



**PROVENANCE**

**Matt Jones**  
**DataONE Provenance Team**

# Provenance in DataONE

---

Phase II Goal: Facilitate reproducible science

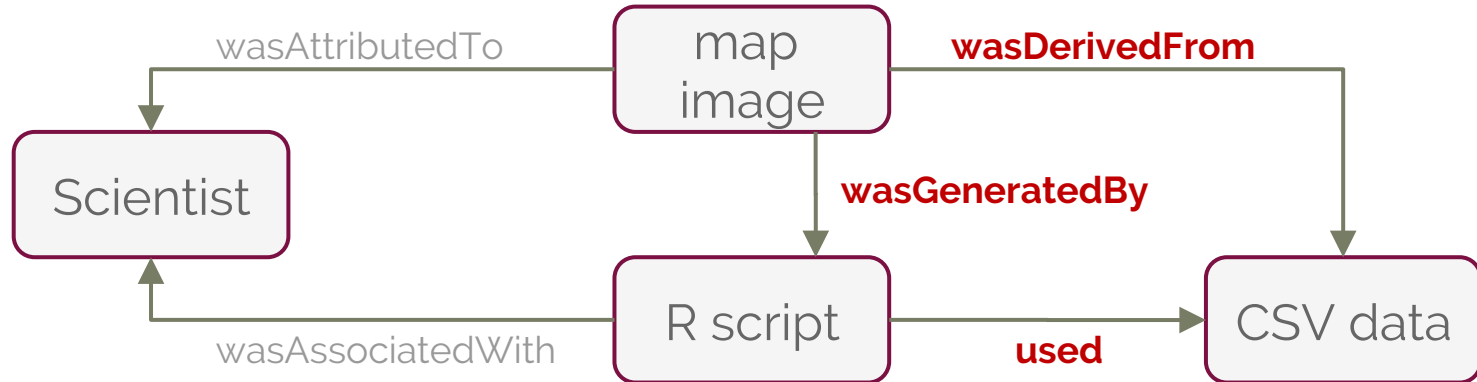
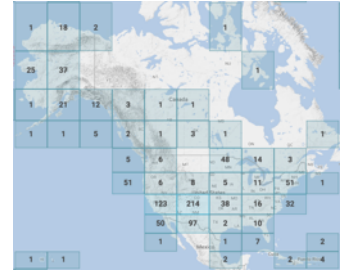
- Track **data derivation** history
- Track data **inputs** and **outputs** of analyses
- Track analysis and model **executions**
- Preserve and document software **workflows**
- Link all of these to **publications**

