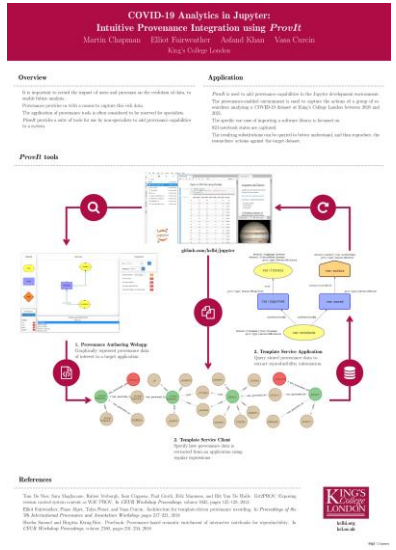


COVID-19 Analytics in Jupyter: Intuitive Provenance Integration using *ProvIt*

Martin Chapman, Elliot Fairweather, Asfand Khan and Vasa Curcin

King's College London

Provenance-enabled Jupyter environment: https://github.com/kclhi/jupyter

How do I properly specify the data I want to capture?

covid.ipynb

Code

[1]:
```
import pandas as pd
from pandas import DataFrame
from pandas import Series
import numpy as np
```

How do I properly specify the data I want to capture?

covid.ipynb

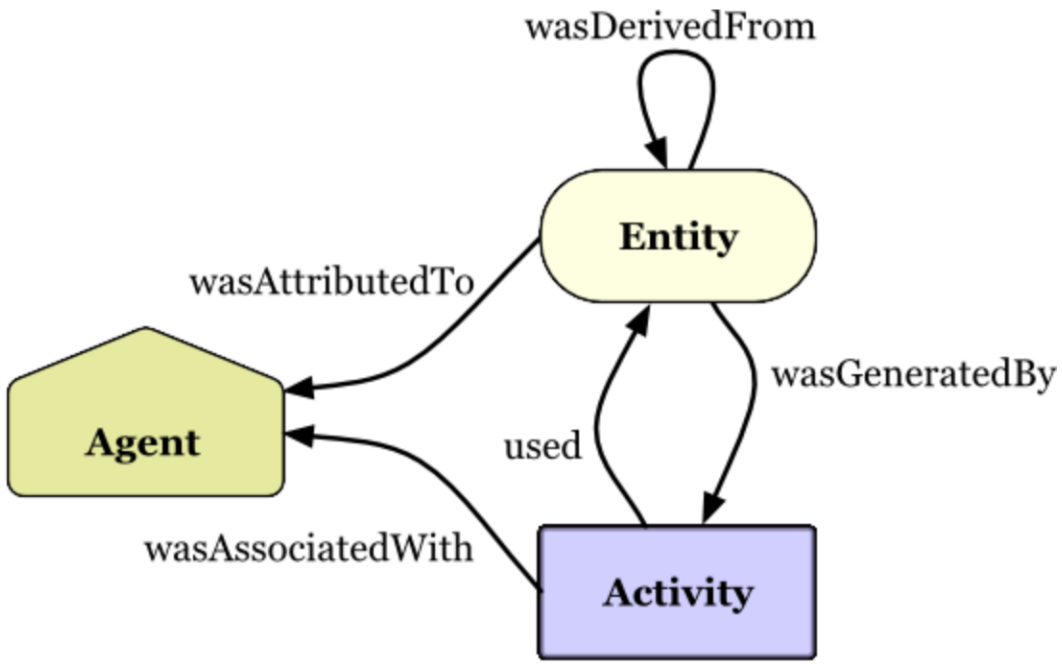
[1]:
```
import pandas as pd
from pandas import DataFr
from pandas import Series
import numpy as np
```

W3C Working Group Note

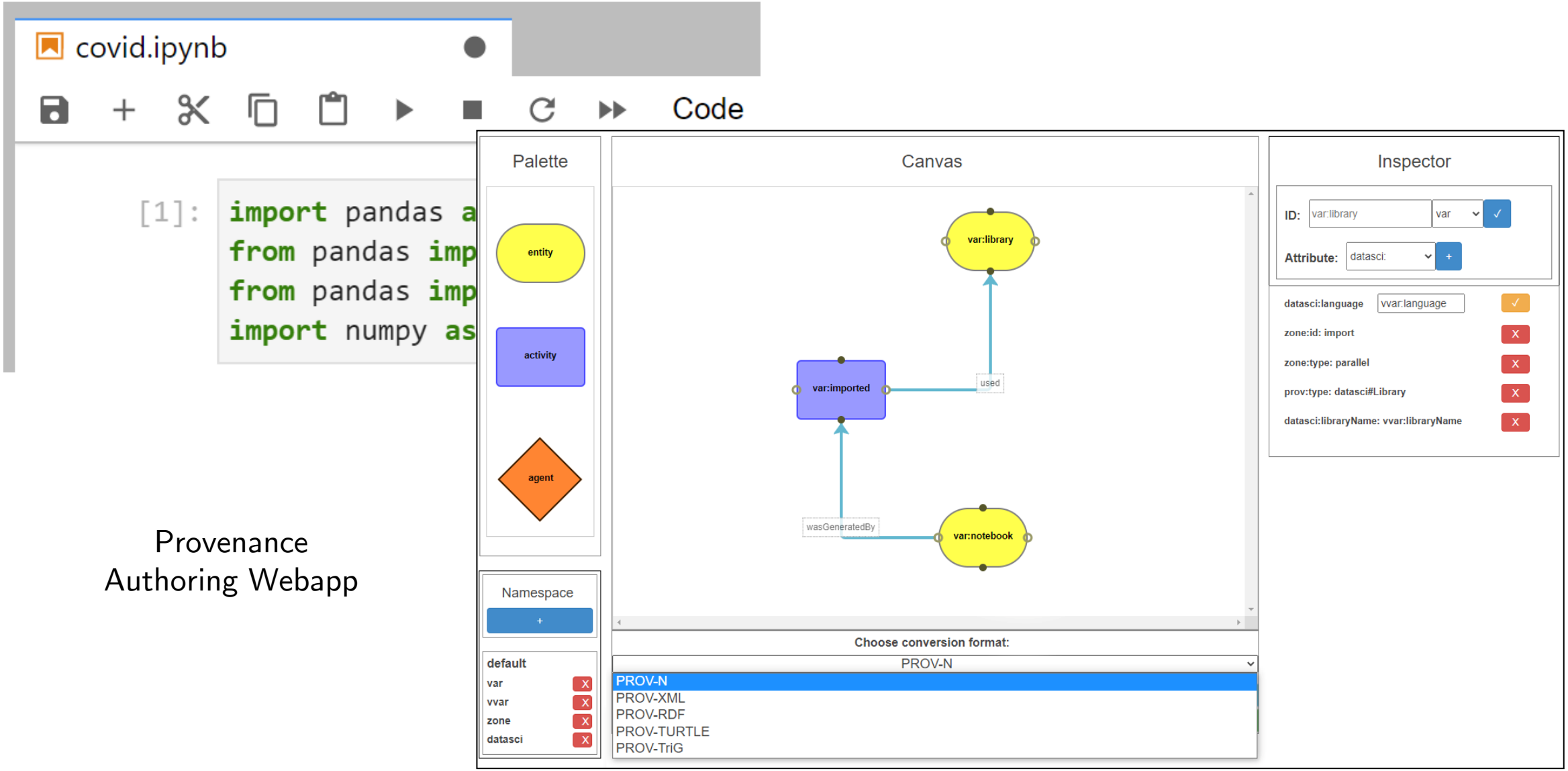
2. Intuitive overview of PROV

This section provides an explanation of the main concepts in PROV. As with the rest of this document, it should be treated as a starting point for understanding the model. The PROV data model document [PROV-DM] provides precise definitions and constraints [PROV-CONSTRAINTS] to be followed.

The following diagram provides a high level overview of the structure of PROV records, limited to some key PROV concepts discussed in this document. Note that because PROV is meant to describe how things were created or delivered, PROV relations are named so they can be used in assertions about the past.

