



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

SWIRRL & C4I Reproducible Notebooks in Climate4Impact

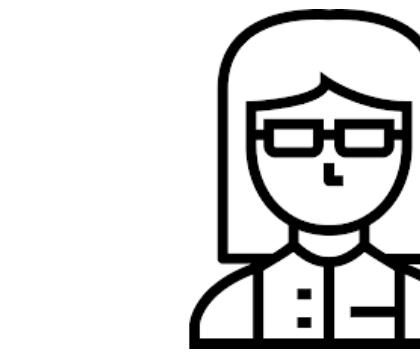
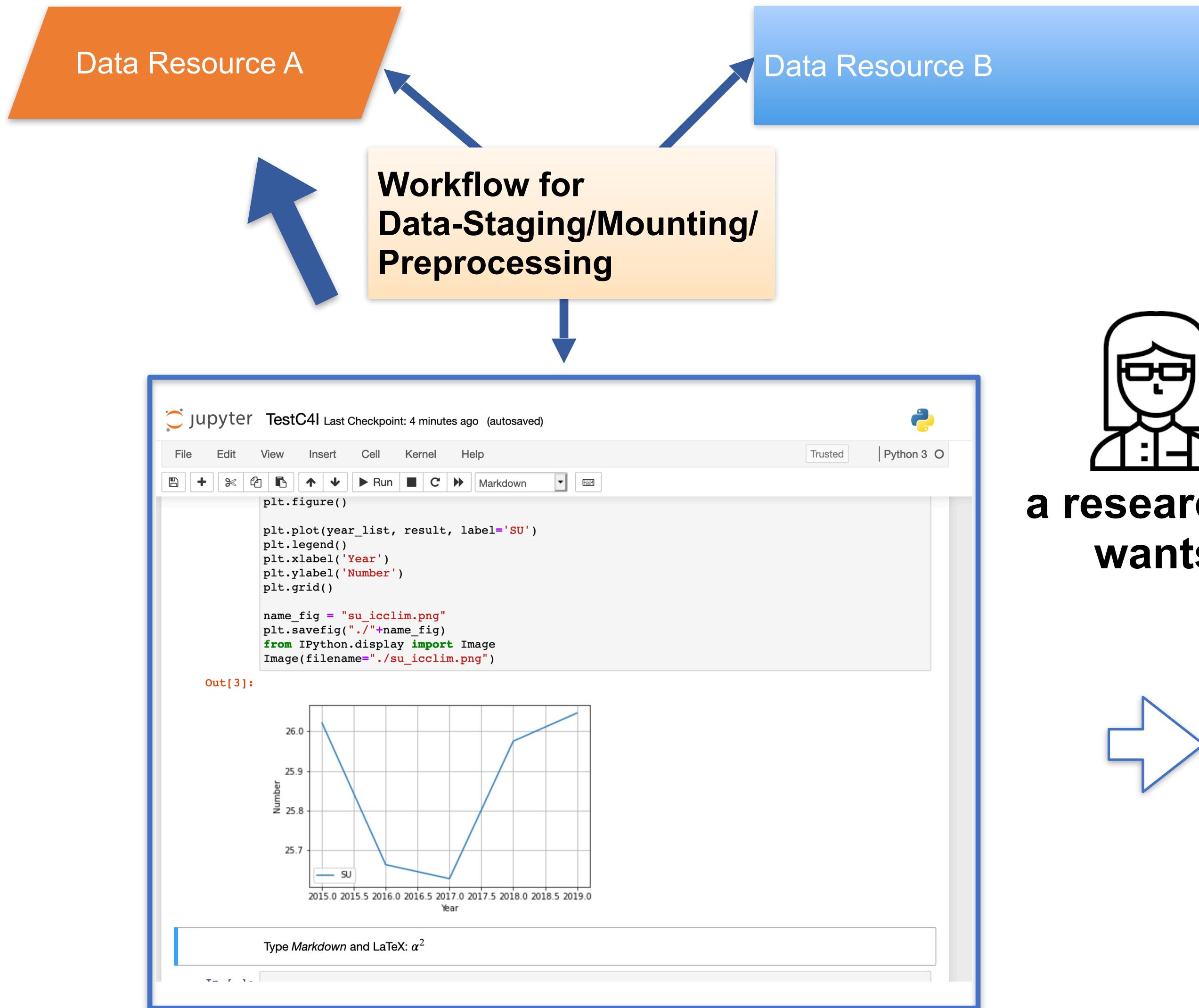
**Alessandro Spinuso, Ian van der Neut
Hans Verhoef, Friedrich Striewski, Mats Veldhuizen**

