Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

SWIRRL
Provenance-Aware Reproducible Workspaces

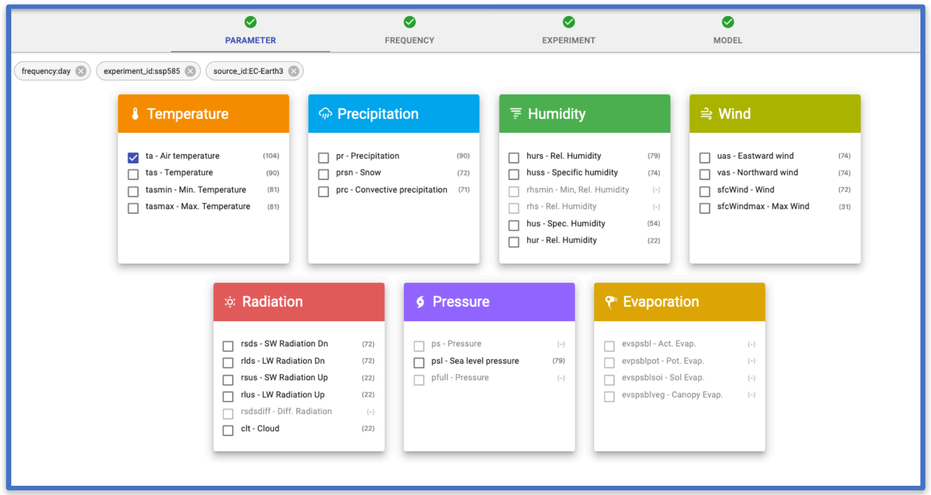
Alessandro Spinuso, Ian van der Neut
Hans Verhoef, Friedrich Striewski, Mats Veldhuizen

R&D Data Technology and Observations

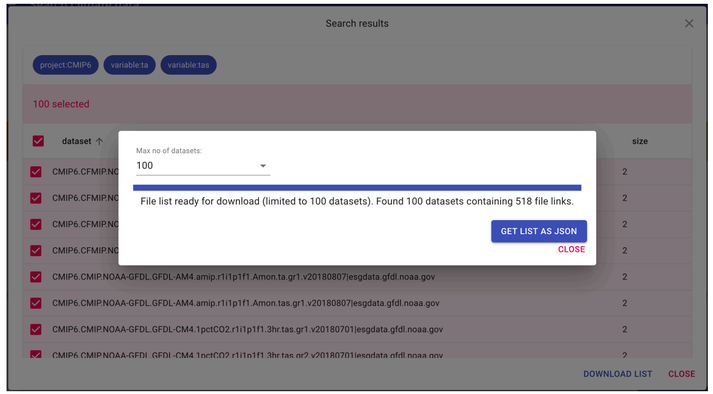
Interactive & reproducible Workspaces

Objective: Extend C4I with Data Driven & Reproducible Workspaces

Climate4Impact Search for CMIP5/6 Cordex Data (Distributed Data)



Incremental data staging/subsetting onto customisable and Reproducible Notebooks (extensible to other tools..)

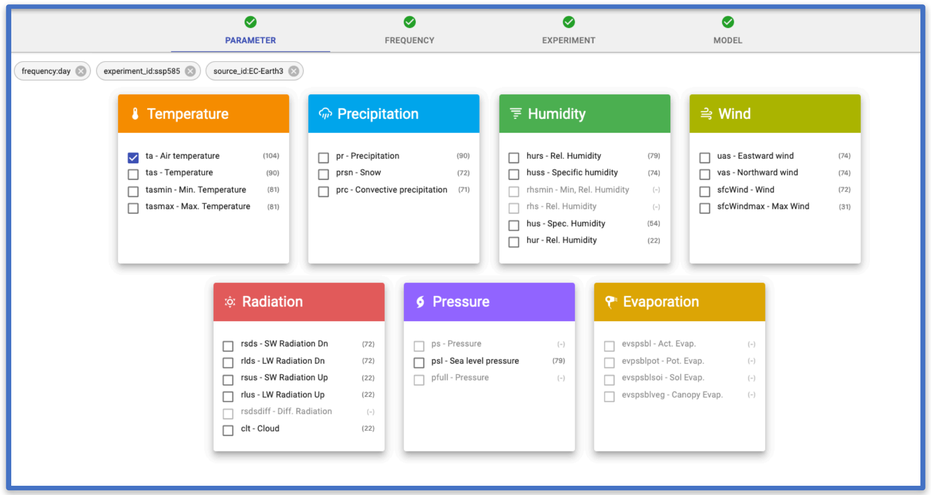


is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

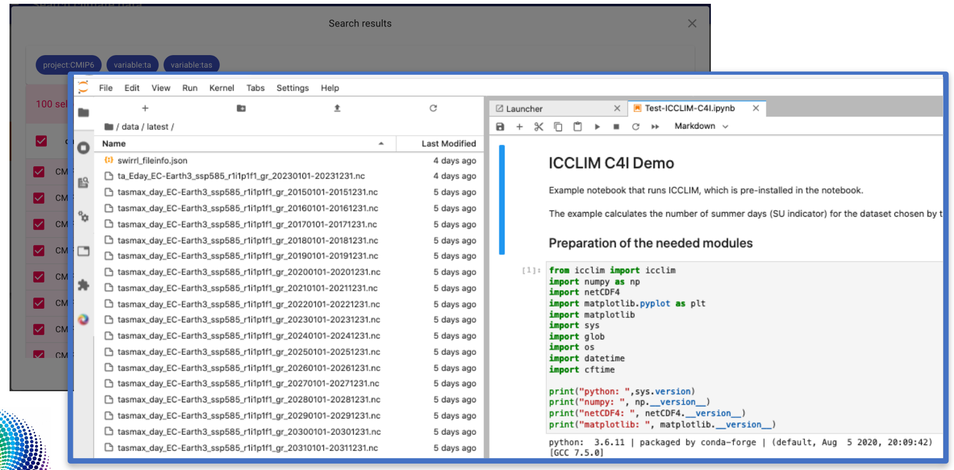
Interactive & reproducible Workspaces

Objective: Extend C4I with Data Driven & Reproducible Workspaces

Climate4Impact Search for CMIP5/6 Cordex Data (Distributed Data)



Incremental data staging/subsetting onto customisable and Reproducible Notebooks (extensible to other tools..)



