

# Developing a Zenodo Jupyter Lab Extension

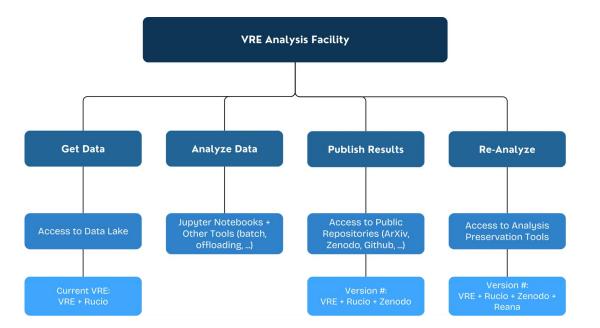
Michael Zengel

08/08/2024

### **The Virtual Research Environment (VRE)**

#### **Grand Purpose of VRE**

- An analysis facility based on the Jupyter framework
- Simplify life for physicists by aggregating software and infrastructure
- Developed by the ESCAPE collaboration
- Goal: End-to-end Scientific Analysis Workflow in a Cloud-Based Environment





### **This project: Context**

**Zenodo:** Open-source database software for sharing results/code

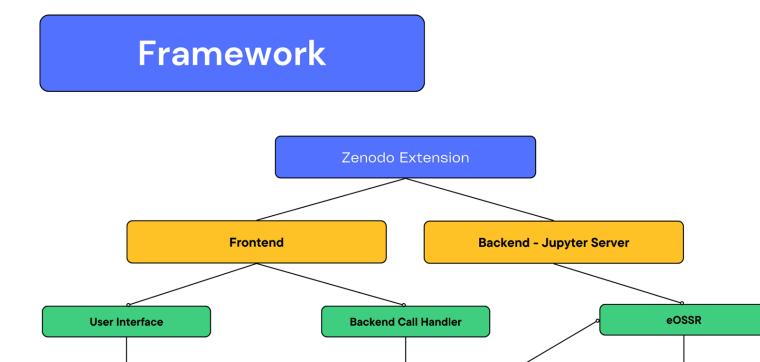
- Goal: Incorporate into VRE via Jupyter Lab, widely used interactive framework
  - Increases speed and ease of downloading and uploading data
  - Removes the need for local storage interaction; fully cloud-based
  - Exploits command line interface (CLI) in a visual way
  - Incorporates the "Publishing Results" step of the VRE end goal
    - https://github.com/vre-hub/zenodo-jupyterlab-extension







### **General Framework**



**API** endpoints

CERN

React Framework

Zenodo REST API

### **General Framework**

#### **Frontend Design**

- Built off of copier extension template
- Developed via Nodejs 20 and React
- Rendered as a Sidebar Widget
- Extends JupyterFrontEnd app

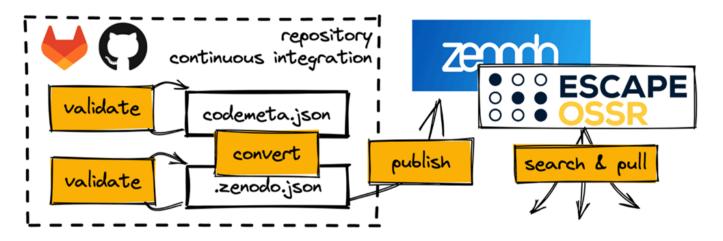
	jupyterlab / extension-template		Q Type [] to search	+ • O 11
<> Code	⊙ Issues 4 1 Pull requests 1 ⊙ Actions	① Security 1 🗠 Insights		
	extension-template (Public)		🛇 Sponsor 💿 Watch 9 🗸	♥ Fork 16 ▼ ☆ Star 49 ▼
	ٷ main → 양 3 Branches 🛇 25 Tags		Add file      Code	About
	👳 joaopalmeiro Update pyproject.toml.jinja (#85)		4ddbfc3 · last week 🕚 75 Commits	A `copier` template for JupyterLab extensions
	🧰 .github	Upgrade to copier 9.2.0 (#69)		template jupyterlab
	template	Update pyproject.toml.jinja (#85)	last week	jupyterlab-extension copier-template
	🗅 .gitignore			Readme     CC0-1.0 license
				Code of conduct
	C README.md	Upgrade to copier 9.2.0 (#69)		4 Security policy
	C copier.yml	Expand wording for has_settings to includ	e mention of s 3 weeks ago	<ul> <li>小 Activity</li> <li>E Custom properties</li> </ul>
	README      Sode of conduct	license 🎄 Security		☆ 49 stars ⓒ 9 watching 몇 16 forks
	JupyterLab extension	emplate		Report repository
	C CI passing			Releases 25