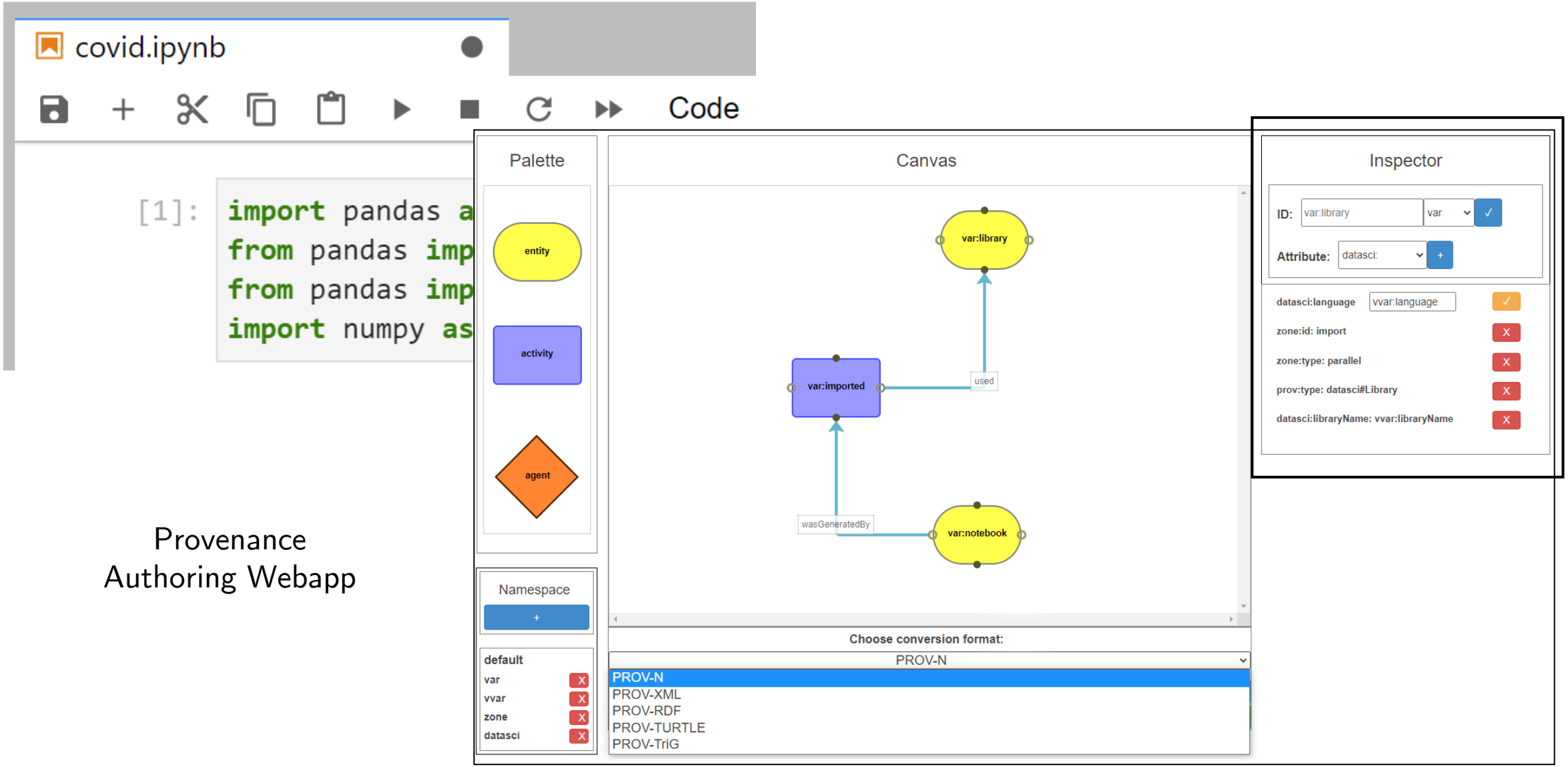


How do I properly specify the data I want to capture?



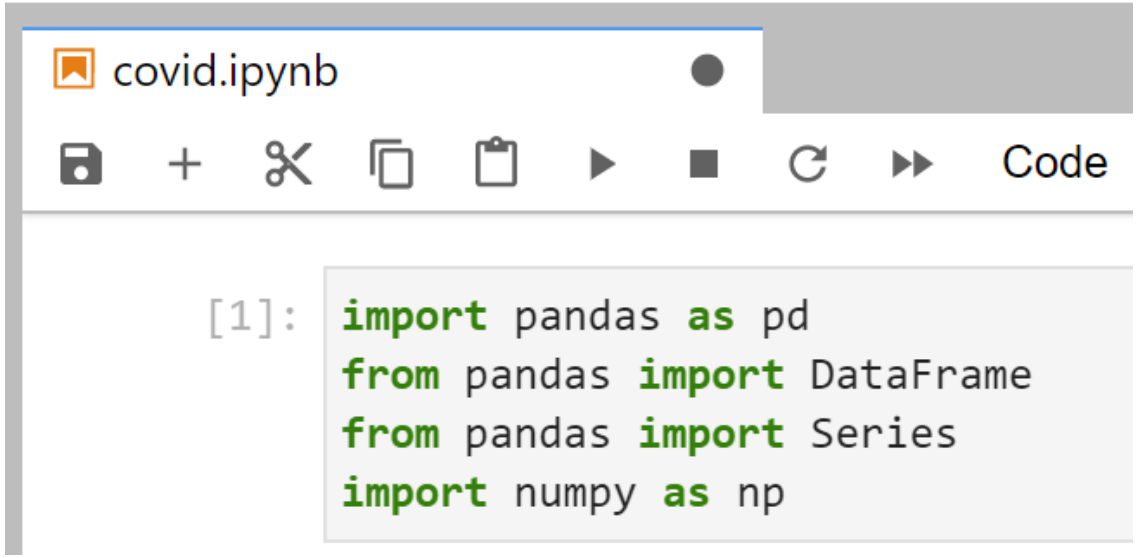
Provenance
Authoring Webapp

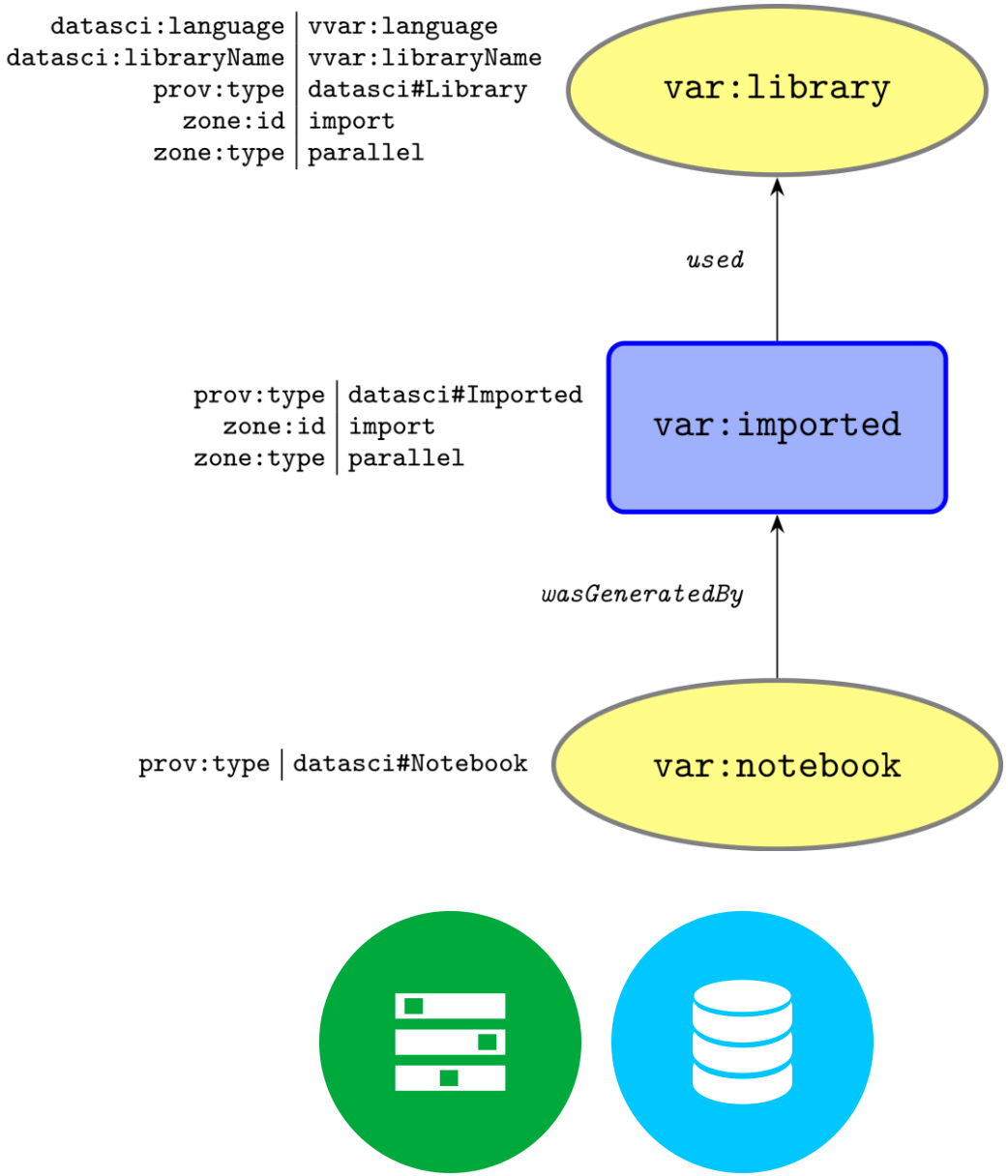
How do I properly specify the data I want to capture?