R&D Data Technology and Observations

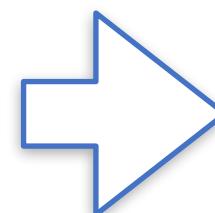


- **GUI usability** (Search, Selection, Subsetting)
- **Flexible analysis features** (Integration of Notebooks, ICCLIM and Workflows/Batch-processing)
- **Automated reproducibility mechanisms** (Datasets Version/GitHub/Binder/Provenance)
- **Making FAIRness of Data and Methods accessible**
- **New front-end technology**

Workspace Use Cases



a researcher
wants



- access distributed raw data
- develop, document and reuse methods for processing and visualisation.
- update/extend raw data and software
- Track changes and rollback
(Traceability/Recovery)
- keep old versions of the data after updates
(Reproducibility)
- snapshot and restore the state of a workspace software
(Reproducibility)

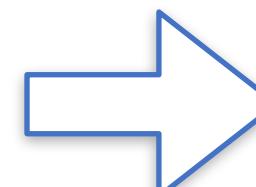
Interactive & reproducible Workspaces



Objective: Extend C4I with Data Driven & Reproducible Workspaces

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

Category	Variable	Count
Temperature	ta - Air temperature	(104)
	tas - Temperature	(90)
	tasmin - Min. Temperature	(81)
	tasmax - Max. Temperature	(81)
Precipitation	pr - Precipitation	(90)
	prsn - Snow	(72)
	prc - Convective precipitation	(71)
Humidity	hurs - Rel. Humidity	(79)
	huss - Specific humidity	(74)
	rhsmin - Min. Rel. Humidity	(-)
	rhs - Rel. Humidity	(-)
	hus - Spec. Humidity	(54)
	hur - Rel. Humidity	(22)
	Wind	uas - Eastward wind
vas - Northward wind		(74)
sfcWind - Wind		(72)
sfcWindmax - Max Wind		(31)
Radiation		rsds - SW Radiation Dn
	rlds - LW Radiation Dn	(72)
	rsus - SW Radiation Up	(22)
	rlus - LW Radiation Up	(22)
	rsdldiff - Diff. Radiation	(-)
	clt - Cloud	(22)
Pressure	ps - Pressure	(-)
	psl - Sea level pressure	(79)
	pfull - Pressure	(-)
Evaporation	evpsbl - Act. Evap.	(-)
	evpsblpot - Pot. Evap.	(-)
	evpsblsoi - Sol Evap.	(-)
	evpsblveg - Canopy Evap.	(-)



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

Search results

project:CMIP6 variable:ta variable:tas

100 selected

Max no of datasets: 100

File list ready for download (limited to 100 datasets). Found 100 datasets containing 518 file links.

size

GET LIST AS JSON CLOSE

DOWNLOAD LIST CLOSE

dataset	size
CMIP6.CFMI.PNC	2
CMIP6.CMIP.NOAA-GFDL_GFDL-AM4.amip.r1i1p1f1.Amon.ta.gr1.v20180807/esgdata.gfdl.noaa.gov	2
CMIP6.CMIP.NOAA-GFDL_GFDL-AM4.amip.r1i1p1f1.Amon.tas.gr1.v20180807/esgdata.gfdl.noaa.gov	2
CMIP6.CMIP.NOAA-GFDL_GFDL-CM4.1pcrC02.r1i1p1f1.3hr.tas.gr1.v20180701/esgdata.gfdl.noaa.gov	2
CMIP6.CMIP.NOAA-GFDL_GFDL-CM4.1pcrC02.r1i1p1f1.3hr.tas.gr2.v20180701/esgdata.gfdl.noaa.gov	2

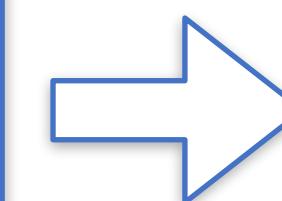
Interactive & reproducible Workspaces



Objective: Extend C4I with Data Driven & Reproducible Workspaces

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

The screenshot shows a search interface for climate data. It has four main tabs: PARAMETER, FREQUENCY, EXPERIMENT, and MODEL. The PARAMETER tab is active, displaying categories like Temperature, Precipitation, Humidity, Wind, Radiation, Pressure, and Evaporation, each with a list of specific variables and counts.



