PROVENANCE

Matt Jones
DataONE Provenance Team

Provenance in DataONE

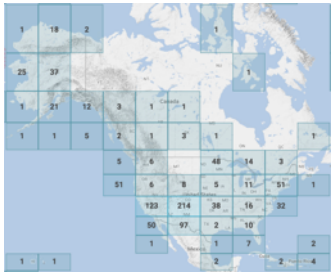
Phase II Goal: Facilitate reproducible science

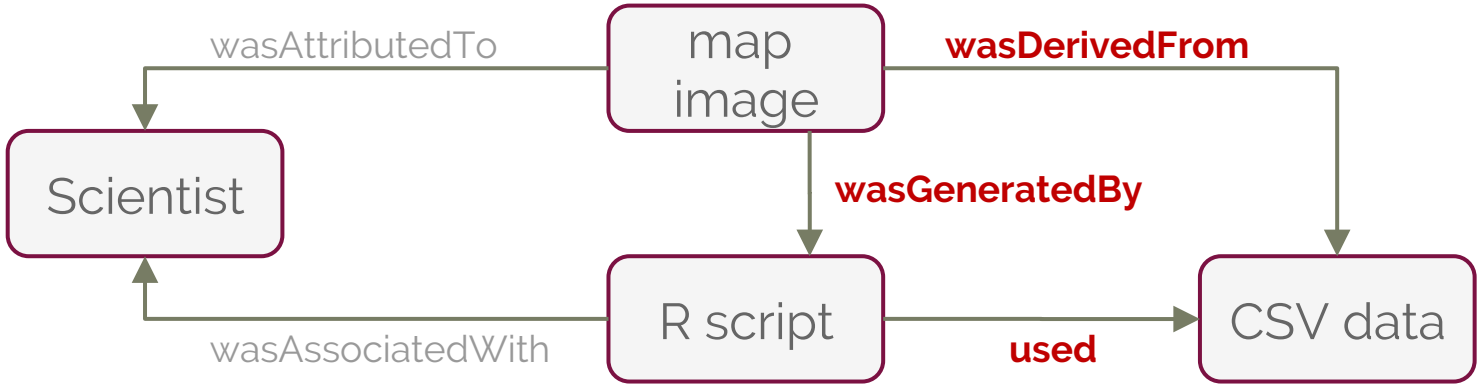
- Track **data derivation** history
- Track data **inputs** and **outputs** of analyses
- Track analysis and model **executions**
- Preserve and document software **workflows**
- Link all of these to **publications**