### **General Framework**

#### **Backend Design**

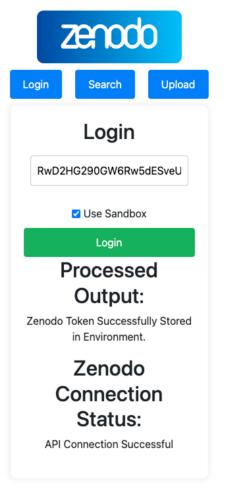
- Jupyter Server Extension (separate from Frontend Extension)
- Hosts API calls
- Runs <u>eOSSR</u> scripts for searching, logging in, and uploading data
  - eOSSR is a python library developed as a part of the ESCAPE Project



#### Logging in to Extension

- Takes in API Access Token
- Validates via deposit query status code
- Stores in env var for use throughout JupyterLab instance
  - Securely only accessible to user within session
- Sandbox functionality (stored for use in uploading)
  - Searching is exclusively non-sandbox

**Note:** This is simply a draft of the application; cosmetic details, such as spacing, will be addressed in the future.



#### Searching

- Uses built in Elasticsearch query string syntax from REST API
  - Searches for DOI, Title, Description, Creators, Communities
- Returns Title, Resource Type, Date Published
- Sorted by Most Recently changed (same as REST API)

zenodo			
Login	Search	Upload	
escape Records  Communities			
V Re	Search	unities	

Title	Resource Type	Date Published
Science Clusters: Position statement on operational commitment to EOSC and Open Research	Publication	2024-03-01
AfterSSHOC: synergies along the journey to EOSC and a view into the future	Presentation	2022-04-11
along the journey to EOSC and a view into	Presentation	2022-04-11



#### Searching

- Uses built in Elasticsearch query string syntax from REST API
- Returns Title, Resource Type, Date Published
- Sorted by Most Recently changed (same as REST API)
- Clicking a Record gives more information
  - Title with link to Zenodo record
  - Authors (with affiliations upon hover)
  - Download links on listed files (WIP)
    - Now on the PC
    - Future: \$HOME directory in Jupyter

```
escape
                                 Communities
                     Records
                               Search
Title
                        Resource Type
                                                Date Published
Science Clusters:
Position statement on
operational
                        Publication
                                                2024-03-01
commitment to EOSC
and Open Research
Title: Science Clusters: Position statement on operational commitment
to EOSC and Open Research
Authors:
                          ENVRI
    Petzold, Andreas; Hienola, Anca; Ewbank, Jonathan;
    Tedds, Jonathan; Lamanna, Giovanni; Bird, Ian; Gotz, Andrew;
    Bodera, Jordi; de Jong, Franciska; Wolff-Boenisch, Bonnie
Files:

    ScienceClusters_PSD3_010324.pdf
```



#### Searching

- Searchable Communities (same Elasticsearch query)
- Returns title and date published sorted by most recently changed

	zenod	0
Login	Search	Upload
esc	ape	
□ Records		
	Search	

Title	Date Published
ESCAPE	2022-12-05
ESCAPE OSSR	2019-06-27
ESCAPE-NET H2020	2019-05-24
Community for ESCAPE demos	2021-04-01
ESCAPE 2018 - Workshop on Energy Scale Calibration in Anti- neutrino Precision Experiments	2018-07-18

#### Searching

- Searchable Communities (same Elasticsearch ٠ query)
- Returns title and date published sorted by ۰ most recently changed
- When clicked: •
  - Allows for searching of records within that ٠ community
- Possible future goal: More advanced search settings

	zenod	
Logir	Search	Upload
	astro	
	Records 🗆 Comm	unities
	Search	
Showi	ng Results from "ESC	APE OSSR"
Title	Resource Type	Date Publish
cds-astro/aladin-lite: 3.4.5	Software	2024-07-22
Access and use of astronomy-related data from Python : a series of Jupyter	Software	2023-01-18



notebooks tutorials

×

ed

#### Uploading

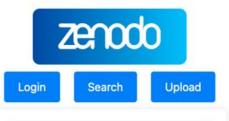
- Takes in basic required info
  - Files to upload (from \$HOME directory), Resource Type, Title, Creator
  - Optional: DOI (otherwise automatic), Description, Creator affiliation, multiple Creators