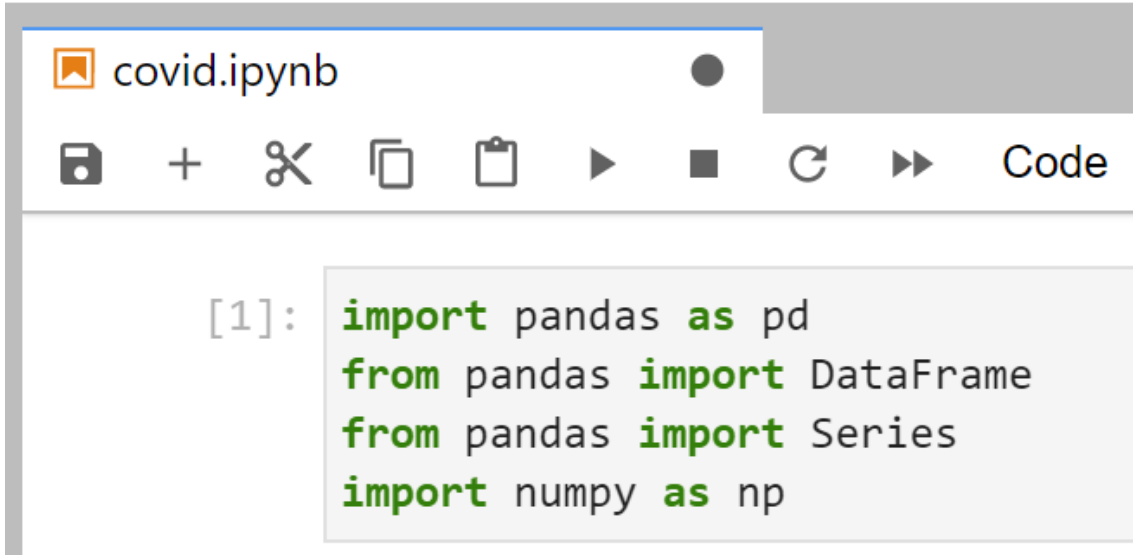
Provenance
Authoring Webapp

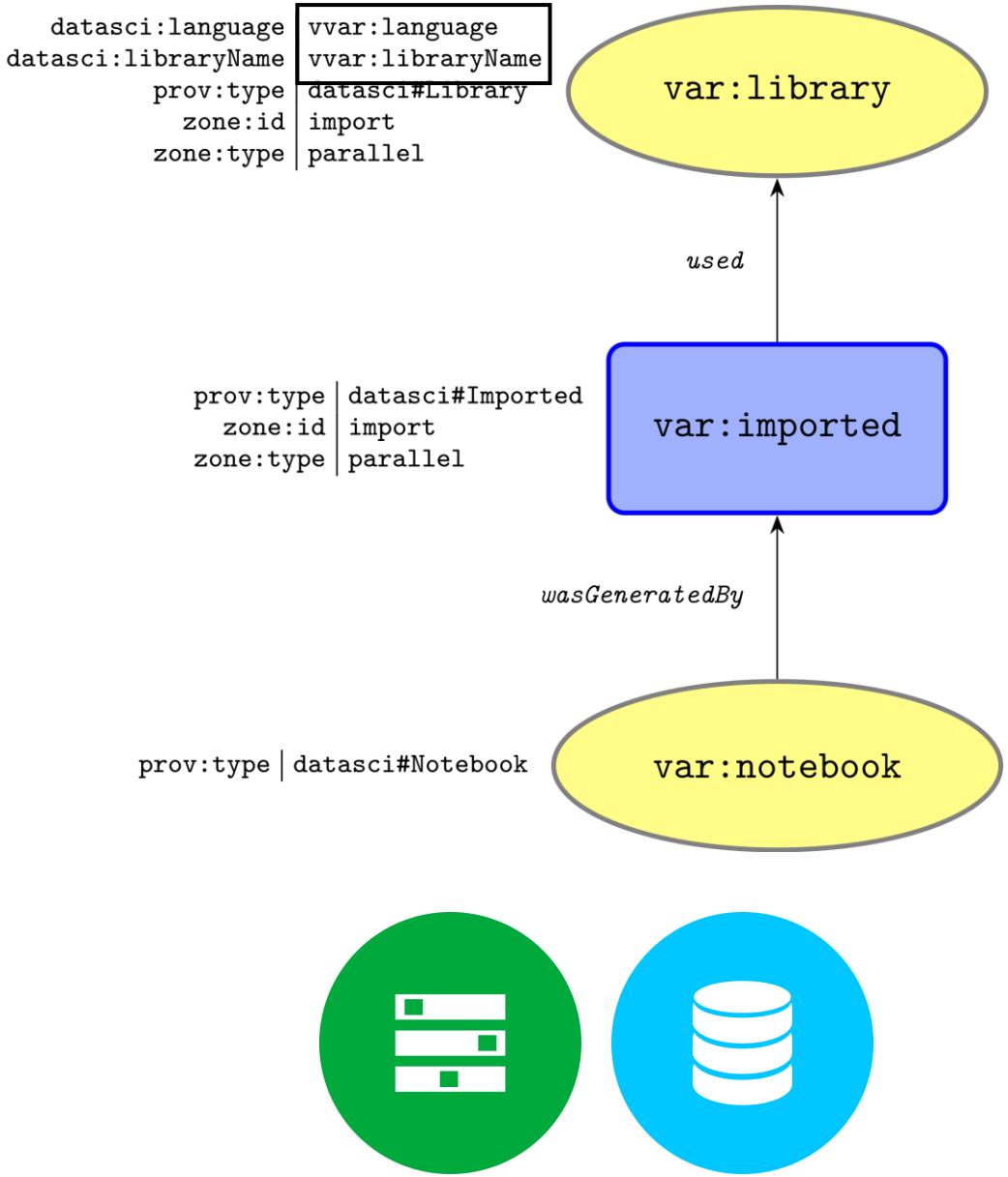
How do I extract this data from a running system?



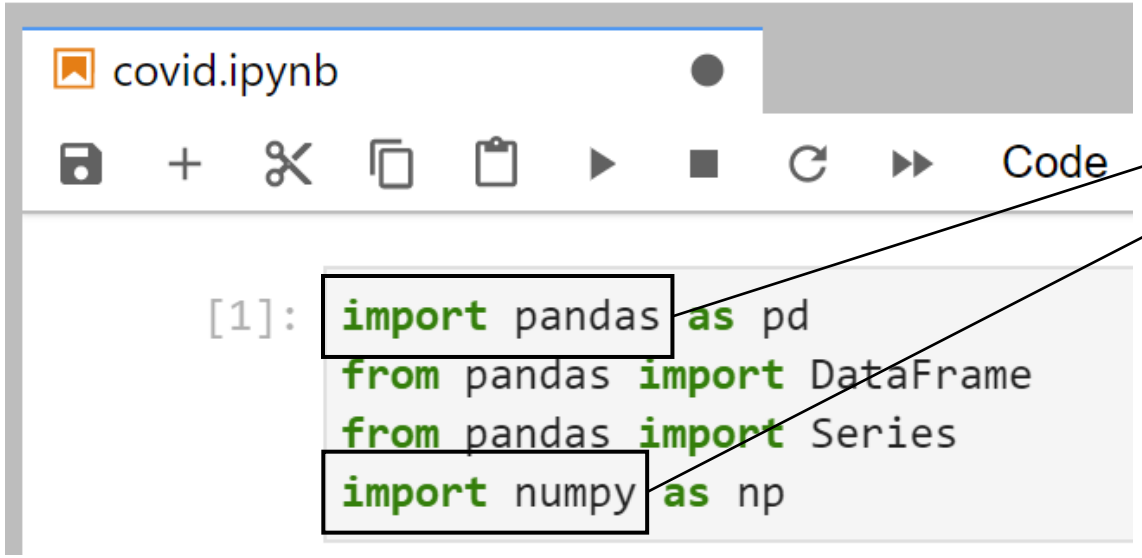


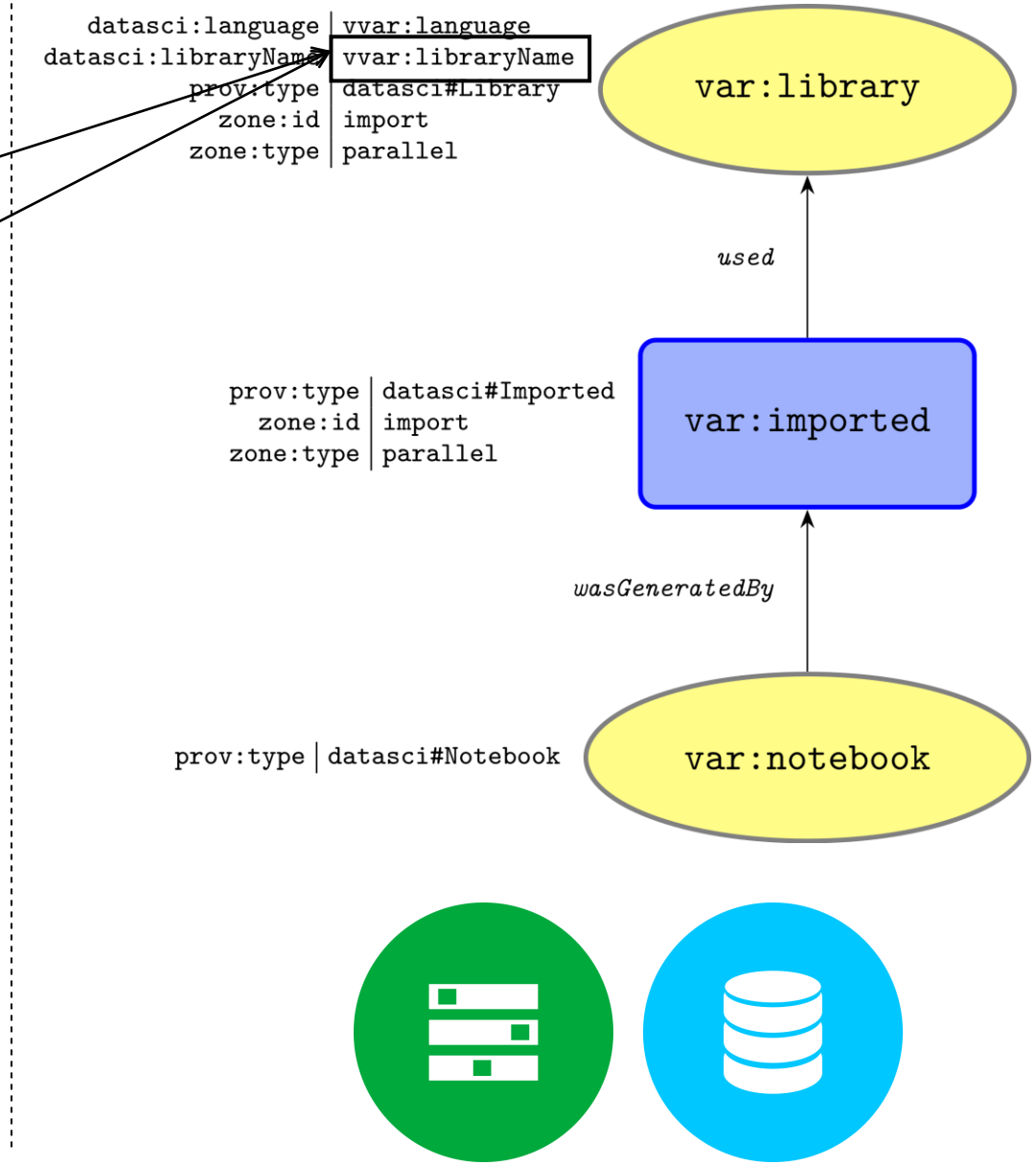
How do I extract this data from a running system?