Provenance for Science Workflows

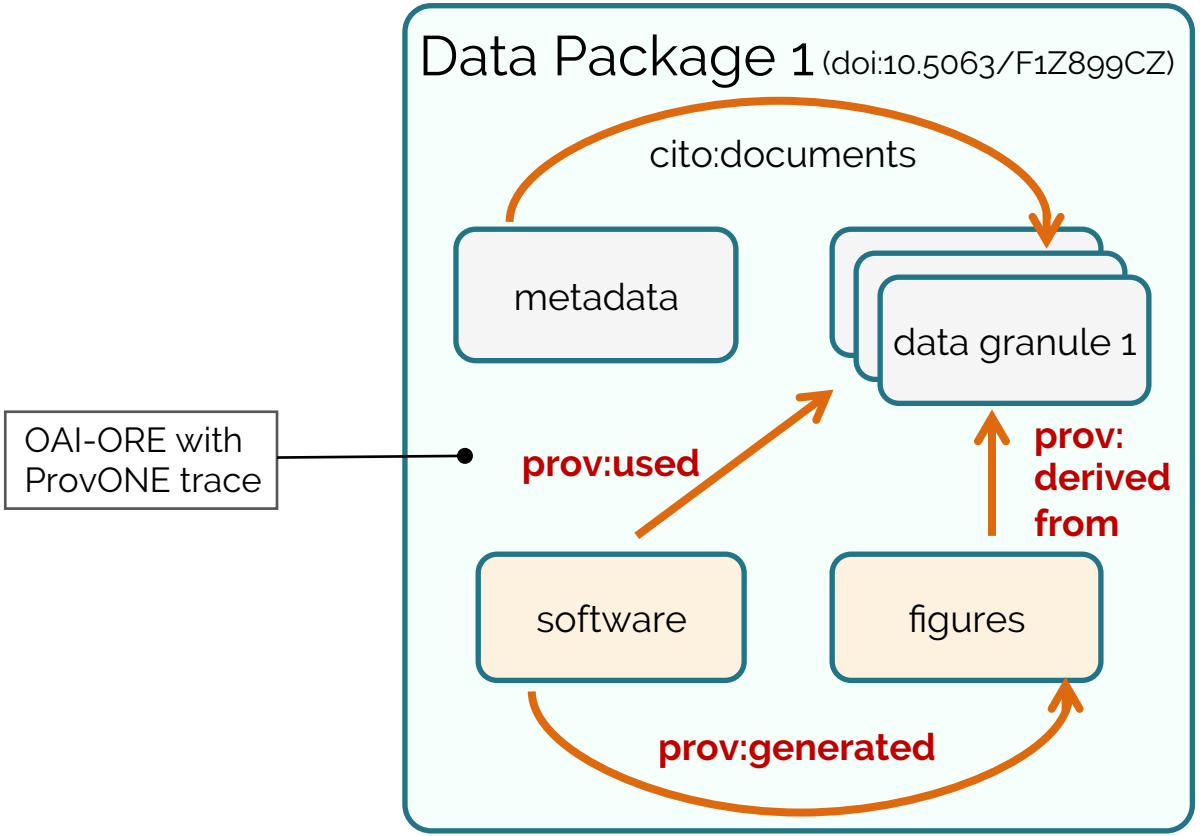
ProvONE – an extension of W3C PROV

See **purl.dataone.org/provone-v1-dev**

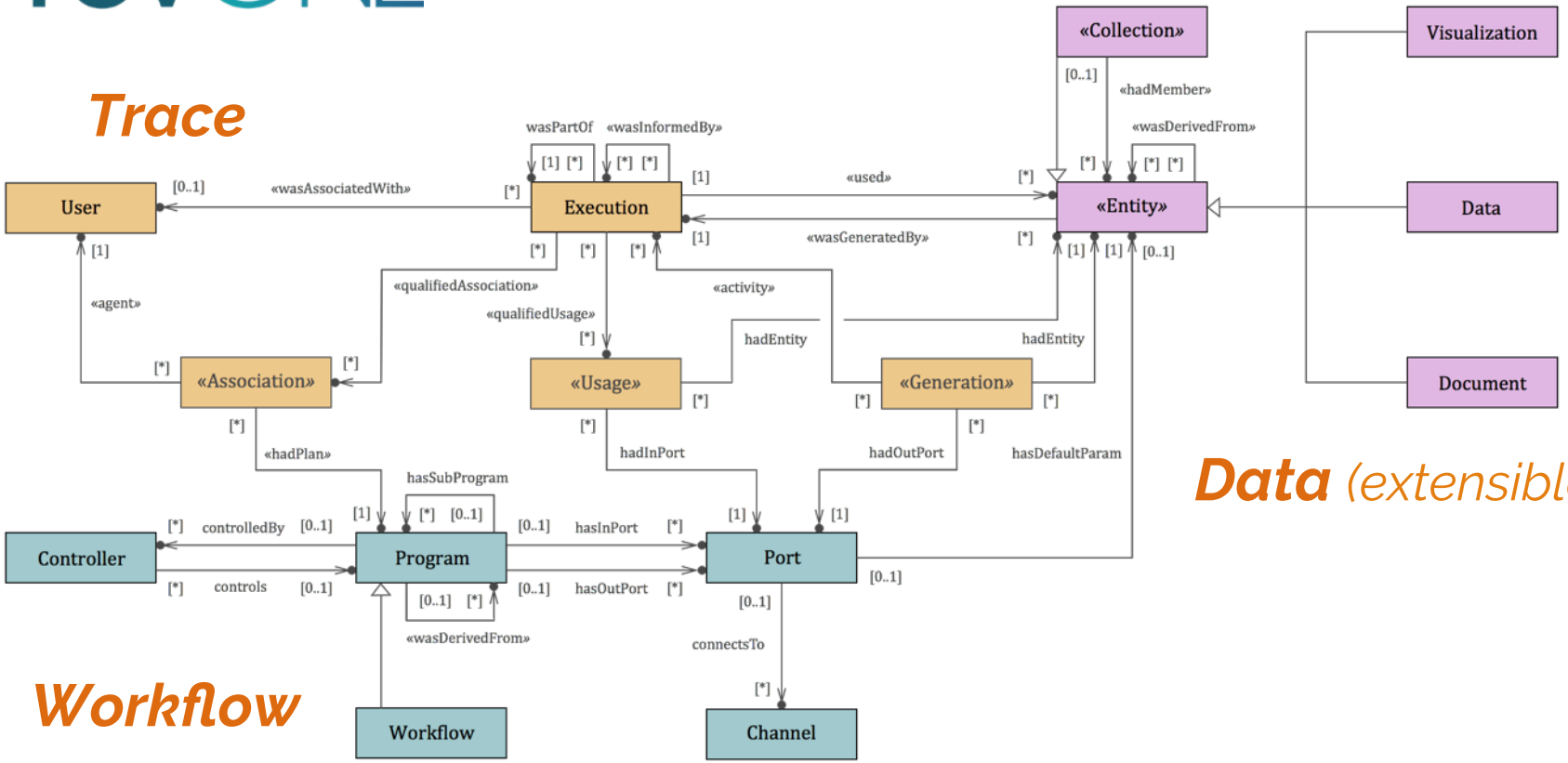




Data Package with Provenance

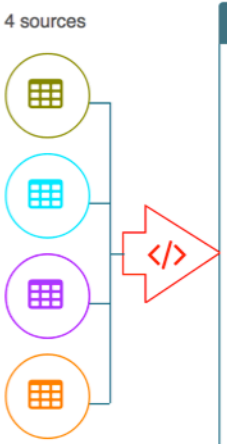


ProvONE