# Provenance for Science Workflows

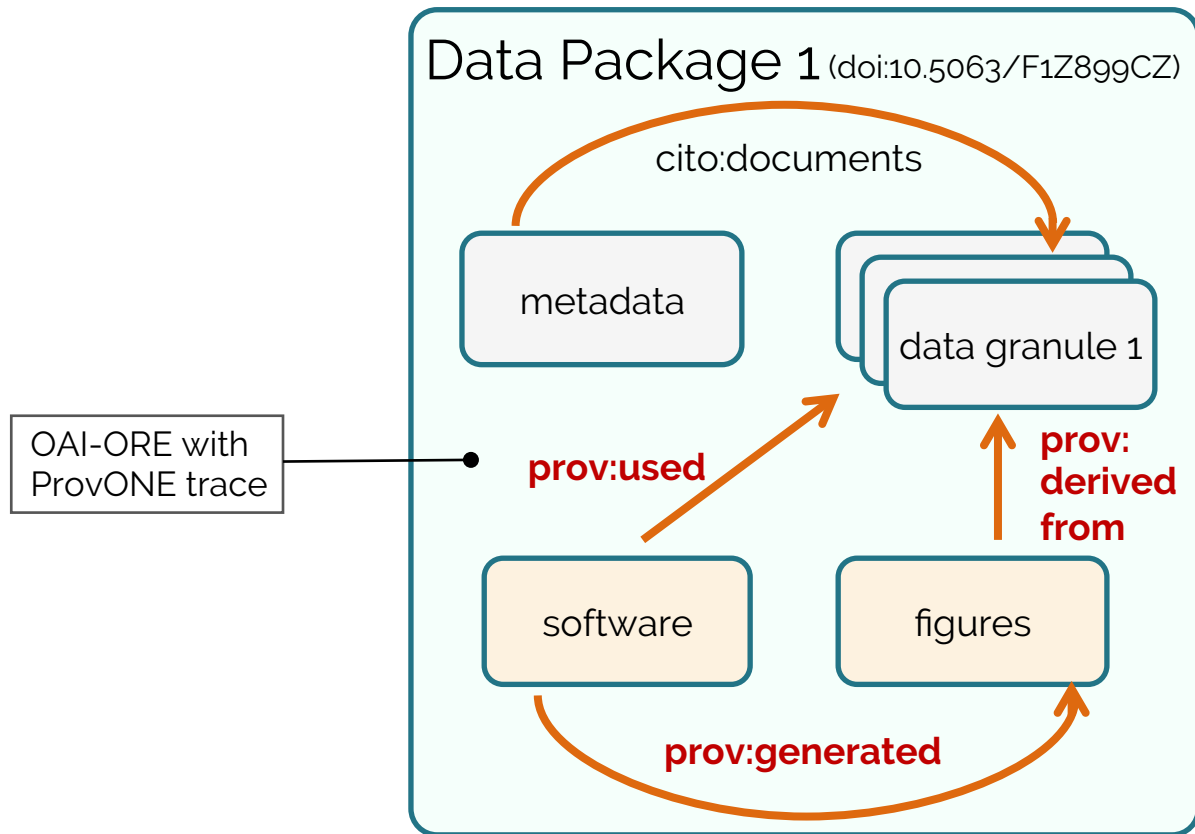


ProvONE – an extension of W3C PROV

See [purl.dataone.org/provone-v1-dev](http://purl.dataone.org/provone-v1-dev)



# Data Package with Provenance

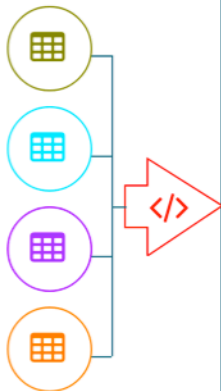




See [purl.dataone.org/provone-v1-dev](http://purl.dataone.org/provone-v1-dev)

## Data Table, Image, and Other Data Details

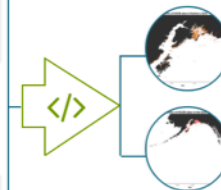
4 sources



### Data Table

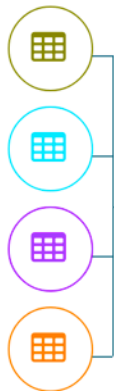
Entity Name	Total_Aromatic_Alkanes_PWS.csv										
	<a href="#">Download</a>										
Description	Combined dataset from PAH, Alkane and Sample tables documenting samples collected after the Exxon Valdez oil spill in Prince William Sound, AK										
Object Name	Total_Aromatic_Alkanes_PWS.csv										
Online Distribution Info	<a href="https://cn.dataone.org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9">https://cn.dataone.org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9</a>										
Size	2801033 byte										
Text Format	<table><tr><td>Number of Header Lines</td><td>1</td></tr><tr><td>Record Delimiter</td><td>#x0A</td></tr><tr><td>Attribute Orientation</td><td>column</td></tr><tr><td colspan="2"><b>Simple Text</b></td></tr><tr><td>Field Delimiter</td><td>,</td></tr></table>	Number of Header Lines	1	Record Delimiter	#x0A	Attribute Orientation	column	<b>Simple Text</b>		Field Delimiter	,
Number of Header Lines	1										
Record Delimiter	#x0A										
Attribute Orientation	column										
<b>Simple Text</b>											
Field Delimiter	,										
Number Of Records	12142										

2 derivations



## Data Table, Image, and Other Data Details

4 sources



**Source Program**

**Total\_PAH\_and\_Alkanes\_GoA\_Hydrocarbons\_Clean.R**

Citation

[View »](#)

This program generated the data you are currently viewing, **Total\_Aromatic\_Alkanes\_PWS.csv**.