SWIRRL-API

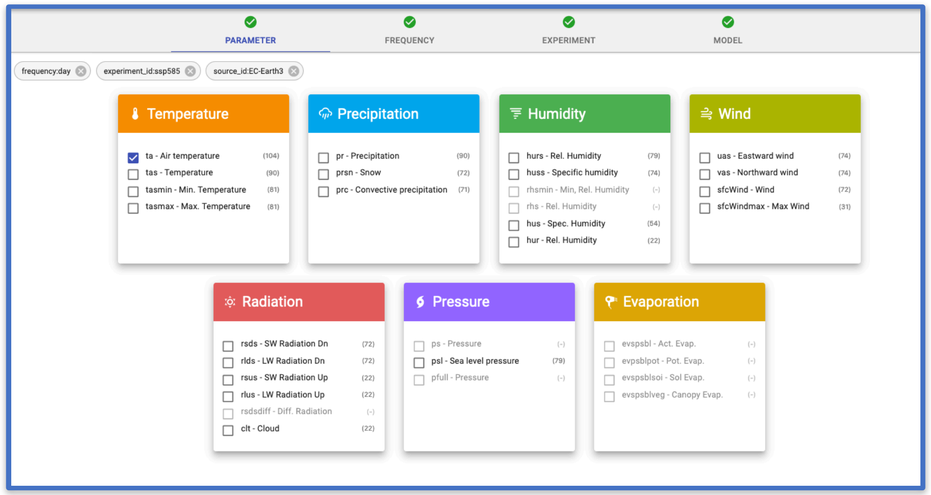
- Trace Changes to Software and Data
- Restore Environments

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

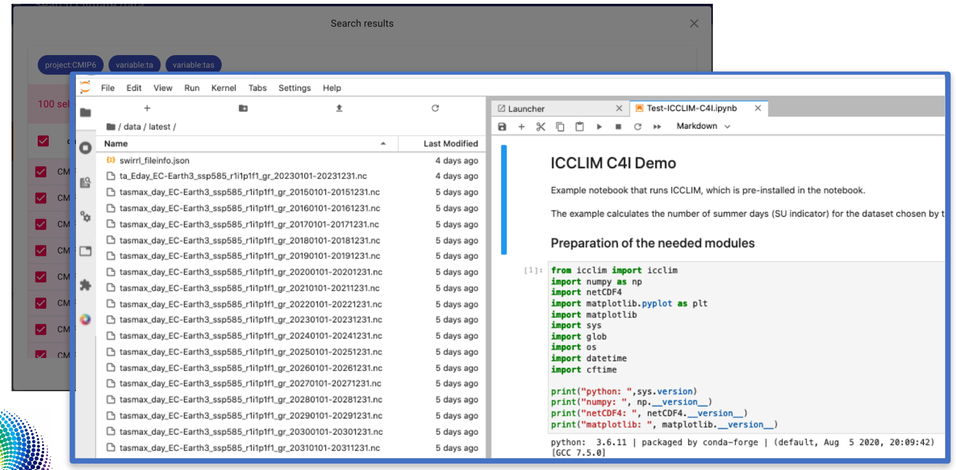
Interactive & reproducible Workspaces

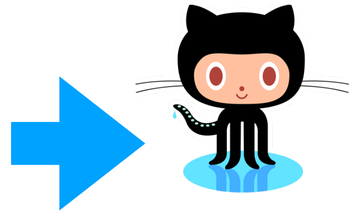
Objective: Extend C4I with Data Driven & Reproducible Workspaces

Climate4Impact Search for CMIP5/6 Cordex Data (Distributed Data)



Incremental data staging/subsetting onto customisable and Reproducible Notebooks (extensible to other tools..)





Software and Environment to Git

SWIRRL-API

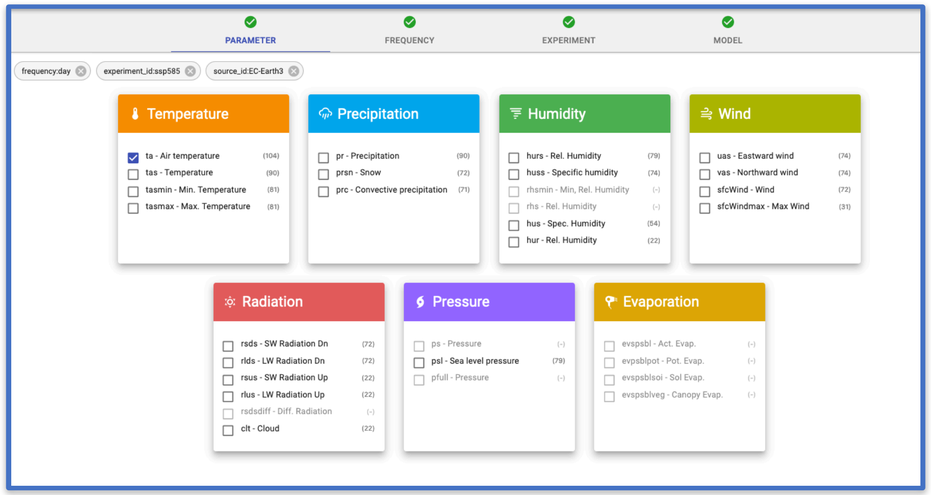
- Trace Changes to Software and Data
- Restore Environments