See purl.dataone.org/provone-v1-dev

Data Table, Image, and Other Data Details

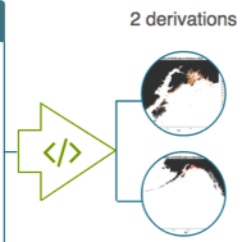


Data Table

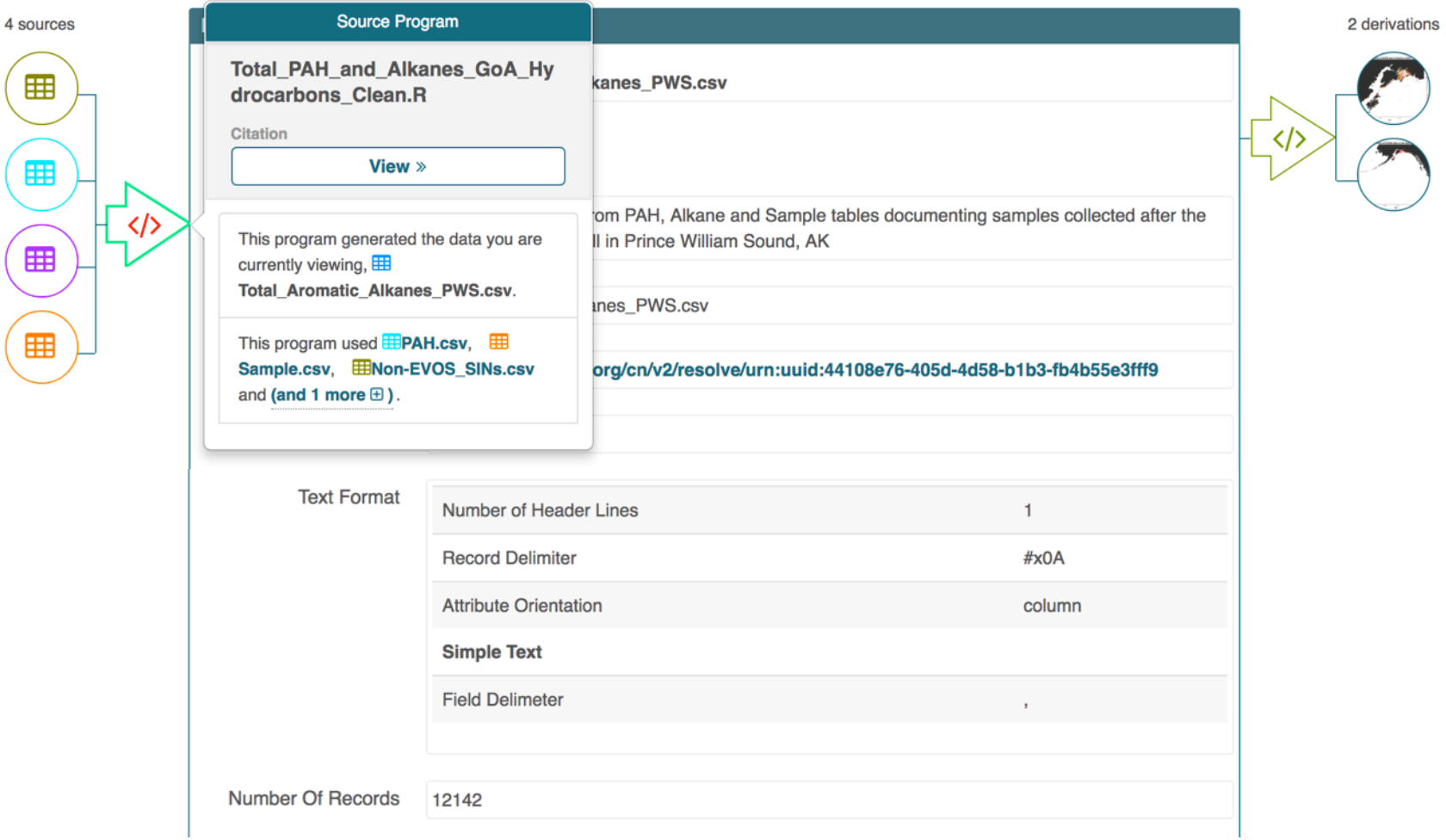
Entity Name	**Total_Aromatic_Alkanes_PWS.csv**
	Download
Description	Combined dataset from PAH, Alkane and Sample tables documenting samples collected after the Exxon Valdez oil spill in Prince William Sound, AK
Object Name	Total_Aromatic_Alkanes_PWS.csv
Online Distribution Info	**https://cn.dataone.org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9**
Size	2801033 byte
Text Format	

