on all properties (project name, upload url, start/finished time and status)

Uploaded project name

Clickable Dataverse URL for the uploaded project

Upload start time

Upload finished time

Upload status (Finished / Error / Canceled)

DATAVERSE INTEGRATION

Dataverse Module in Github repository

- GUI module (~pasta_eln/GUI/dataverse)
 - ~pasta_eln/GUI/dataverse/main_dialog.py (Main UI Dialog)
 - ~pasta_eln/GUI/dataverse/config_dialog.py (Dataverse Configuration Dialog)
 - ~pasta_eln/GUI/dataverse/edit_metadata_dialog.py (Metadata editor Dialog)
 - ~pasta_eln/GUI/dataverse/upload_config_dialog.py (Upload Configuration Dialog)
- Core Module (~pasta_eln/dataverse)
 - ~pasta_eln/dataverse/database_api.py (Database abstraction layer)
 - ~pasta_eln/dataverse/data_upload_task.py (Contains upload & publish logic)
 - ~pasta_eln/dataverse/pasta_config_reader_factory.py (Configuration singleton)
 - ~pasta_eln/dataverse/progress_updater_thread.py (Thread updating progress for the upload)
 - ~pasta_eln/dataverse/upload_queue_manager.py (Queue manager for the upload tasks)

THANKS

JITHU MURUGAN