SWIRRL-API

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

The screenshot shows the SWIRRL API environment. On the left is a file browser with a tree view of datasets. On the right is a Jupyter Notebook titled "Test-ICCLIM-C4I.ipynb". The notebook contains code for an ICCLIM demo, including imports for icclim, numpy, netCDF4, and matplotlib, and prints for system and software versions.

```
from icclim import icclim
import numpy as np
import netCDF4
import matplotlib.pyplot as plt
import matplotlib
import sys
import glob
import os
import datetime
import cftime

print("python: ", sys.version)
print("numpy: ", np.__version__)
print("netCDF4: ", netCDF4.__version__)
print("matplotlib: ", matplotlib.__version__)

python: 3.6.11 | packaged by conda-forge | (default, Aug 5 2020, 20:09:42)
[GCC 7.5.0]
```

Trace Changes to Software and Data
Restore Environments

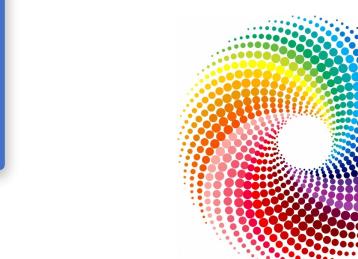
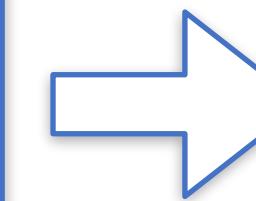
Interactive & reproducible Workspaces



Objective: Extend C4I with Data Driven & Reproducible Workspaces

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

The screenshot shows a search interface for Climate4Impact. It has four main categories: PARAMETER, FREQUENCY, EXPERIMENT, and MODEL. Under PARAMETER, there are sections for Temperature, Precipitation, Humidity, Wind, Radiation, Pressure, and Evaporation. Each section lists variables and their counts. For example, under Temperature, 'ta - Air temperature' has a count of 104.



SWIRRL-API

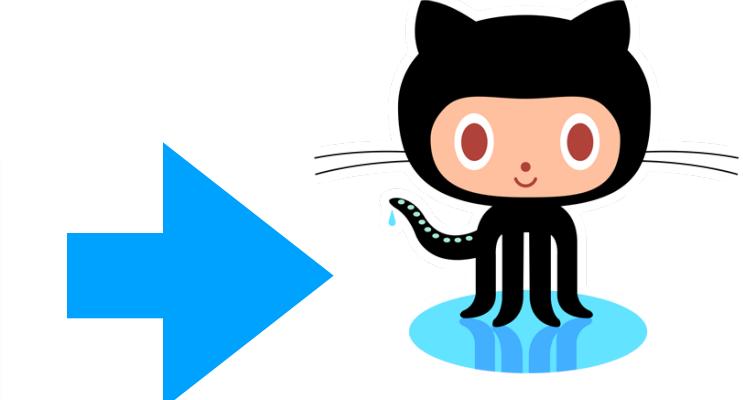
Incremental data staging/subsetting onto customisable
and Reproducible Notebooks (extensible to other tools..)

The screenshot shows a Jupyter Notebook environment. On the left is a file browser titled 'Search results' showing a list of netCDF files. On the right is a code cell containing Python code for an ICCLIM demo. The code imports icclim, numpy, netCDF4, and matplotlib, and prints system information.

```
from icclim import icclim
import numpy as np
import netCDF4
import matplotlib.pyplot as plt
import matplotlib
import sys
import glob
import os
import datetime
import cftime

print("python: ", sys.version)
print("numpy: ", np.__version__)
print("netCDF4: ", netCDF4.__version__)
print("matplotlib: ", matplotlib.__version__)

python: 3.6.11 | packaged by conda-forge | (default, Aug 5 2020, 20:09:42)
[GCC 7.5.0]
```



Software and
Environment
to Git

Trace Changes to Software and Data
Restore Environments

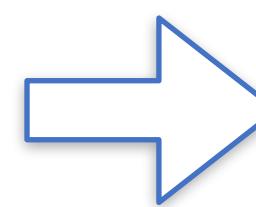
Interactive & reproducible Workspaces



Objective: Extend C4I with Data Driven & Reproducible Workspaces

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

The screenshot shows a search interface for Climate4Impact. It has four main sections: **PARAMETER**, **FREQUENCY**, **EXPERIMENT**, and **MODEL**. The **PARAMETER** section is expanded, showing categories like Temperature, Precipitation, Humidity, Wind, Radiation, Pressure, and Evaporation, each with a list of specific variables and counts.