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

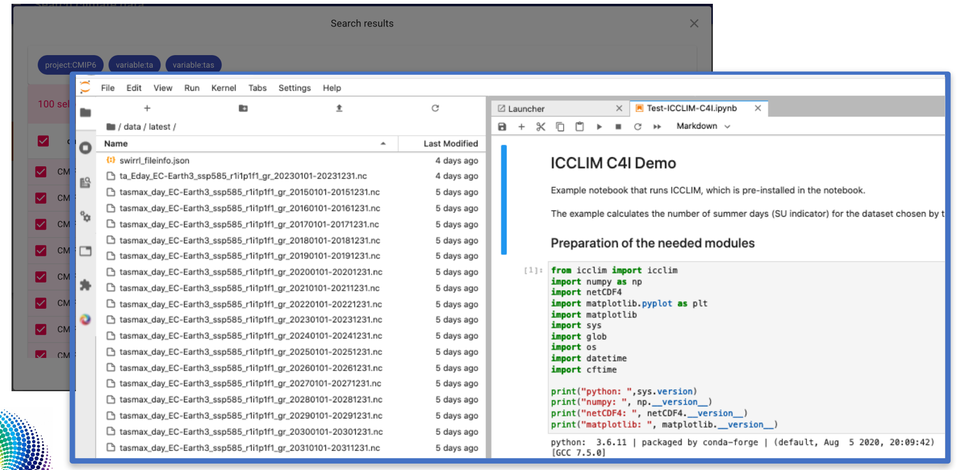
Interactive & reproducible Workspaces

Objective: Extend C4I with Data Driven & Reproducible Workspaces

Climate4Impact Search for CMIP5/6 Cordex Data (Distributed Data)



Incremental data staging/subsetting onto customisable and Reproducible Notebooks (extensible to other tools..)



SWIRRL-API

- Trace Changes to Software and Data
- Restore Environments

Software and Environment to Git

Data

MyBinder
Reproduce

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Provenance-aware Workspaces

SWIRRL-API

A Web API (high-level piece of infrastructure) to:

Manage **Working Sessions offering Notebook and Visualisation Services**

Run Workflows (CWL) for data staging and preprocessing onto the Working Session

Keep data staging history

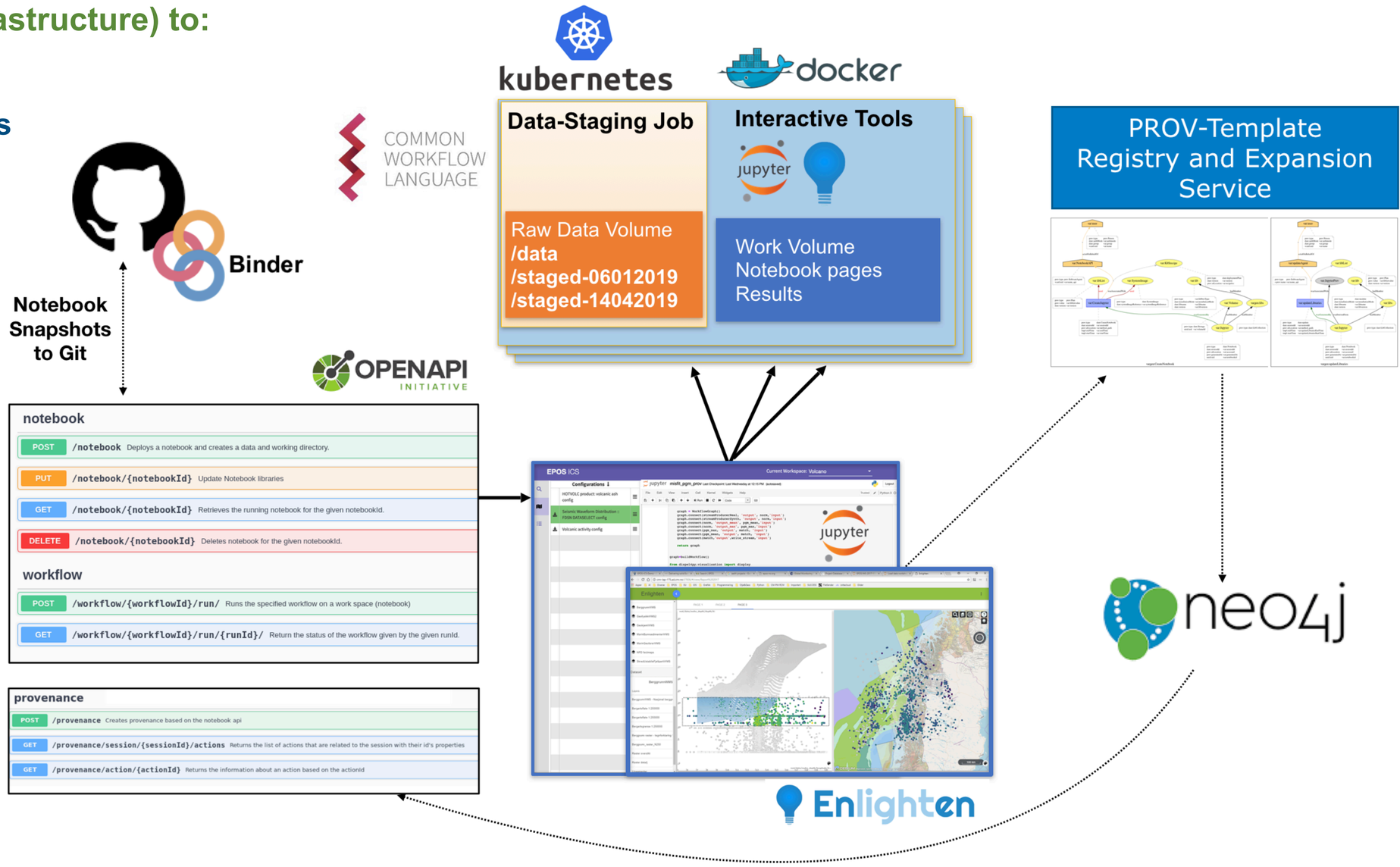
Provenace-aware

Restore SW Environments to to a state in the past

Ondemand Binder Snapshots to GitHub (Environment, methods, data references)

Jupyter Lab Extension!

https://gitlab.com/KNMI-OSS/swirrl/swirrl-api
https://zenodo.org/record/4264852#.X7ZeqNv_qNZ