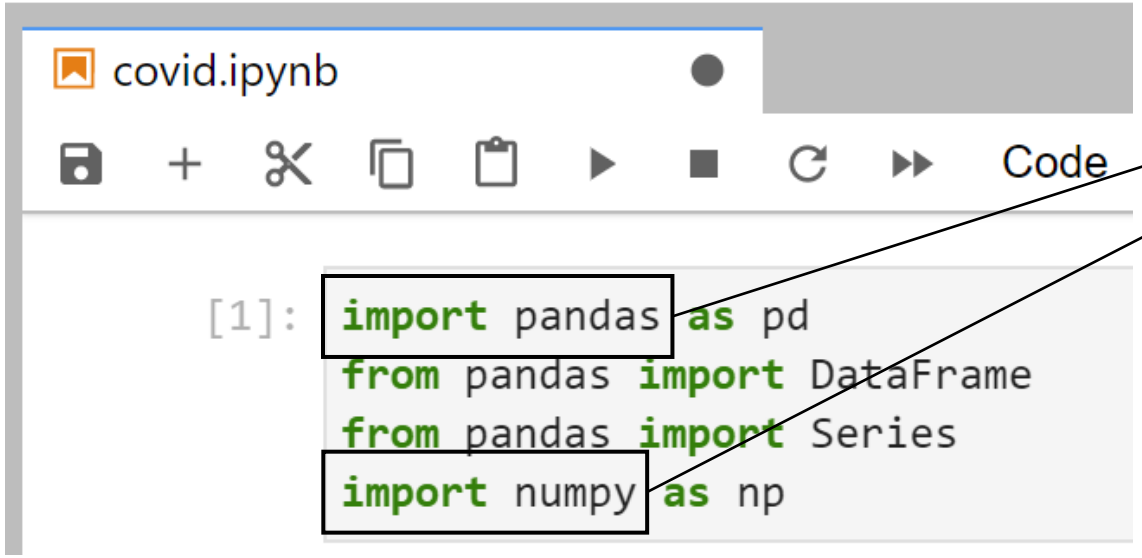


How do I extract this data from a running system?

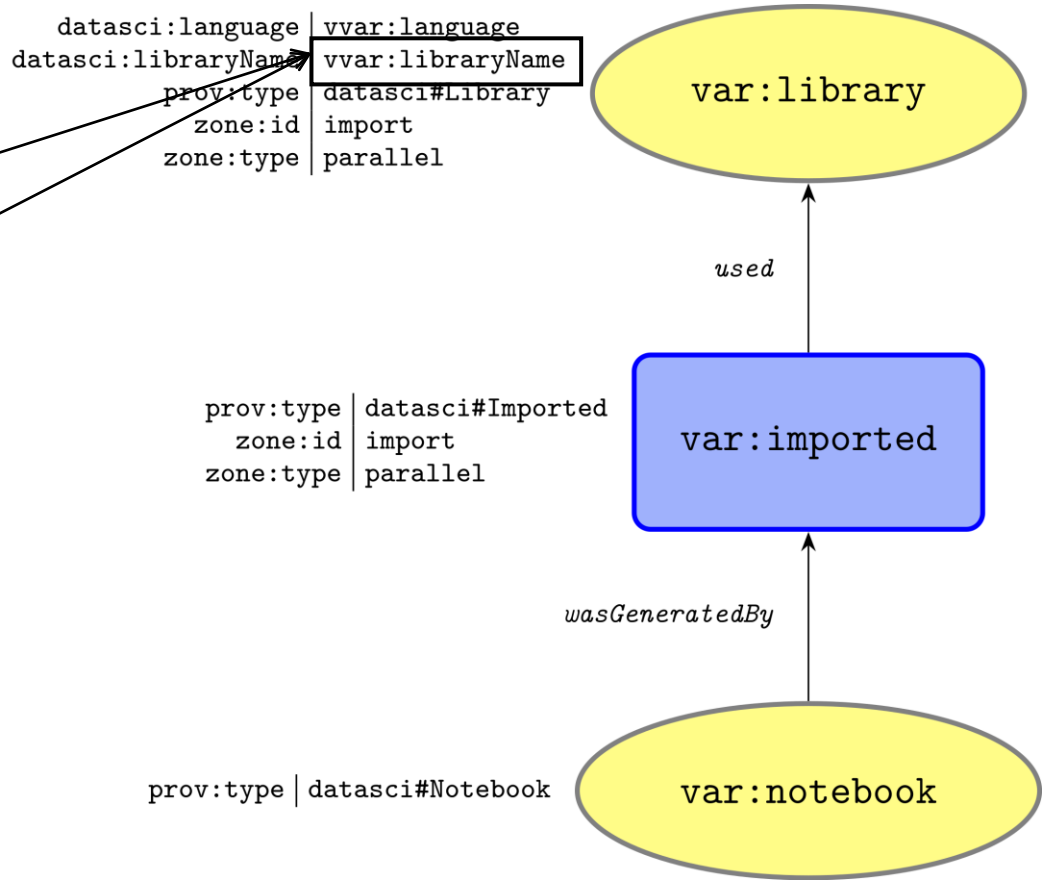




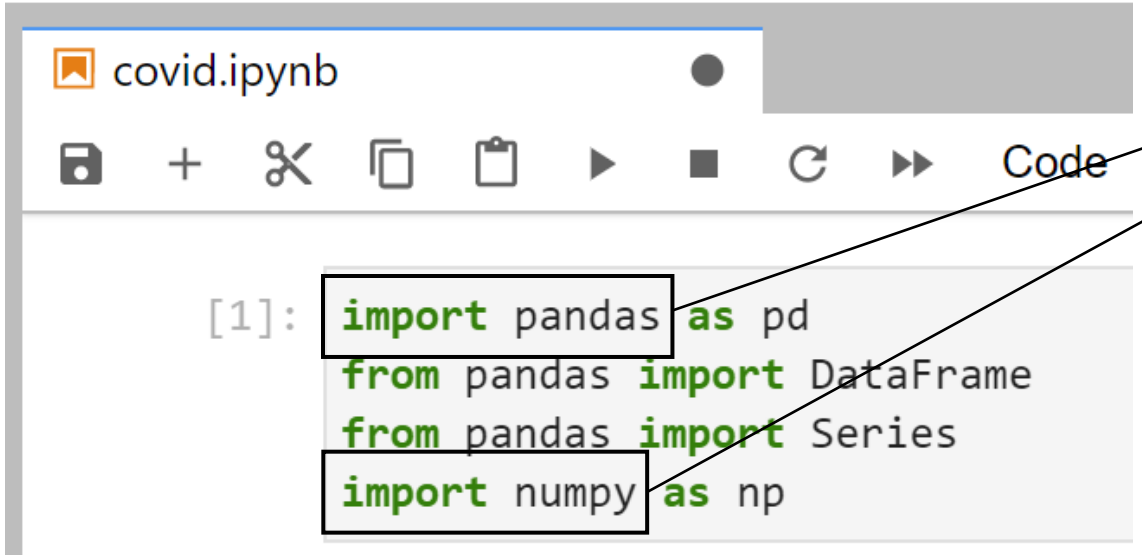
How do I extract this data from a running system?