Number of Header Lines	1
Record Delimiter	#x0A
Attribute Orientation	column
Simple Text	
Field Delimeter	,

Number Of Records	12142

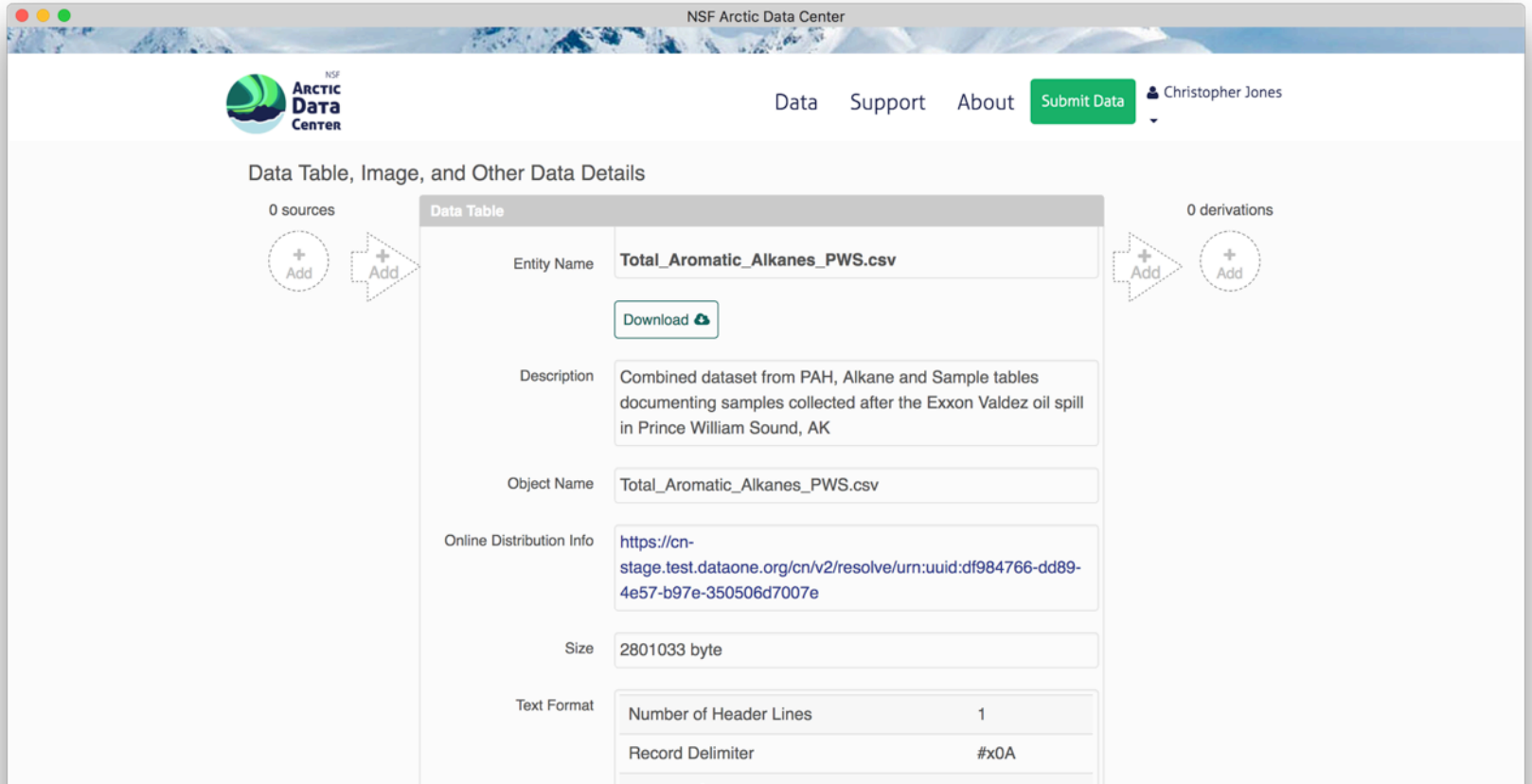


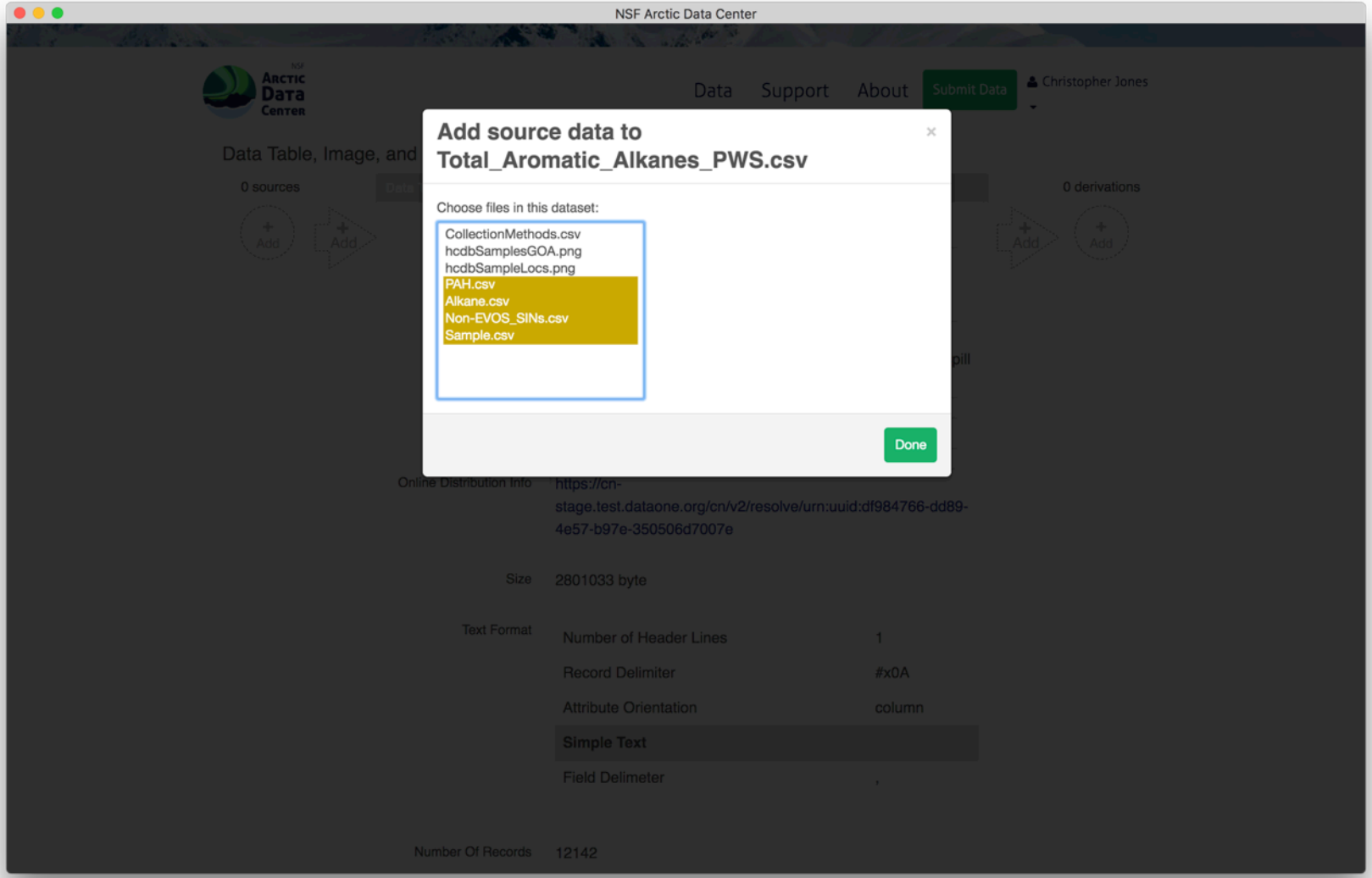
Data Table, Image, and Other Data Details