SWIRRL-API

FAIR ENVRI

notebook

provide/manage a customisable notebook environment

POST	`/notebook`	Deploys a notebook and creates a data and working directory.
POST	`/notebook/{id}/snapshot`	Creates a notebook snapshot
PUT	`/notebook/{id}/restorelibs/{activityId}`	Restore Notebook libraries from a previous update
PUT	`/notebook/{id}`	Update Notebook libraries
GET	`/notebook/{id}`	Retrieves the running notebook for the given notebookId.
DELETE	`/notebook/{id}`	Deletes notebook for the given notebookId.

docker

SWIRRL-API

notebook

provide/manage a customisable notebook environment

POST	`/notebook`	Deploys a notebook and creates a data and working directory.
POST	`/notebook/{id}/snapshot`	Creates a notebook snapshot
PUT	`/notebook/{id}/restorelibs/{activityId}`	Restore Notebook libraries from a previous update
PUT	`/notebook/{id}`	Update Notebook libraries
GET	`/notebook/{id}`	Retrieves the running notebook for the given notebookId.
DELETE	`/notebook/{id}`	Deletes notebook for the given notebookId.

workflow

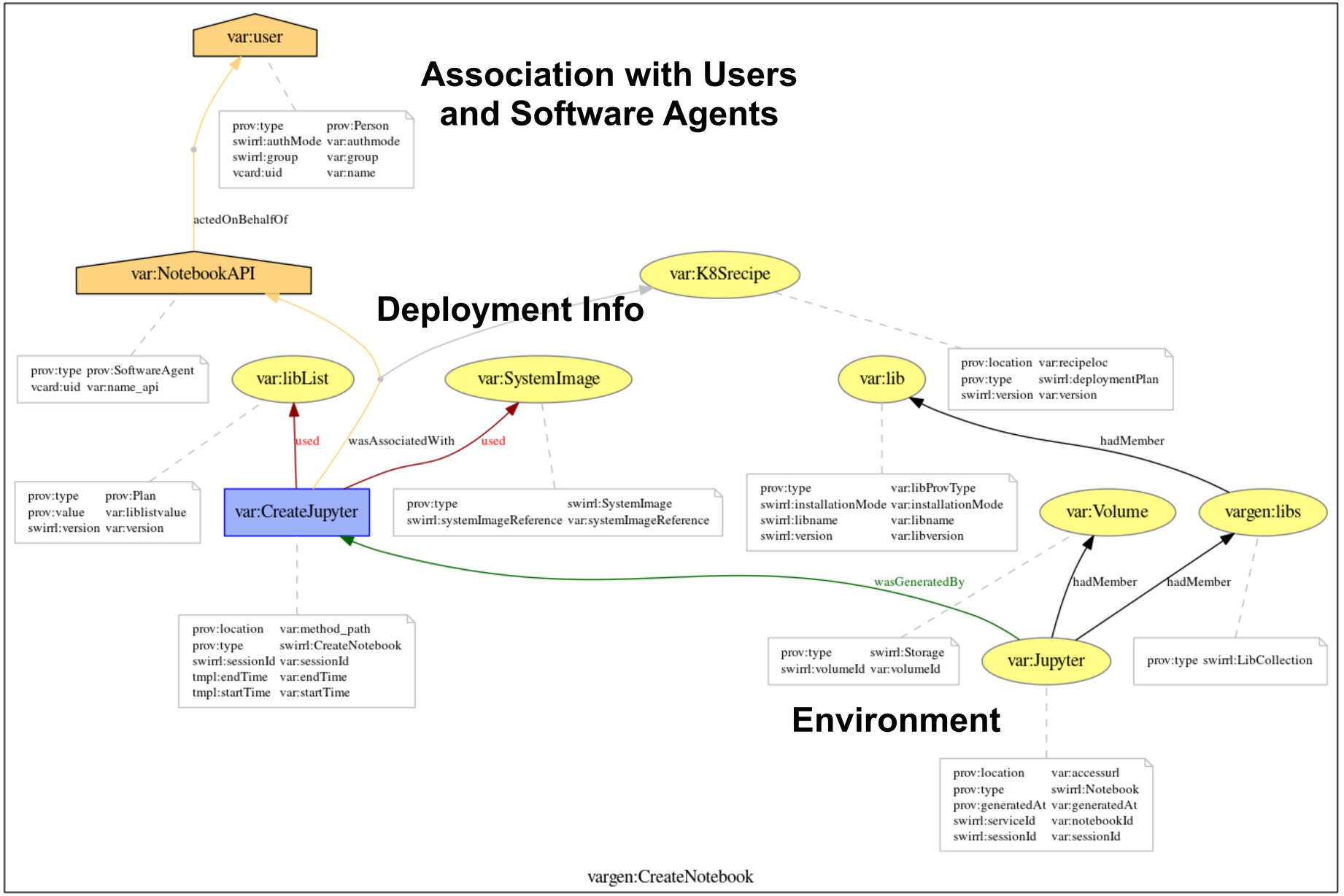
data staging / data preparation / processing

POST	`/workflow/{workflowId}/run/`	Runs the specified workflow on a work space (notebook)
GET	`/workflow/{workflowId}/run/{runId}/`	Return the status of the workflow given by the given runId.