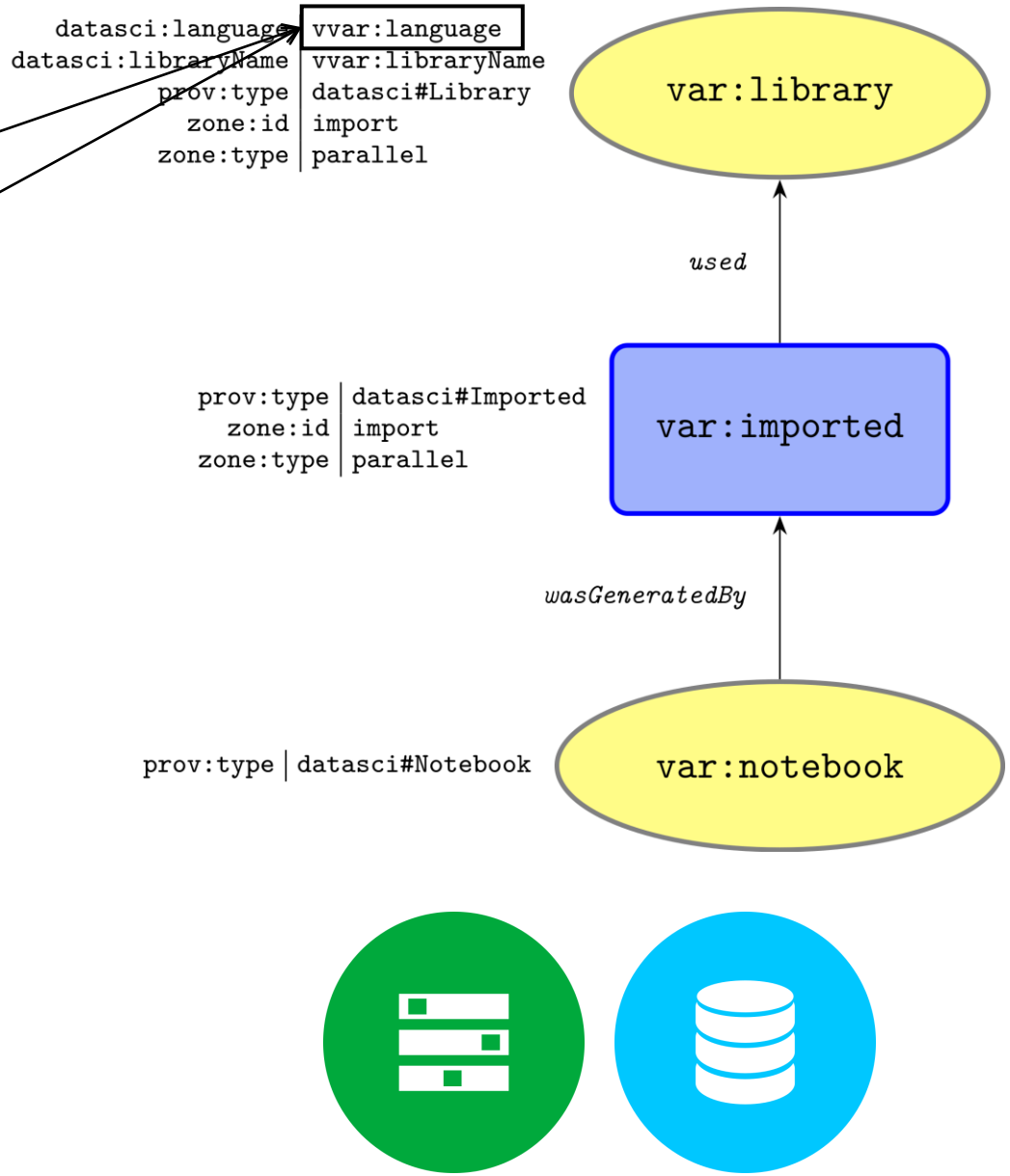


Pattern: import [ ]
Action: extract



How do I extract this data from a running system?

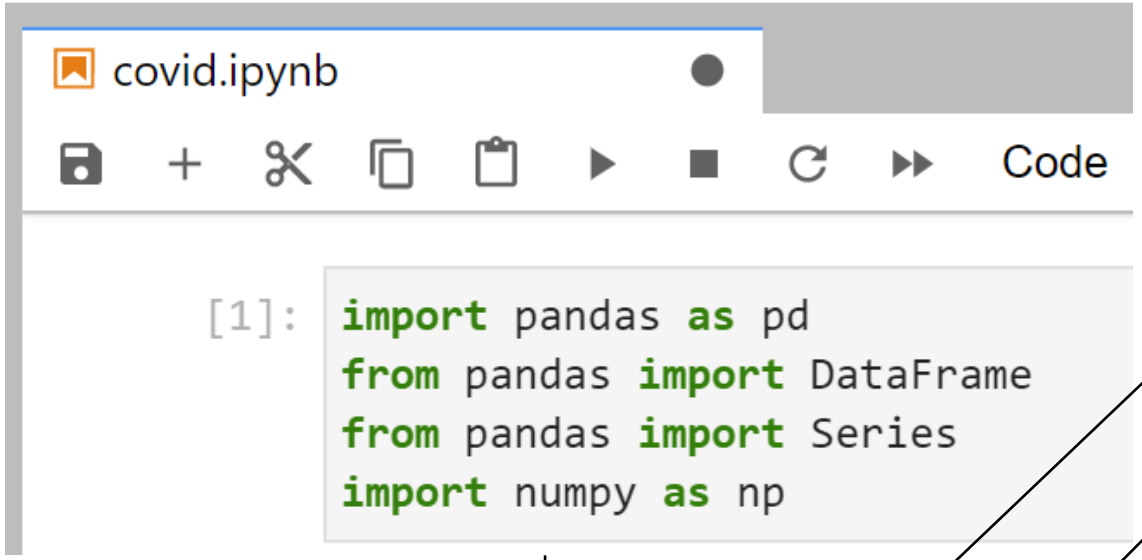




Pattern: import [ ]
Action: extract

Pattern: import [ ]
Action: use string 'python'

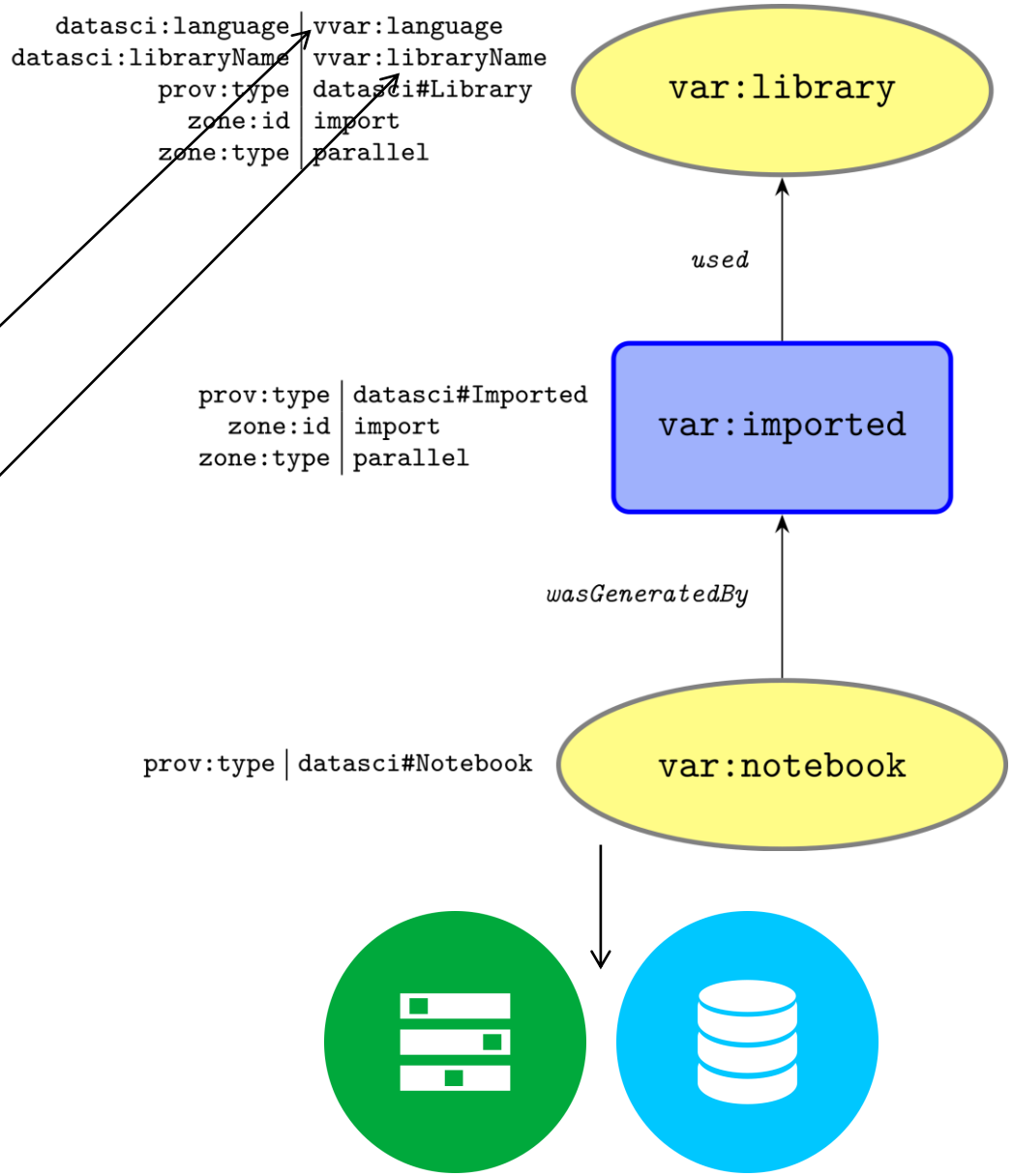
How do I extract this data from a running system?