SWIRRL-API

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

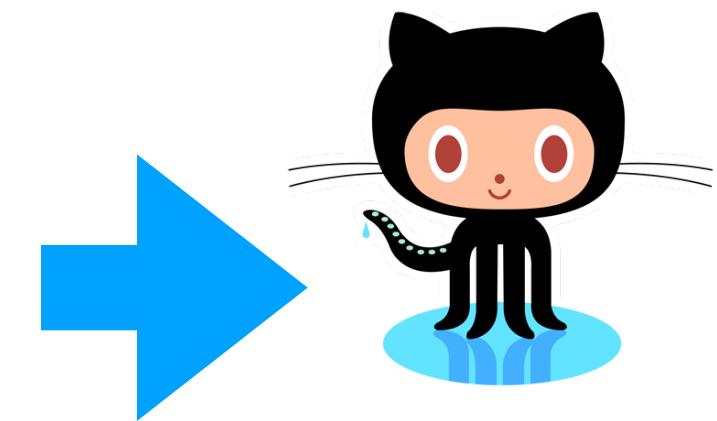
The screenshot shows a Jupyter Notebook interface. On the left is a file browser showing a directory structure with many netCDF files. On the right is a code cell containing Python code for setting up the ICCLIM environment, followed by environment variable prints.

```
from icclim import icclim
import numpy as np
import netCDF4
import matplotlib.pyplot as plt
import matplotlib
import sys
import glob
import os
import datetime
import cftime

print("python: ", sys.version)
print("numpy: ", np.__version__)
print("netCDF4: ", netCDF4.__version__)
print("matplotlib: ", matplotlib.__version__)

python: 3.6.11 | packaged by conda-forge | (default, Aug 5 2020, 20:09:42)
[GCC 7.5.0]
```

Trace Changes to Software and Data
Restore Environments



Software and
Environment
to Git



MyBinder
Reproduce

SWIRRL Jupyter Lab Extension



- Monitor Jobs

- Snapshot Controls

- Trace Activities and trigger rollback actions

The screenshot shows the SWIRRL extension integrated into a Jupyter Lab environment. The sidebar on the left contains sections for 'Notebook idle' (with a GitHub integration for user 'aspinuso'), 'Snapshot' (with a 'CREATE SNAPSHOT' button), and 'Activities' (listing recent operations like Create, Update, and multiple Snapshot events). The main area features a terminal window showing execution logs and a code cell displaying Python code for calculating spatial averages and visualizing results with a line plot.

Notebook idle

Github
aspinuso

SWITCH USER

Please review your access using [this link](#) to revoke your access tokens.

Snapshot

Create a snapshot of your notebook and save it in your git repository.

SU-Indicator CREATE SNAPSHOT

Snapshot created: [repository url](#).

Activities

Activity log of this notebook. Restore the notebook to a previous state.

LOAD ACTIVITIES

Type	Created at	Action
Create	2020-11-18 16:59	RESTORE
Update	2020-11-18 17:08	RESTORE
Snapshot	2020-11-18 17:17	
Snapshot	2020-11-18 17:20	
Snapshot	2020-11-18 17:26	

Terminal 1

```
Test-ICCLIM-C41.ipynb Terminal 1
```

```
2020-11-18 17:33:03,933 ****
2020-11-18 17:33:06,908 Loading data: chunk 1/8 ...
2020-11-18 17:33:52,622 Loading data: chunk 2/8 ...
2020-11-18 17:34:42,806 Loading data: chunk 3/8 ...
2020-11-18 17:35:28,758 Loading data: chunk 4/8 ...
2020-11-18 17:35:39,935 Loading data: chunk 5/8 ...
2020-11-18 17:35:49,512 Loading data: chunk 6/8 ...
2020-11-18 17:36:18,272 Loading data: chunk 7/8 ...
2020-11-18 17:36:32,132 Loading data: chunk 8/8 ...
2020-11-18 17:36:41,195 ****
2020-11-18 17:36:41,196 *
2020-11-18 17:36:41,199 * icclim V4.2.14 *
2020-11-18 17:36:41,201 *
2020-11-18 17:36:41,203 *
2020-11-18 17:36:41,203 * Wed Nov 18 17:36:41 2020 GMT
2020-11-18 17:36:41,207 *
2020-11-18 17:36:41,207 * END EXECUTION
2020-11-18 17:36:41,208 *
2020-11-18 17:36:41,209 * CP SECS = 205.487
2020-11-18 17:36:41,210 *
2020-11-18 17:36:41,212 ****
```

Code

```
Calculate spatial average
```

```
[12]: var = np.reshape(var, (var.shape[0], -1))
result = np.mean(var, axis=1)

print(result)
```

```
[26.02153 25.663391 25.628197 25.975288 26.046402]
```

```
Visualise the results
```

```
[13]: plt.figure()
plt.plot(year_list, result, label='SU')
plt.legend()
plt.xlabel('Year')
plt.ylabel('Number')
plt.grid()

name_fig = "su_icclim.png"
plt.savefig("./"+name_fig)
from IPython.display import Image
Image(filename="./su_icclim.png")
```

Figure: A line plot titled 'Visualise the results' showing the spatial average over time. The x-axis is labeled 'Year' and ranges from 2015.0 to 2019.0. The y-axis is labeled 'Number' and ranges from 25.7 to 26.0. The line shows a sharp drop from approximately 26.02 in 2015 to about 25.75 in 2016, followed by a steady increase back to around 25.98 by 2019. A legend entry 'SU' is present at the bottom left of the plot area.