COMMON WORKFLOW LANGUAGE

docker

“Create” a Notebook



“Update” a Notebook Environment

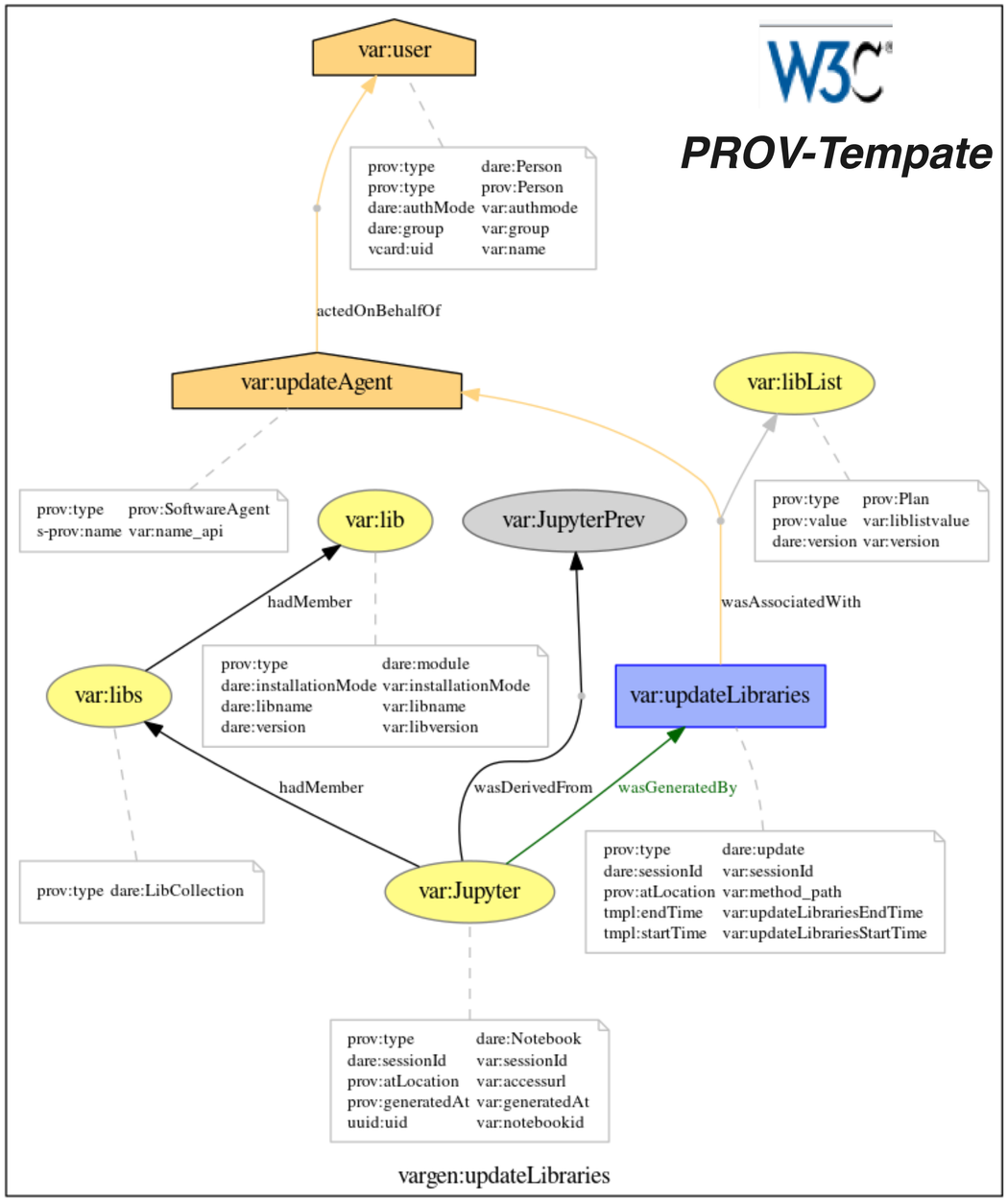
```
document
  prefix vargen <http://openprovenance.org/vargen#>
  prefix s-prov <http://s-prov/ns/#>
  prefix pre_0 <http://www.w3.org/2001/XMLSchema>
  prefix dare <http://project-dare.eu/ns#>
  prefix d-prov <http://d-prov.org/#>
  prefix dcterms <http://purl.org/dc/terms/>
  prefix vcard <http://www.w3.org/2006/vcard/ns#>
  prefix var <http://openprovenance.org/var#>
  prefix tmpl <http://openprovenance.org/tmpl#>
  prefix foaf <http://xmlns.com/foaf/0.1/>
  prefix uuid <urn:uuid:>

  bundle vargen:updateLibraries
    prefix vargen <http://openprovenance.org/vargen#>
    prefix s-prov <http://s-prov/ns/#>
    prefix dare <http://project-dare.eu/ns#>
    prefix vcard <http://www.w3.org/2006/vcard/ns#>
    prefix var <http://openprovenance.org/var#>
    prefix tmpl <http://openprovenance.org/tmpl#>
    prefix dcterms <http://purl.org/dc/terms/>
    prefix uuid <urn:uuid:>

    entity(var:Jupyter, [prov:generatedAt='var:generatedAt', uuid:uid='var:notebookid', prov:atLocat
    entity(var:lib, [dare:libname='var:libname', dare:installationMode='var:installationMode', dare:
    entity(var:libList, [dare:version='var:version', prov:type='prov:Plan', prov:value='var:liblistv
    entity(var:libs, [prov:type='dare:LibCollection'])
    wasDerivedFrom(var:Jupyter, var:JupyterPrev, -, -, -)
    wasAssociatedWith(var:updateLibraries, var:updateAgent, var:libList)
    activity(var:updateLibraries, -, -, [prov:atLocation='var:method_path', tmpl:startTime='var:upda
    actedOnBehalfOf(var:updateAgent, var:user, -)
    wasGeneratedBy(var:Jupyter, var:updateLibraries, -)
    agent(var:user, [vcard:uid='var:name', dare:authMode='var:authmode', dare:group='var:group', pro
    agent(var:updateAgent, [prov:type='prov:SoftwareAgent', s-prov:name='var:name_api'])
    hadMember(var:Jupyter, var:libs)
    hadMember(var:libs, var:lib)
  endBundle
endDocument
```