	.condarc	23 E
	.docker	7/31/2024 384 E
	.anyconnect	2/20/2022 623 E
	Select	
Digi	tal Object Identifier	
	Resource type *	
Select type		
	Title *	
	Description	
	Creators *	
Creator name	Affiliation	
	Add prostor	
	Add creator	
	Next	



#### Uploading

- Takes in basic required info
  - Files to upload (from \$HOME directory), Resource Type, Title, Creator
  - Optional: DOI (otherwise automatic), Description, Creator affiliation, multiple Creators
- Confirmation Page of Info to Upload
- "Confirm" does the following:
  - Creates deposit
  - Sets metadata (WIP)
  - Adds files to deposit (WIP)



### Confirmation

Title: Testing

Resource Type: software

DOI: (automatic)

Description: None given.

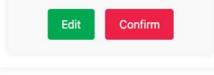
#### Creators:

• Michael (CERN)

#### Files:

/Users/michaelzengel/.condarc

Sandbox: No





#### Install

You will need NodeJS >= 20 for these steps.

Now, install yarn : npm install -g corepack corepack enable

Install the Python dependencies from within the main project directory: python -m pip install -r requirements.txt

Install Yarn Dependencies:

jlpm

Install and Build the Extension:

python -m pip install .

Enable the Extension: jupyter server extension enable zenodo\_jupyterlab.server

Now open a local instance of Jupyter Lab, and it should be present on the sidebar.

#### Docker

Rather than manually cloning the repository, it is possible to run the extension in a Docker container. To do this, use the following command:

docker run -d -p 8888:8888 ghcr.io/vre-hub/zenodo-jupyterlab-extension:<version>

All available versions can be found here

Now the instance of Jupyter Lab with the extension installed and enabled should be avilable on localhost:8888

### Usage

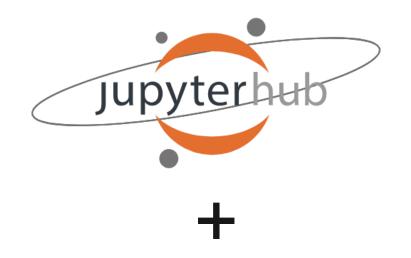
- Downloadable and Installable via git repository
- For Developers:
  - Easy installation in development mode (-ve after install command)
  - Simple building of front-end via jlpm
  - Up to date docker image available for download
    - Automatically downloads dependencies, software, installs, and runs Jupyter Lab with extension active
    - Can easily be added to Jupyter Hub distributions



•

### **Future Steps**

- Continued Development of software
  - Implementation of downloads to the Jupyter \$HOME directory
  - Ability to upload files to Zenodo Records
  - Advanced search settings
  - Improved cosmetic design
- Presentation of results at ADASS
  - Astronomical Data Analysis Software & Systems







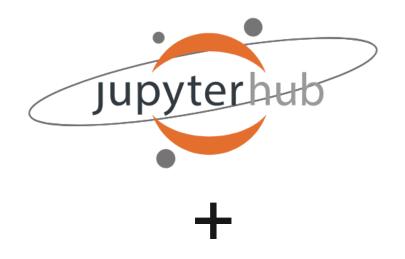
### **Conclusions**

#### The Software:

- Jupyter Lab Extension
- Provides Visual interface between users and the Zenodo Service
- Easily integrated into existing VRE

#### Why it's useful:

- Capability for **fully cloud-based** interaction (downloading and uploading) with Zenodo
- Adds another step into the VRE based analysis workflow
- Allows for more seamless downloading and uploading of results and software
- Applicable to any Jupyter-based environment







# Thank you! Any Questions?



### **ESCAPE** ESCAPE project & ESCAPE Data Lake



- Addressed RI's needs in Data Management, Access and Analysis for Astro-particle, Radio-astronomy, Gravitational Waves, Cosmology and Particle Physics.
- Provided a fully working common data infrastructure "The ESCAPE Data Lake" to test novel data management tools and models, giving the RI's the opportunity to influence and steer its development.
- Expanded **collaborations** and fostered involvement with other Scientific Communities. Maintained and strengthened collaborations with related EC initiatives and projects.
- ESCAPE finished Jan '23 and become an open collaboration [link] -> link





4