SWIRRL Jupyter Lab Extension



- Monitor Jobs

- Snapshot Controls

- Trace Activities and trigger rollback actions

The screenshot displays the SWIRRL Jupyter Lab Extension integrated into a Jupyter Notebook environment. The main interface is a GitHub-style code repository viewer for the repository 'newsnap-154a6631'. The 'Code' tab is selected, showing a single commit from 'swirrl-api' made 1 hour ago with the commit hash 'a0483e9'. The commit message is 'newsnap'. The repository contains several files and folders, all of which are labeled with the tag 'newsnap'. The sidebar on the left shows a tree view of the notebook structure and a table of activity logs.

Type	Date
Create	2020-1
Update	2020-1
Snapshot	2020-1
Snapshot	2020-1
Snapshot	2020-1



Demo

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

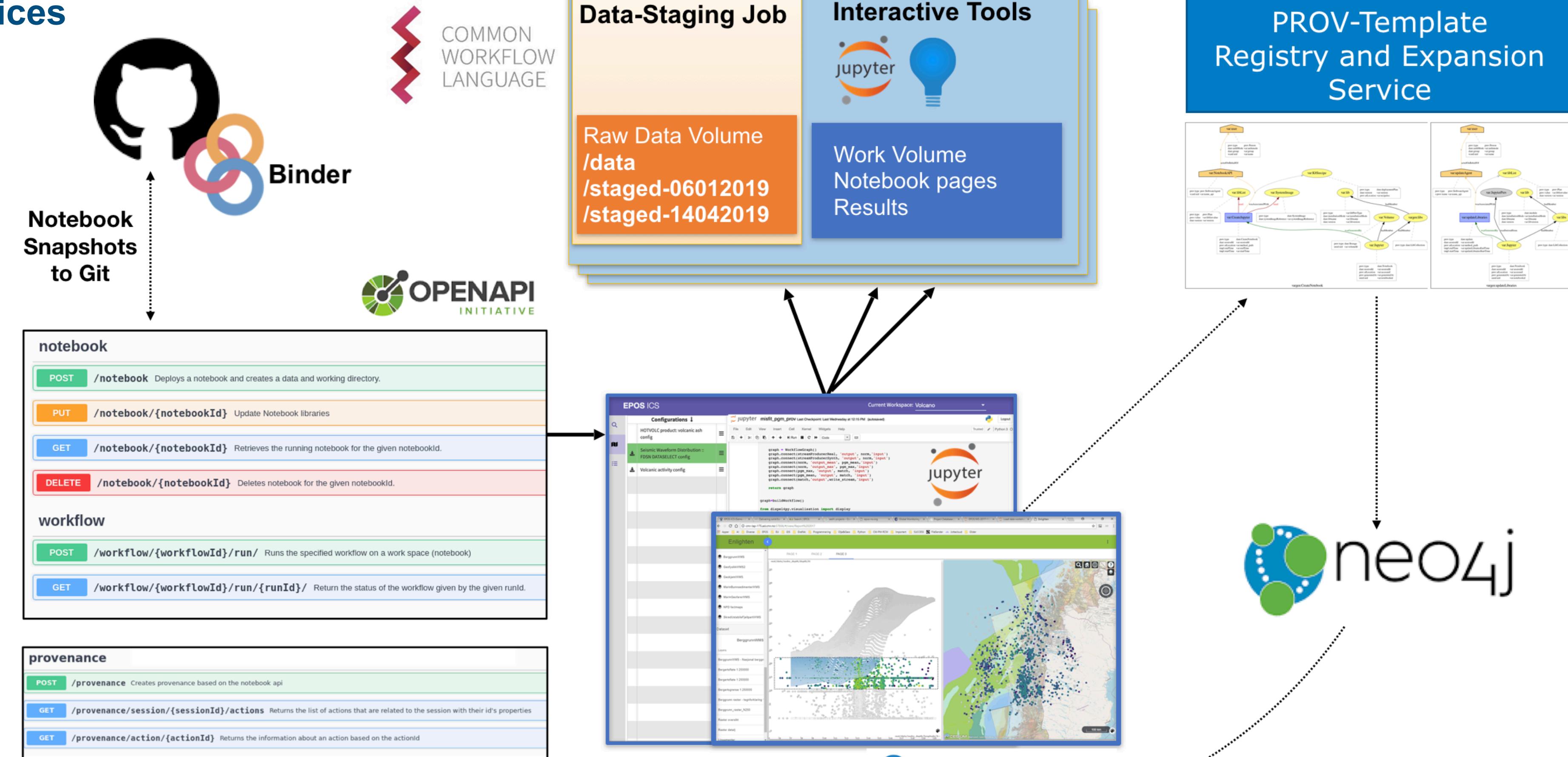
Restore SW Environments to
to a state in the past