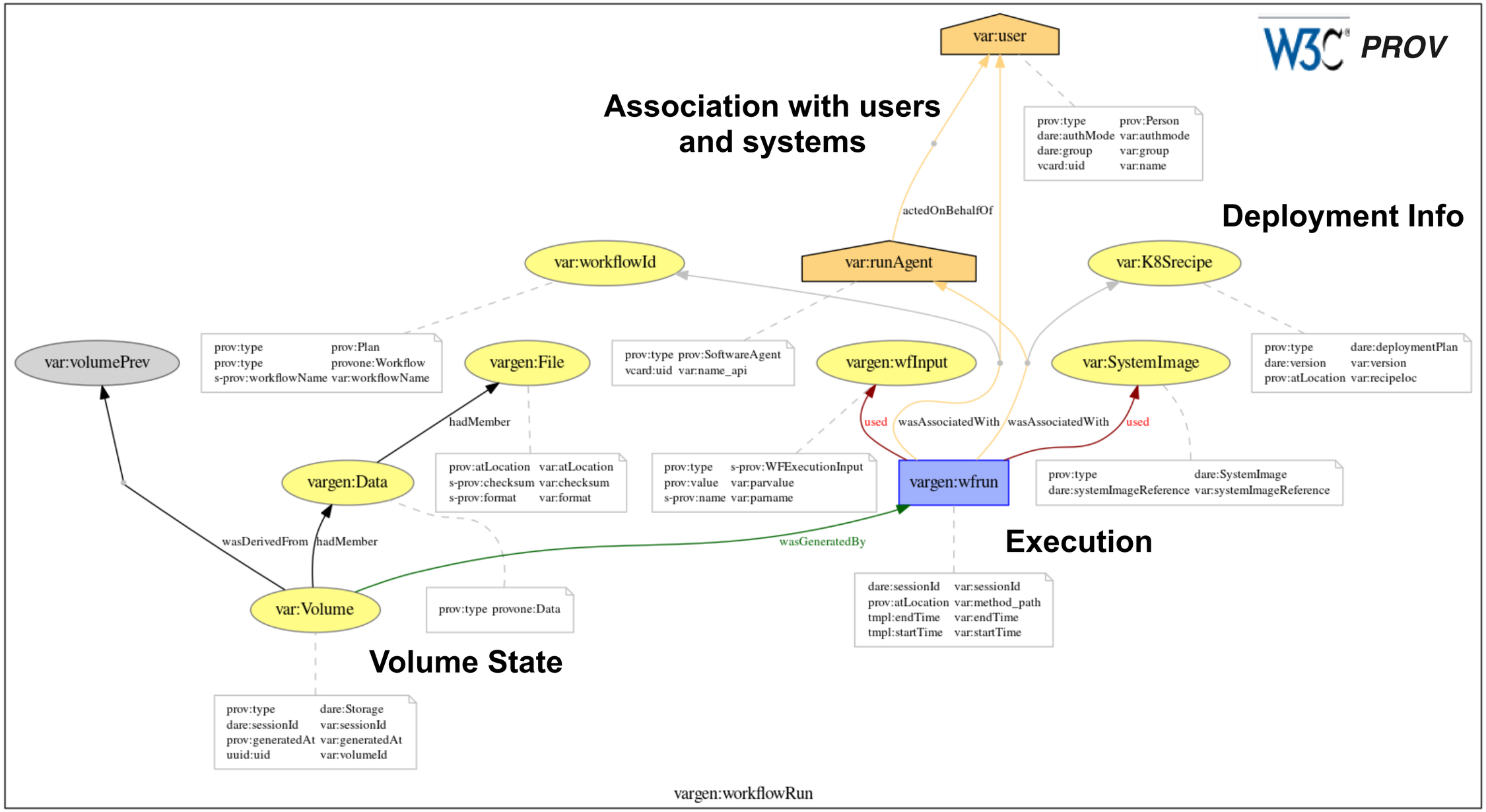
https://openprovenance.org/store/documents/1968