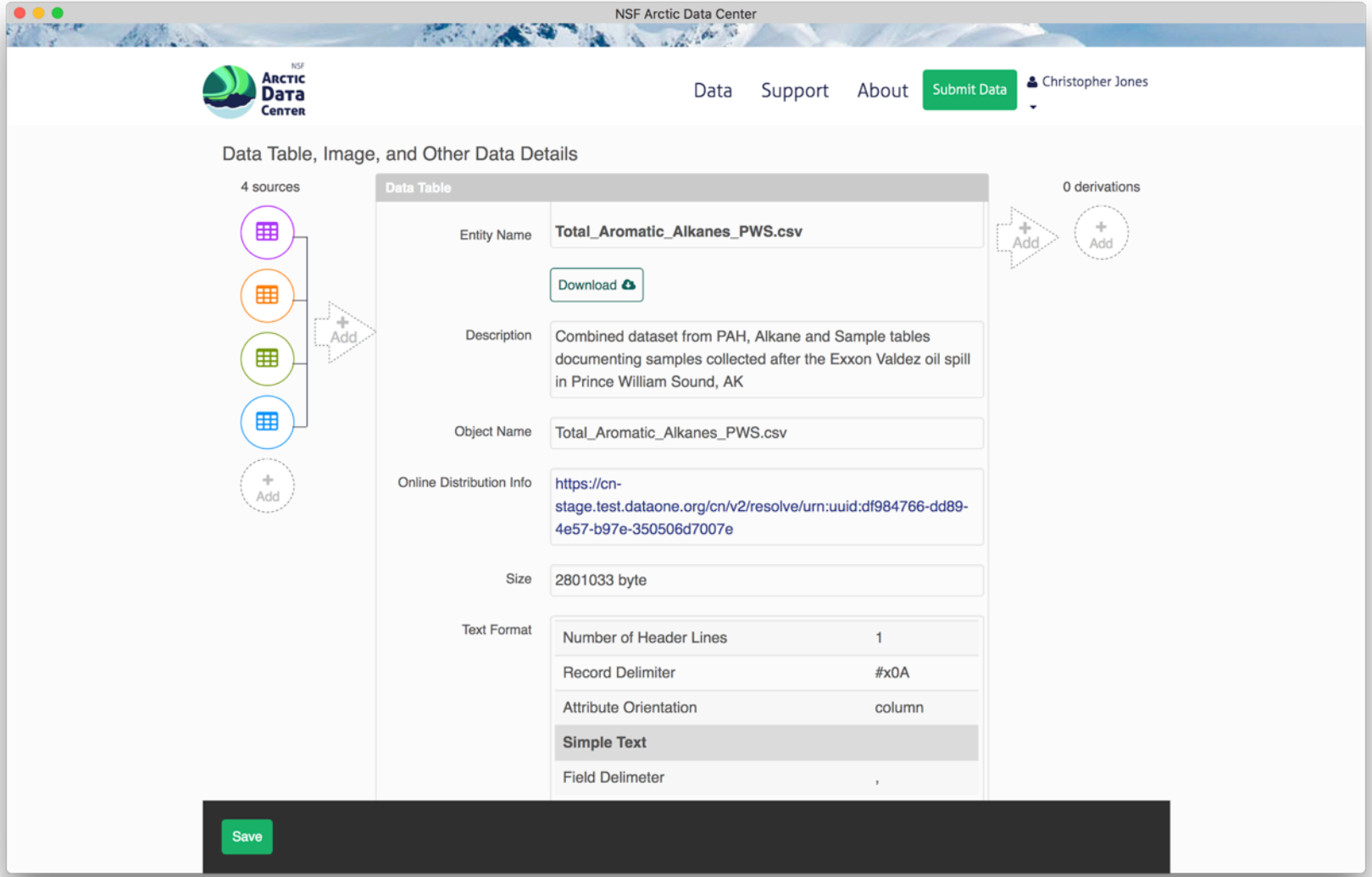


Web Provenance Editor

Deployed by Arctic Data Center