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



“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/tmp#>
  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/tmp#>
  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

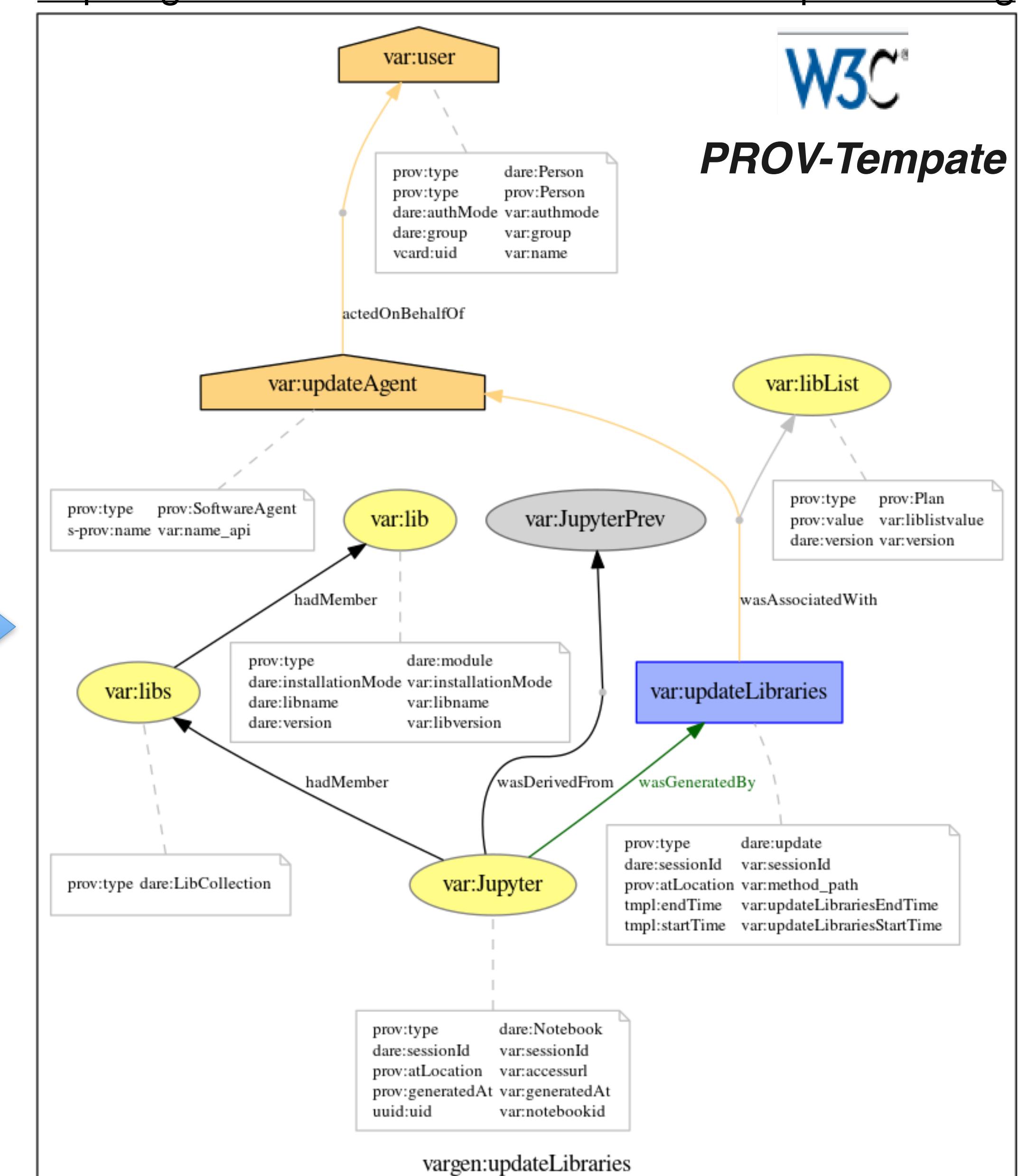
```

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

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



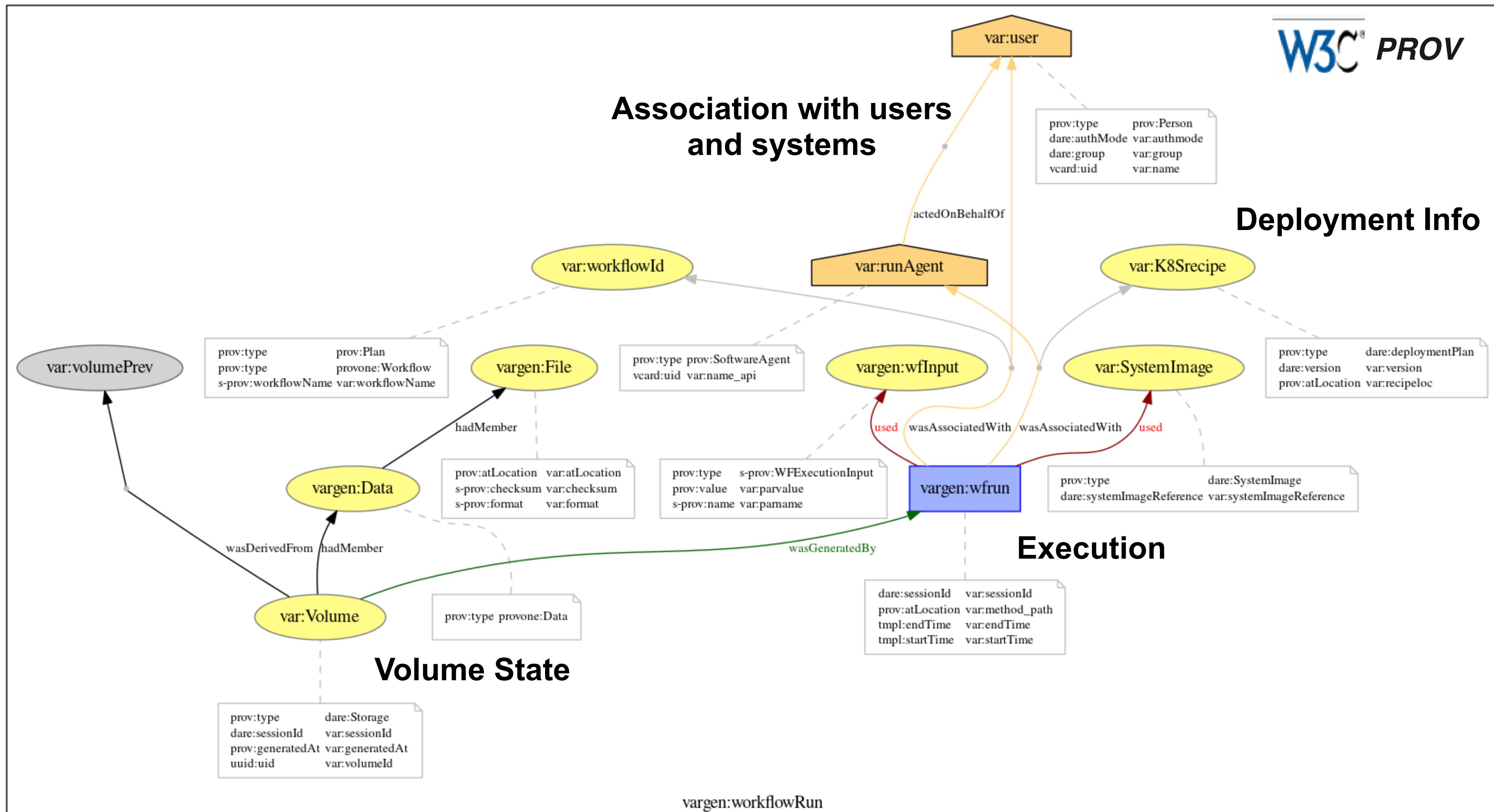