https://github.com/EnvriPlus-PROV/ProvTemplateCatalog

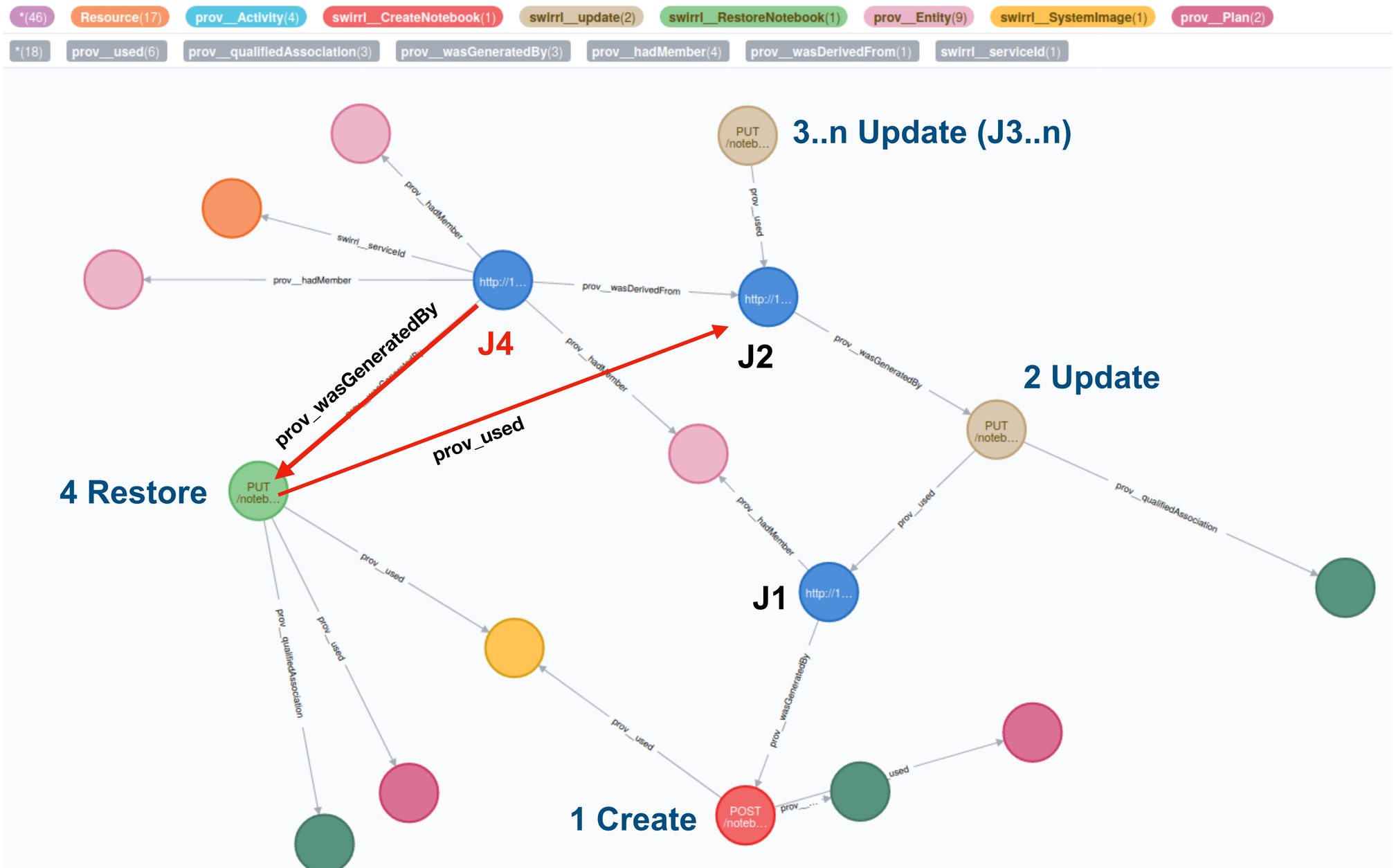


Luc Moreau et al. A Templating System to Generate Provenance
https://eprints.soton.ac.uk/405025/1/provtemplate.pdf

“Run” a Workflow/Container



Restoring Notebook Libraries (Prov Graph)



SWIRRL-API (Provenance)

Provenance Storage, Acquisition and Query

provenance

POST	`/provenance`	Saves provided provenance template expansion.
GET	`/provenance/session/{sessionId}/activities`	Returns the list of activities that are related to the session with their id's properties
GET	`/provenance/activity/{activityId}`	Returns the information about an activity based on the activityId

JSON-LD

SWIRRL-API (Provenance)

Provenance Storage, Acquisition and Query

provenance

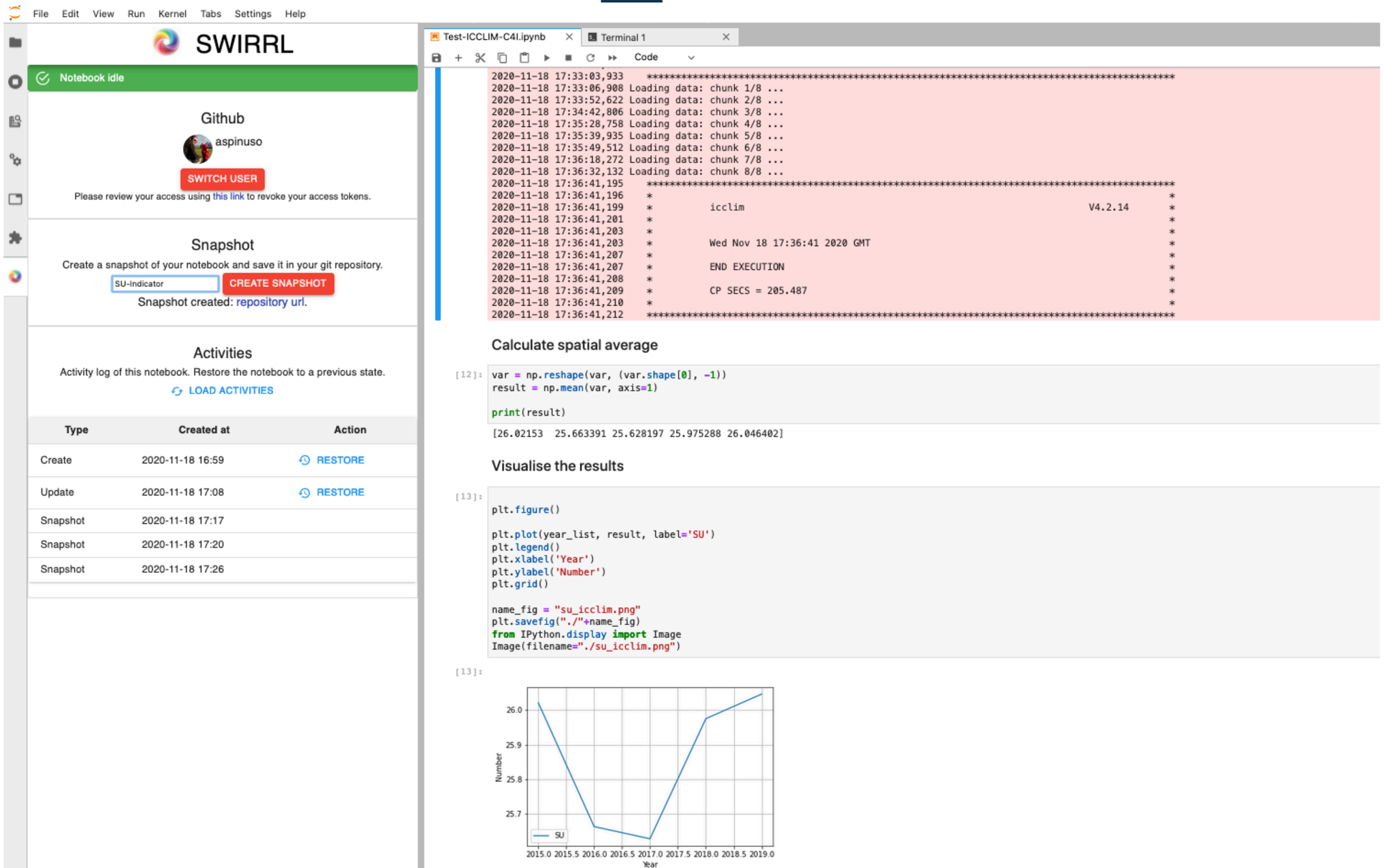
POST	/provenance	Saves provided provenance template expansion.
GET	/provenance/session/{sessionId}/activities	Returns the list of activities that are related to the session with their id's properties
GET	/provenance/activity/{activityId}	Returns the information about an activity based on the activityId