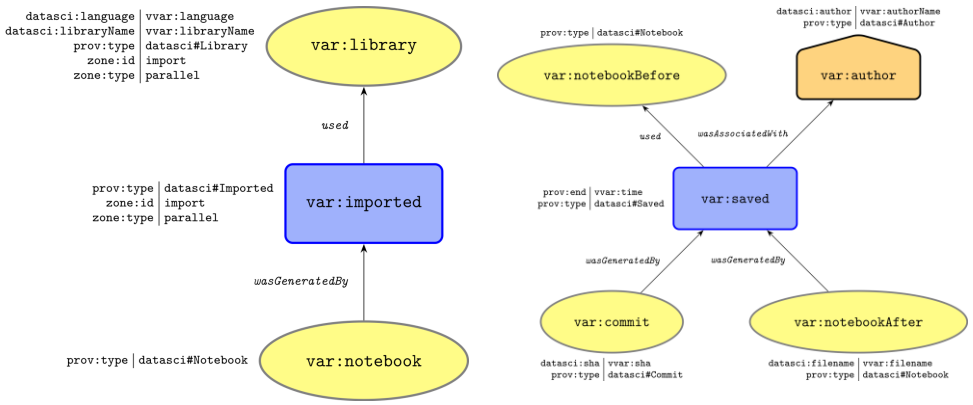
Pattern: import [ ]
Action: extract

Pattern: import [ ]
Action: use string 'python'

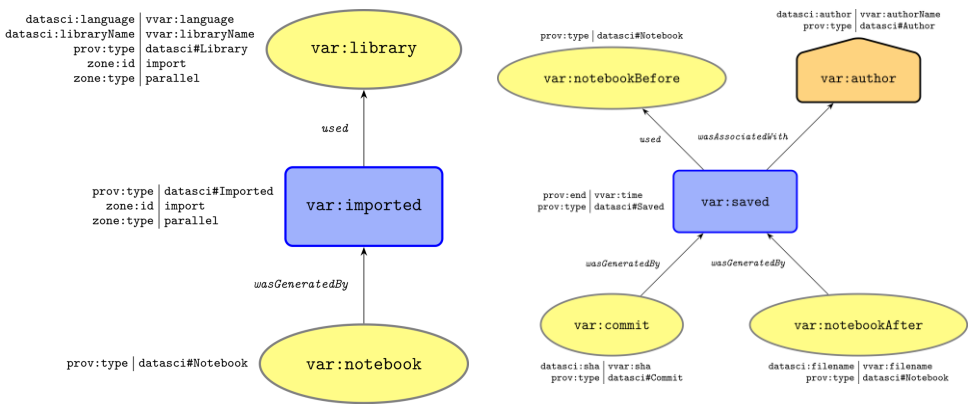
Template Service Client

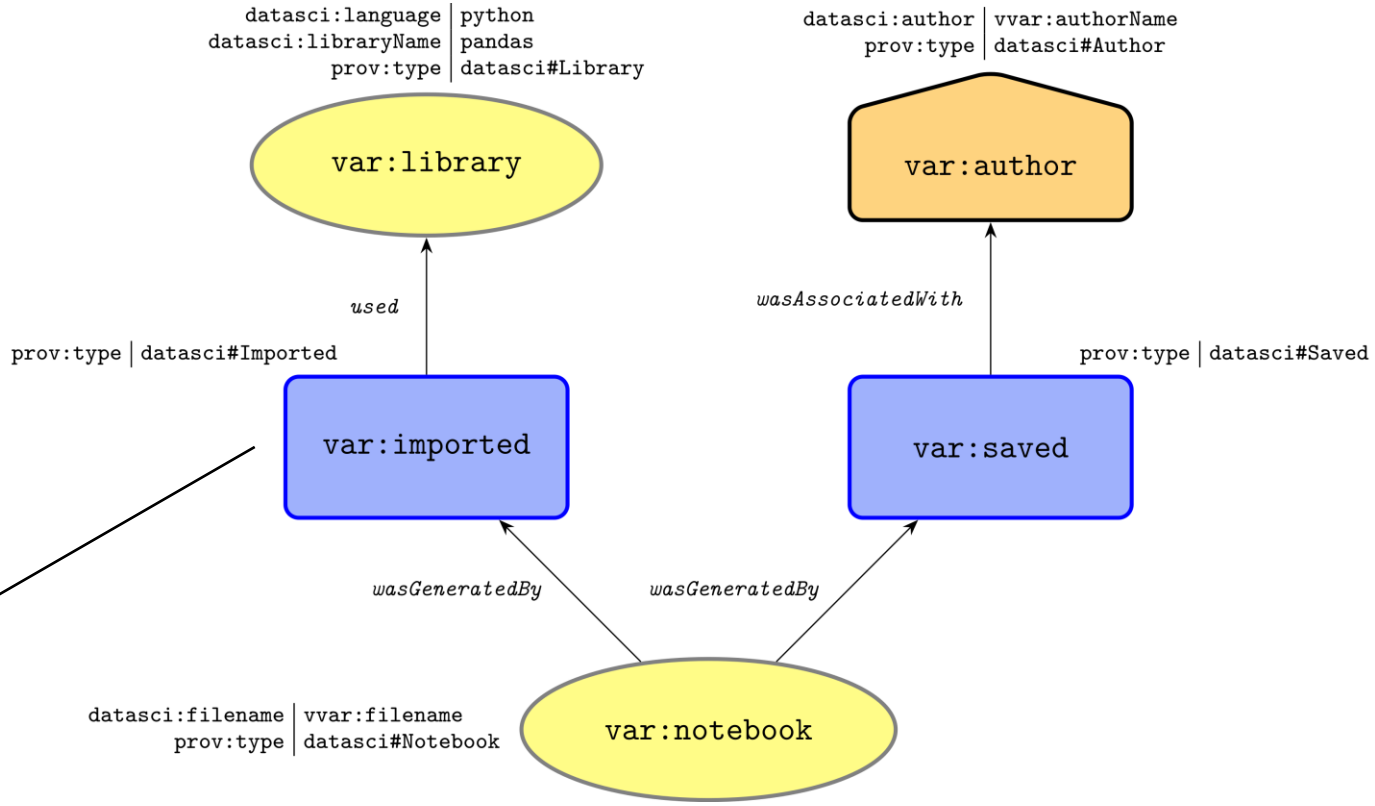


How do I get my data back, once it's been stored?



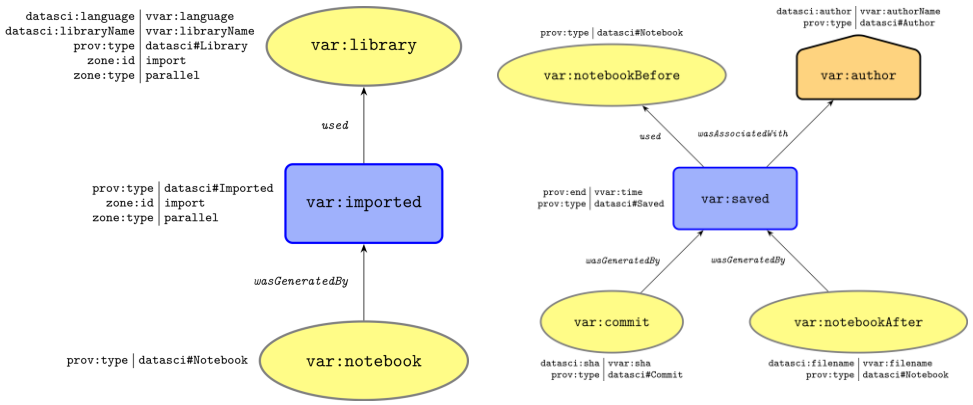
How do I get my data back, once it's been stored?