PROV-Tempate



“Run” a Workflow/Container



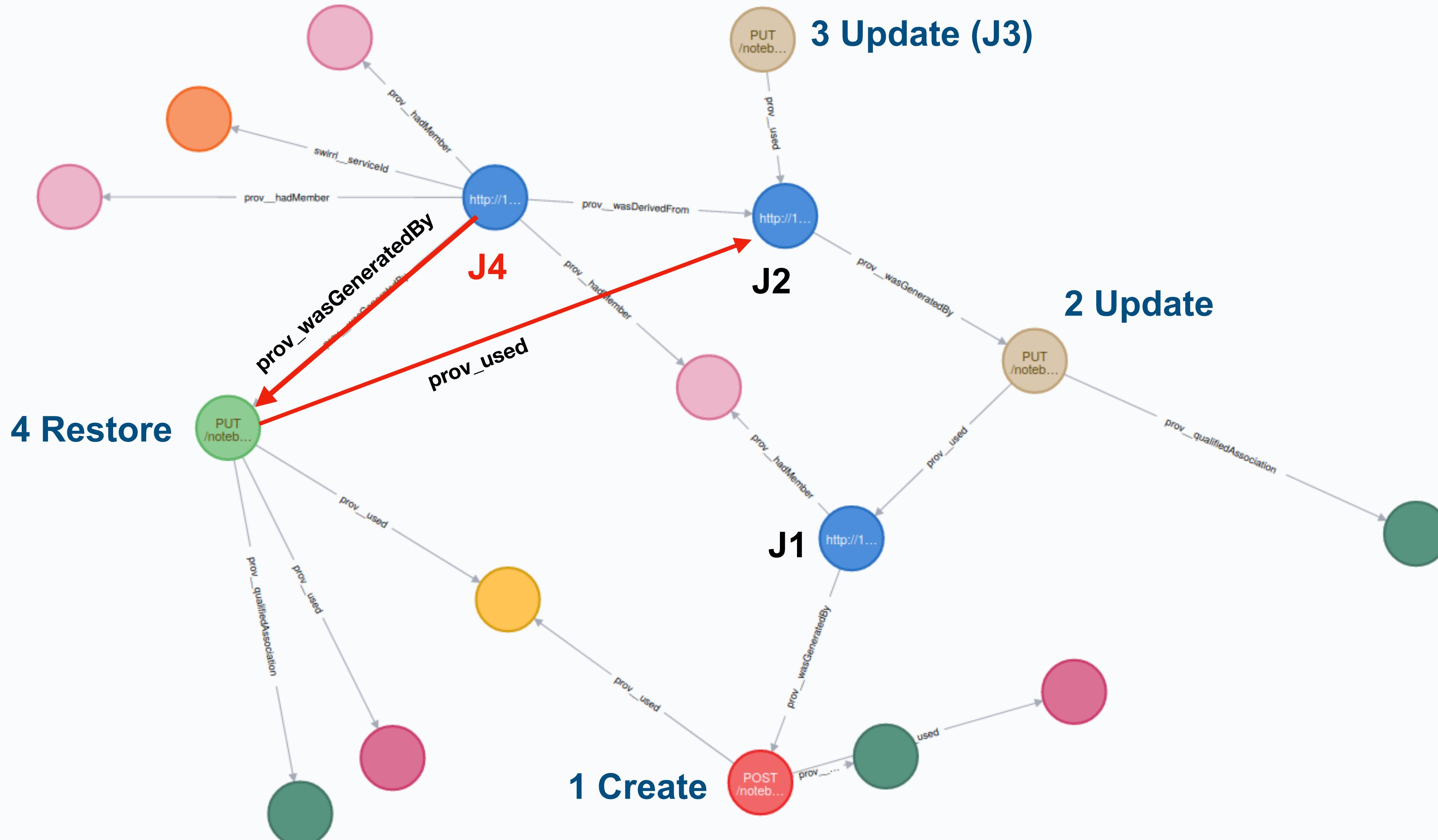
W3C PROV



Restoring Notebook Libraries (Prov Graph)



*(46) Resource(17) prov_Activity(4) swirrl_CreateNotebook(1) swirrl_update(2) swirrl_RestoreNotebook(1) prov_Entity(9) swirrl_SystemImage(1) prov_Plan(2)
*(18) prov_used(6) prov_qualifiedAssociation(3) prov_wasGeneratedBy(3) prov_hadMember(4) prov_wasDerivedFrom(1) swirrl_serviceId(1)



Ongoing and Future Work



- **Integration of OpenDAP data-staging workflows with remote subsetting**
- **Improved Deployment Performances**
- **Manage multiple Workspaces per User**
- **Integration of Datasets' metadata Pages (DOI)**
- **Selection of Model's Members**
- **Activities view and Restoring actions in Jupiter Lab**
- ***Alpha* testers (volunteers?)**
- **AAI C4I/SWIRRL**