This program used **PAH.csv**, **Sample.csv**, **Non-EVOS\_SINs.csv** and **(and 1 more )**.

**Text Format**

Number of Header Lines	1
Record Delimiter	#x0A
Attribute Orientation	column
<b>Simple Text</b>	
Field Delimiter	,

Number Of Records12142

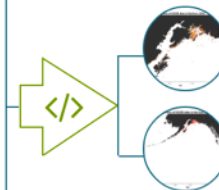
**kanes\_PWS.csv**

from PAH, Alkane and Sample tables documenting samples collected after the ... in Prince William Sound, AK

**anes\_PWS.csv**

[org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9](https://www.data.gov/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9)

2 derivations



# Web Provenance Editor

Deployed by Arctic Data Center

The screenshot displays the 'Web Provenance Editor' interface, titled 'NSF Arctic Data Center'. The header includes the Arctic Data Center logo, navigation links for 'Data', 'Support', and 'About', a 'Submit Data' button, and a user profile for 'Christopher Jones'. The main content area is titled 'Data Table, Image, and Other Data Details'. On the left, under '0 sources', there are two 'Add' buttons. The central 'Data Table' section displays details for 'Total\_Aromatic\_Alkanes\_PWS.csv', including a 'Download' button, a description of the dataset, the object name, online distribution info (a URL), size (2801033 byte), and text format (Number of Header Lines: 1, Record Delimiter: #x0A). On the right, under '0 derivations', there are two 'Add' buttons.

NSF Arctic Data Center

Data Support About Submit Data Christopher Jones

Data Table, Image, and Other Data Details

0 sources

0 derivations

**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



[Data](#)[Support](#)[About](#)[Submit Data](#)[Christopher Jones](#)

Data Table, Image, and

0 sources



Data

## 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](#)

0 derivations



Online Distribution Info

<https://crn-stage.test.dataone.org/crn/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 Delimiter

,

Number Of Records 12142



## Data Table, Image, and Other Data Details

4 sources



## 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, AKObject 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
<b>Simple Text</b>	
Field Delimeter	,

0 derivations

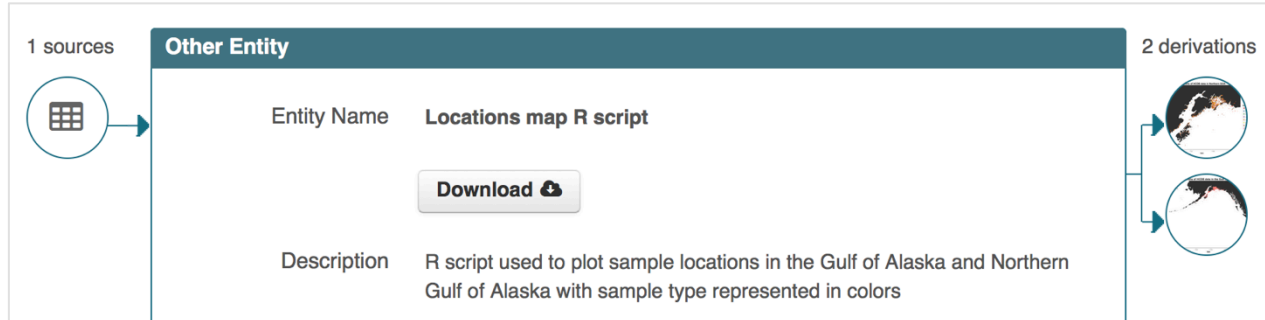
[Save](#)

# 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