JSON-LD


```
{"@graph": [
   {"swirrl:serviceId": “15c0e8a0-19f4-43ff-a619-b99845c01280",
    "swirrl:sessionId": “a99f05e9-7570-46a3-b046-3c442ab1e23d",
    "@type": [
        "Resource",
        "prov:Activity",
        "swirrl:CreateNotebook"],
    “prov:person": "Mock User",
    "@id": "urn:uuid:375fe1aa-7957-4f19-81de-6549962fdbb4",
    "prov:endedAtTime": "2020-12-01T14:32:05.000Z",
    "prov:startedAtTime": "2020-12-01T14:31:21.337Z"},
   {"swirrl:serviceId": "15c0e8a0-19f4-43ff-a619-b99845c01280",
    "swirrl:sessionId": "a99f05e9-7570-46a3-b046-3c442ab1e23d",
    "@type": [
        "Resource",
        "prov:Activity",
        "swirrl:update"],
    “prov:person": "Mock User",
    "@id": "urn:uuid:daddea98-eb60-40d2-98e6-b2f026348da7",
    "prov:endedAtTime": "2020-12-01T14:32:46.137Z",
    "prov:startedAtTime": "2020-12-01T14:32:36.252Z"
],"@context": {..}}
```

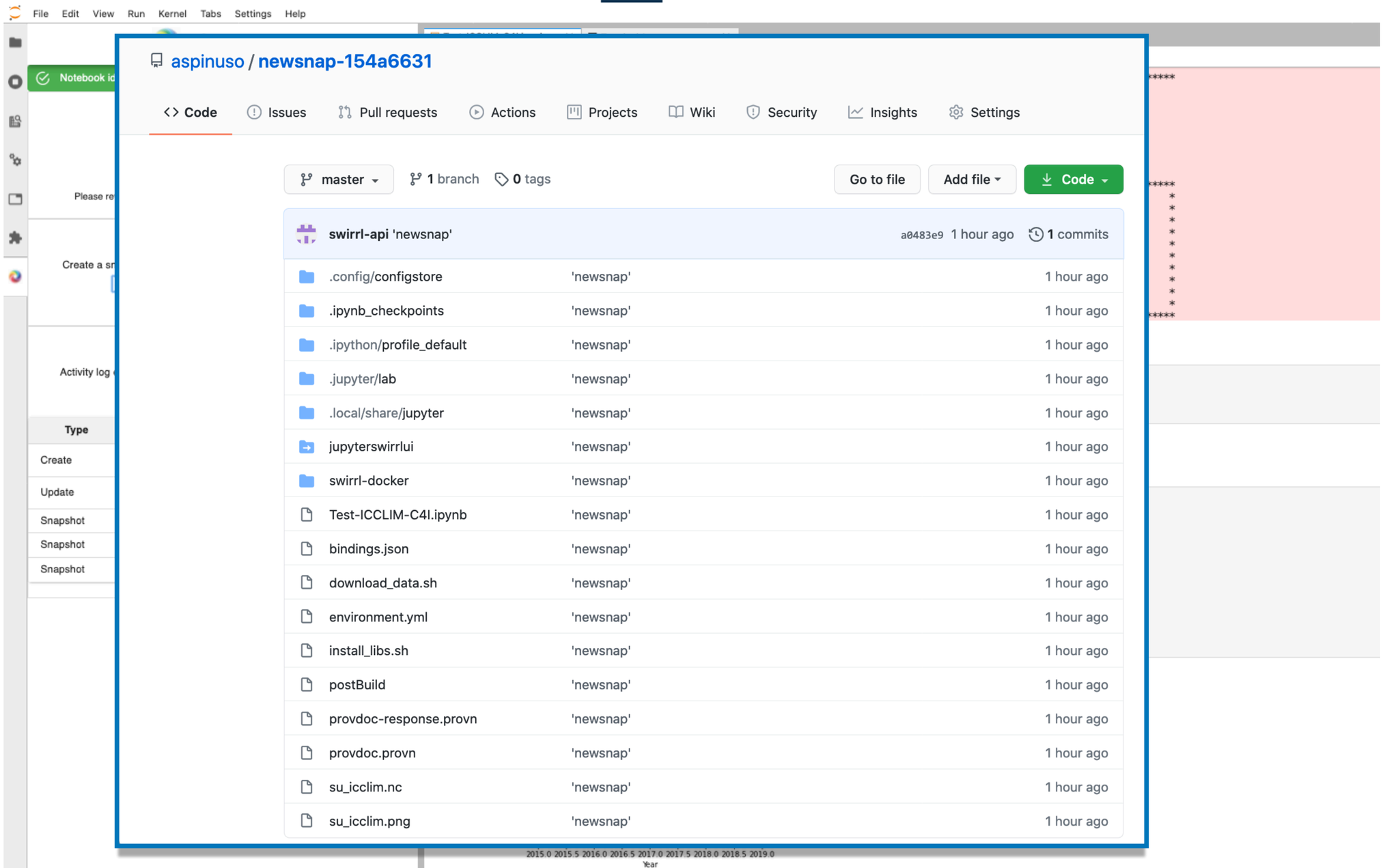
SWIRRL Jupyter Lab Extension

- *Monitor Jobs*
- *Snapshot Controls*
- *Trace Activities and trigger rollback actions*



SWIRRL Jupyter Lab Extension

- *Monitor Jobs*
- *Snapshot Controls*
- *Trace Activities and trigger rollback actions*



Other Resources and Demo

Binder Repository for Template Expansion Demo

Used for KNMI internal Communication/Tutorial on integrating provenance in python scripts for climate studies processes

https://gitlab.com/KNMI-OSS/swirrl/climate-scenarios-binder/-/blob/master/Template%20Expansion%20CS%20Single%20Calculation.ipynb