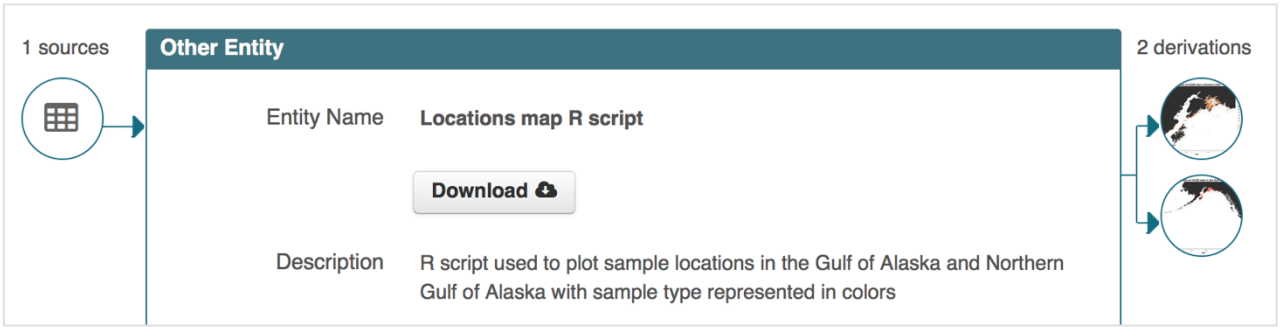
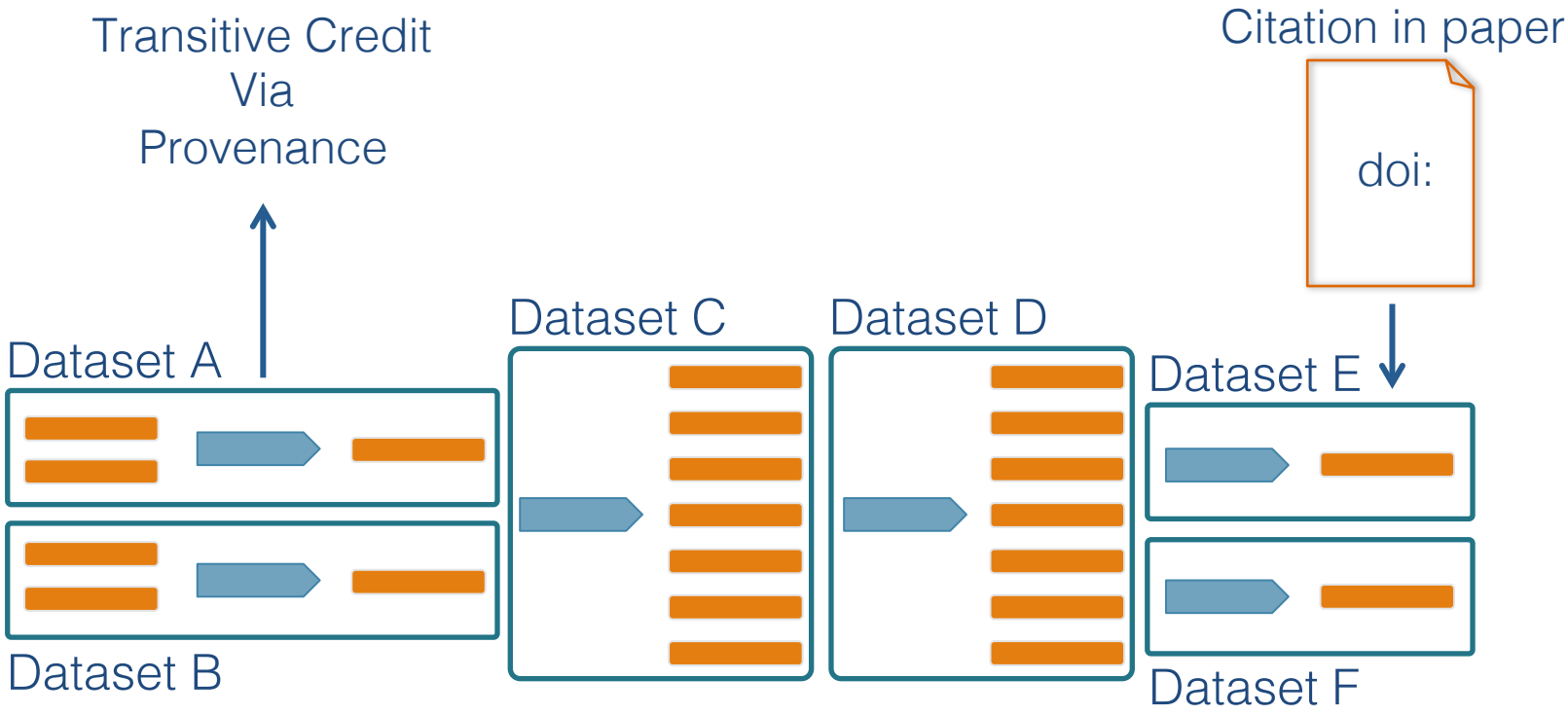