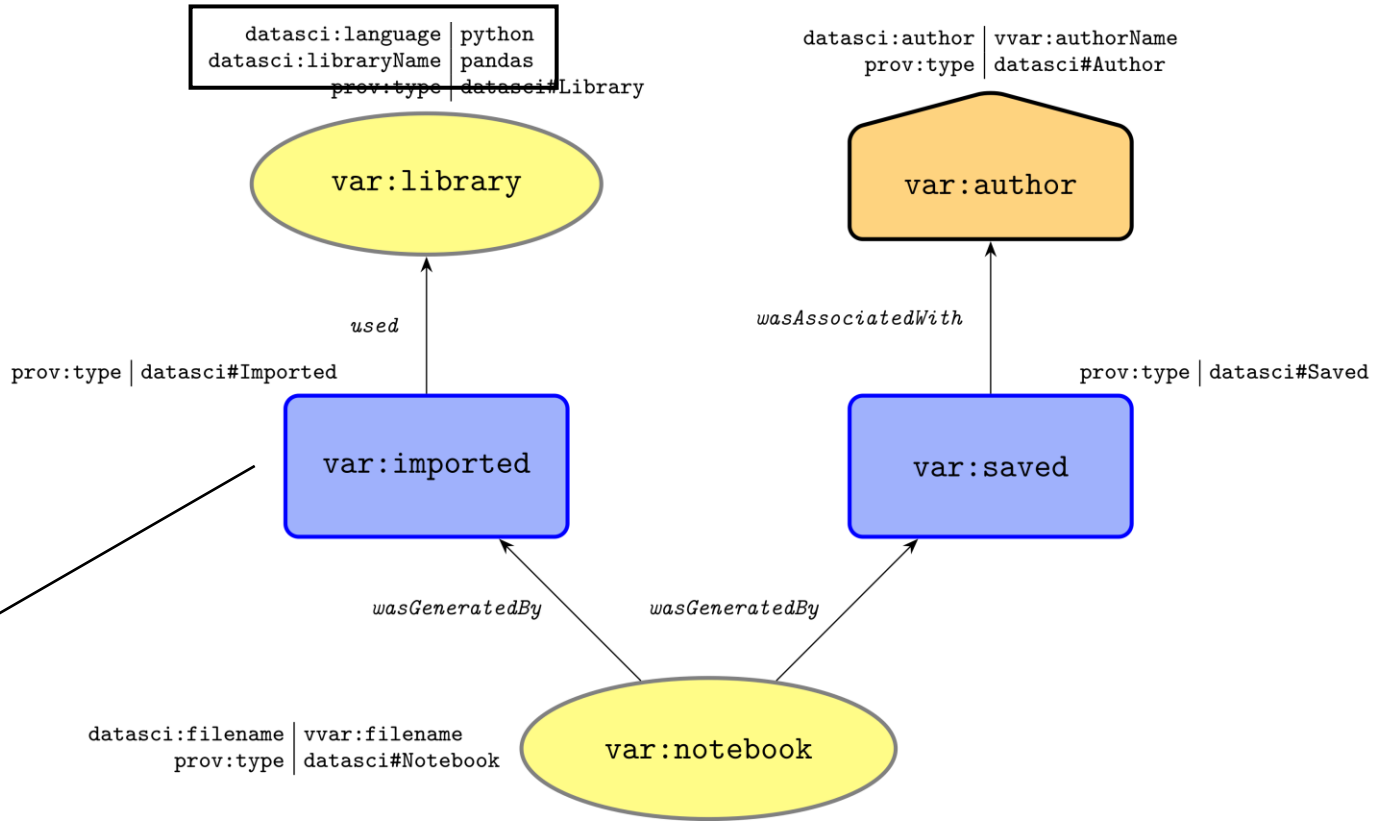




Query Templates

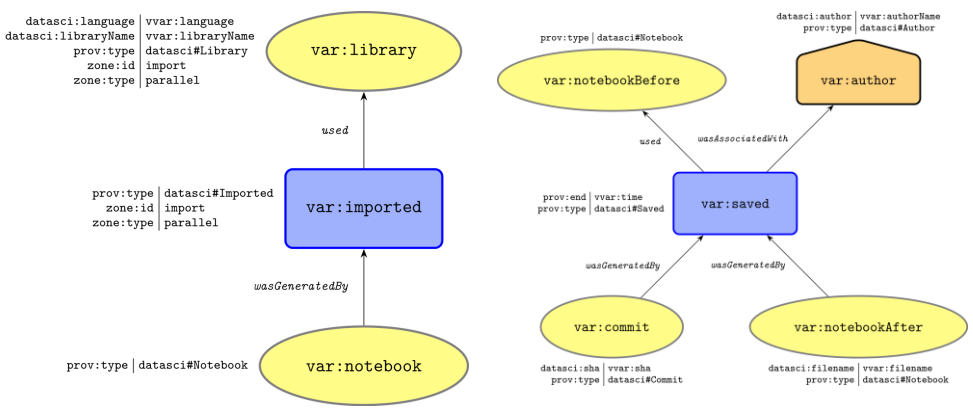
How do I get my data back, once it's been stored?

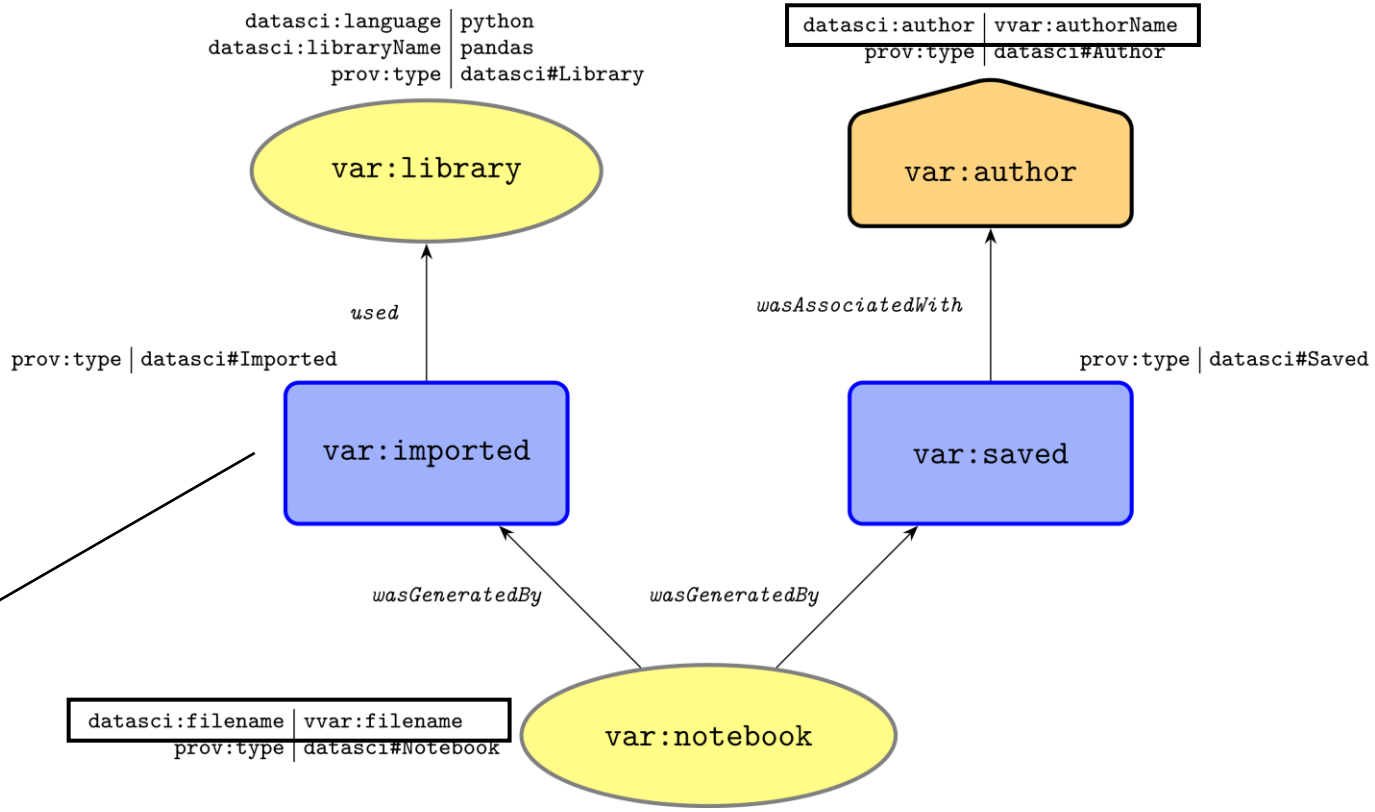




Query Templates

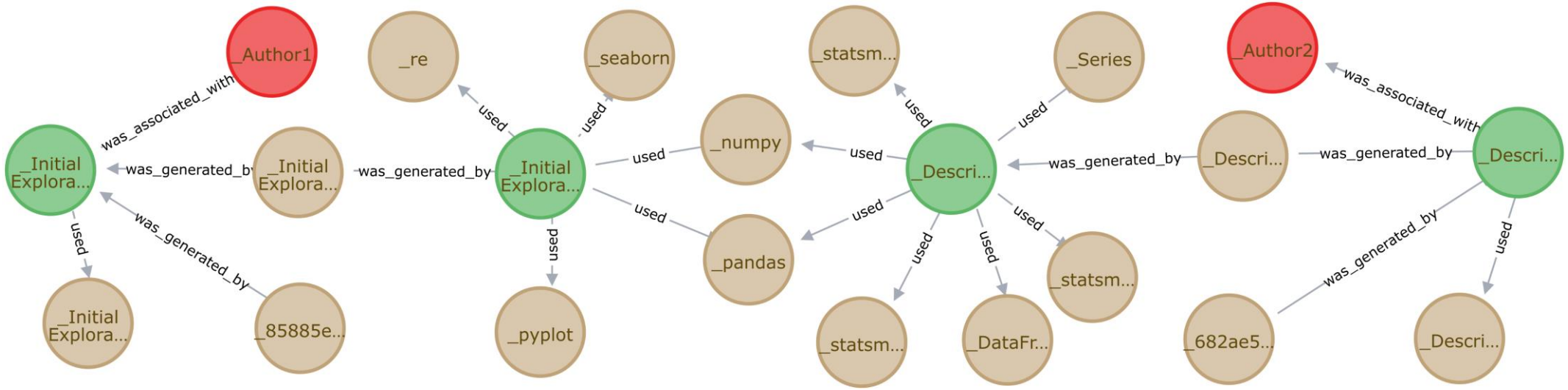
How do I get my data back, once it's been stored?