NSF Arctic Data Center
Arctic Data Center
Data
Support
About
Submit Data
Christopher Jones
Data Table, Image, and
0 sources
0 derivations
Add
Add
Add
Add
Add source data to
Total_Aromatic_Alkanes_PWS.csv
Choose files in this dataset:
CollectionMethods.csv
hcdbSamplesGOA.png
hcdbSampleLocs.png
PAH.csv
Alkane.csv
Non-EVOS_SINs.csv
Sample.csv
Done
Online Distribution Info
https://cn-stage.test.dataone.org/cn/v2/resolve/urn:uuid:df984766-dd89-4e57-b97e-350506d7007e
Size
2801033 byte
Text Format
Number of Header Lines
1
Record Delimiter
#x0A
Attribute Orientation
column
Simple Text
Field Delimeter
,
Number Of Records
12142


NSF Arctic Data Center
NSF Arctic Data Center
Data
Support
About
Submit Data
Christopher Jones
Data Table, Image, and Other Data Details
4 sources
Add
Add
Data Table
Entity Name
Total_Aromatic_Alkanes_PWS.csv
Download
Description
Combined dataset from PAH, Alkane and Sample tables documenting samples collected after the Exxon Valdez oil spill in Prince William Sound, AK
Object Name
Total_Aromatic_Alkanes_PWS.csv
Online Distribution Info
https://cn-stage.test.dataone.org/cn/v2/resolve/urn:uuid:df984766-dd89-4e57-b97e-350506d7007e
Size
2801033 byte
Text Format
Number of Header Lines
1
Record Delimiter
#x0A
Attribute Orientation
column
Simple Text
Field Delimeter
,
0 derivations
Add
Add
Save

R ‘recordr’ library

Record or Assert provenance information

```
1  # Generate map of locations by type
2  library(recordr)
3  recordr <- new("Recordr")
4  pkg <- record(recordr, "./hcdbSites.R", "loc-by-type-png")
```



Citing multi-generational workflows