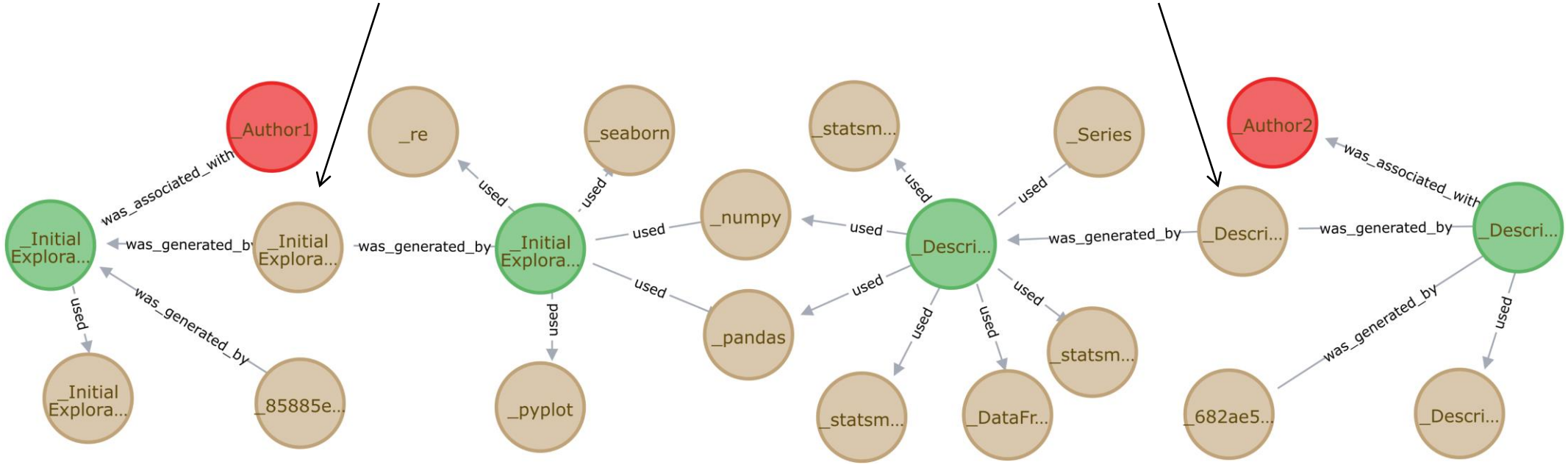
Query Templates

COVID-19 Analytics in Jupyter

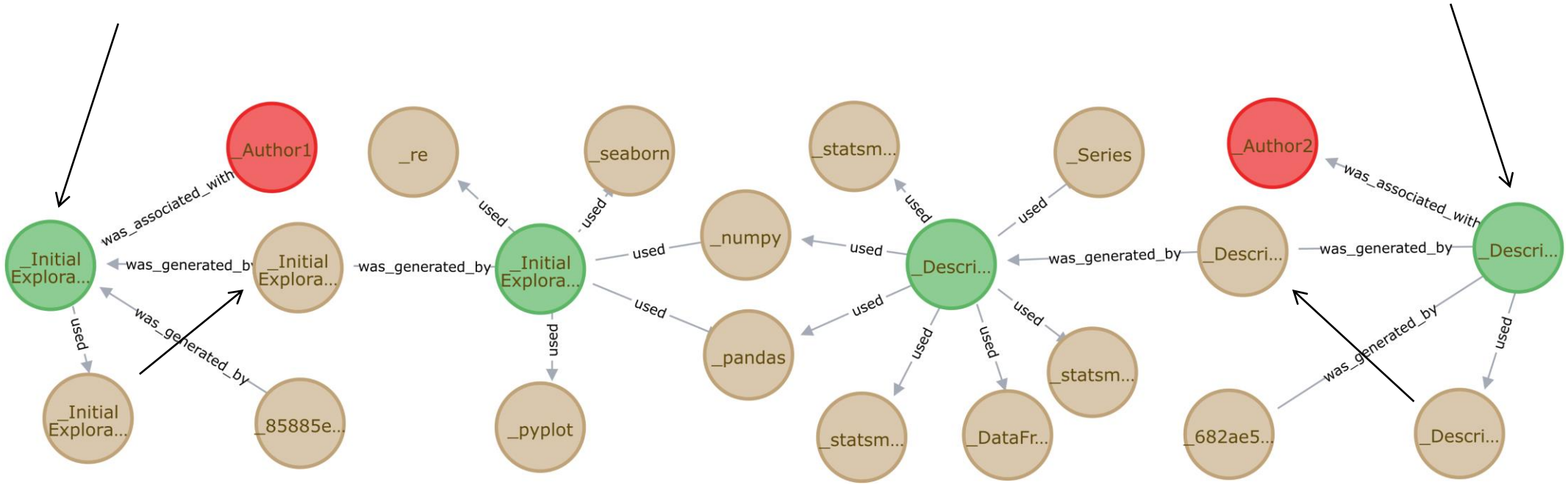


Data relating to 1468 patients who tested positive at Guy's and St. Thomas' NHS Foundation Trust (GSTT), analysed by a group of researchers between April 2020 and February 2021 at King's College London.

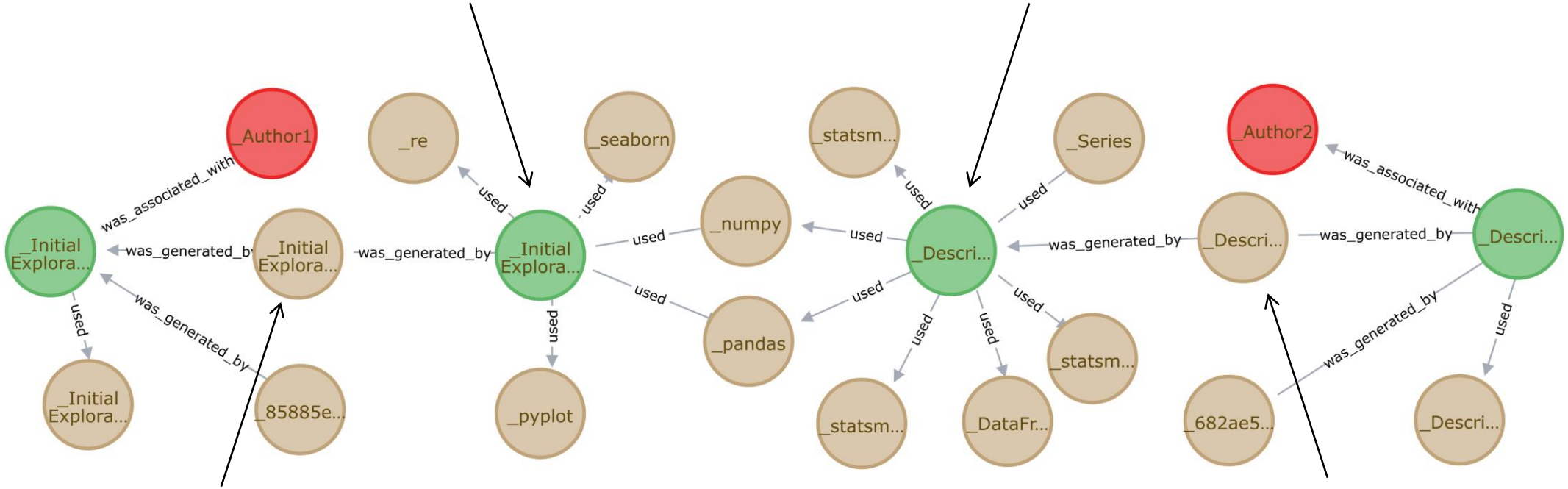
COVID-19 Analytics in Jupyter



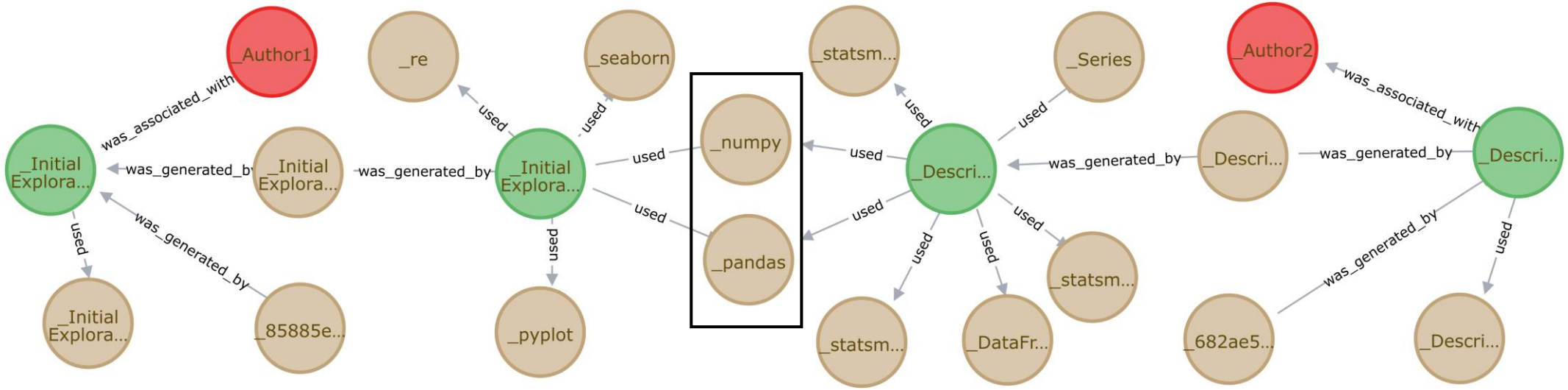
COVID-19 Analytics in Jupyter



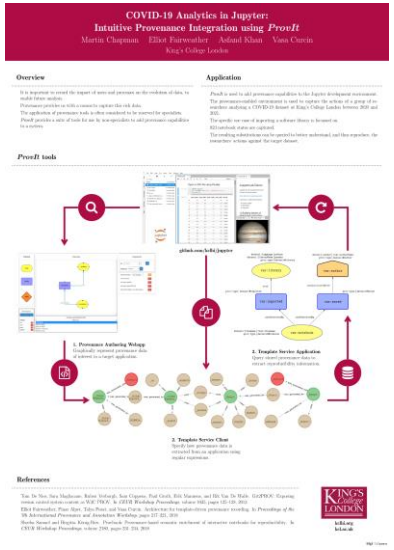
COVID-19 Analytics in Jupyter



COVID-19 Analytics in Jupyter



When this is expanded across multiple saves, we can see what was imported when, and by whom; we can also see more complex information, such as common libraries



COVID-19 Analytics in Jupyter: Intuitive Provenance Integration using *ProvIt*

Martin Chapman, Elliot Fairweather, Asfand Khan and Vasa Curcin

King's College London

Provenance-enabled Jupyter environment: https://github.com/kclhi/jupyter