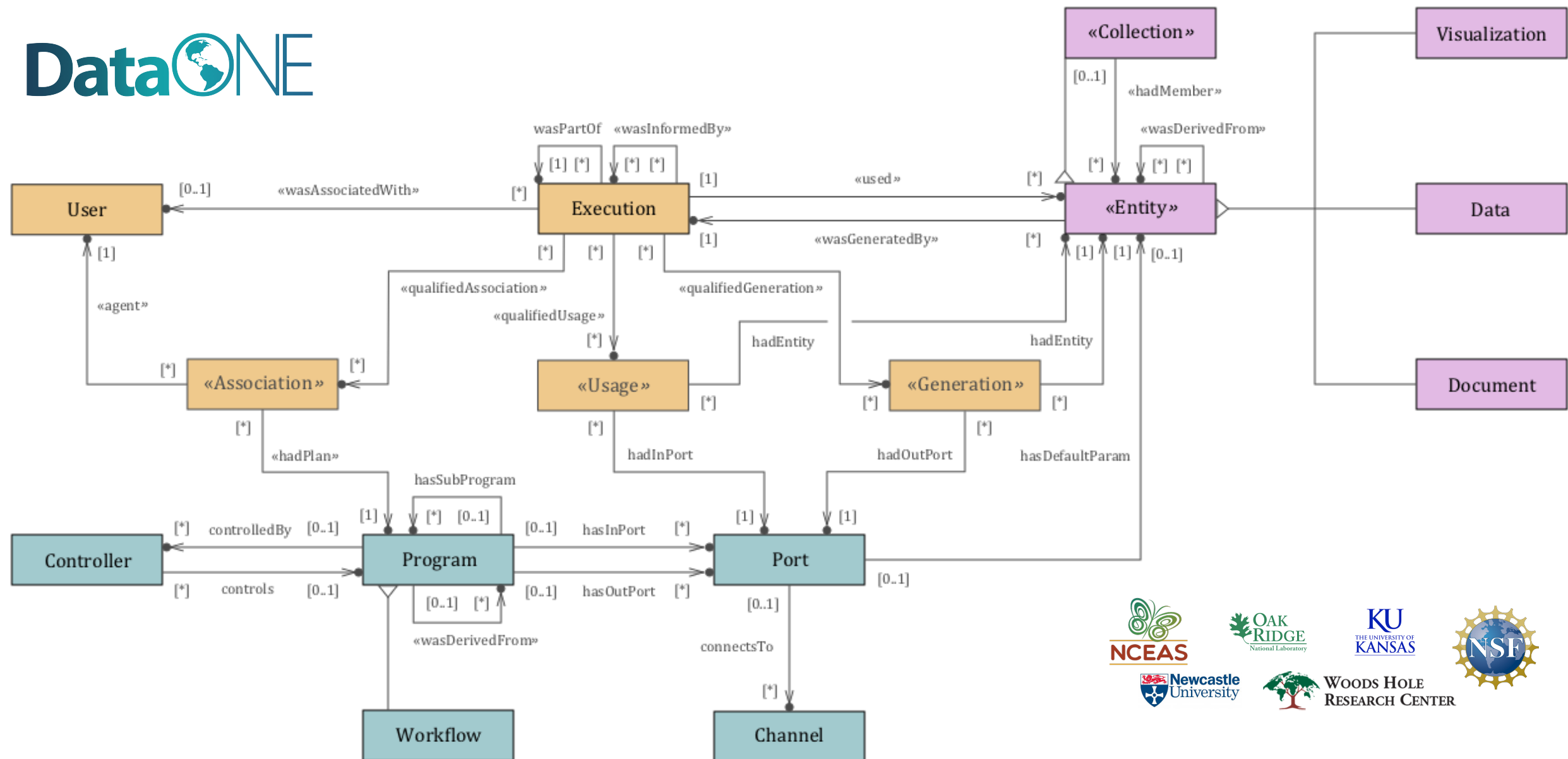


# ProvONE: extending PROV to support the DataONE community

Yang Cao<sup>1</sup>, Christopher Jones<sup>2</sup>, Víctor Cuevas-Vicenttín<sup>3</sup>, Matthew B. Jones<sup>2</sup>, Bertram Ludäscher<sup>1</sup>, Timothy McPhillips<sup>1</sup>, Paolo Missier<sup>4</sup>, Christopher Schwalm<sup>5</sup>, Peter Slaughter<sup>2</sup>, Dave Viegla<sup>6</sup>, Lauren Walker<sup>2</sup>, Yaxing Wei<sup>7</sup>

## 1. The model: (a) Process structure, (b) Data dependencies, (c) Common data structures



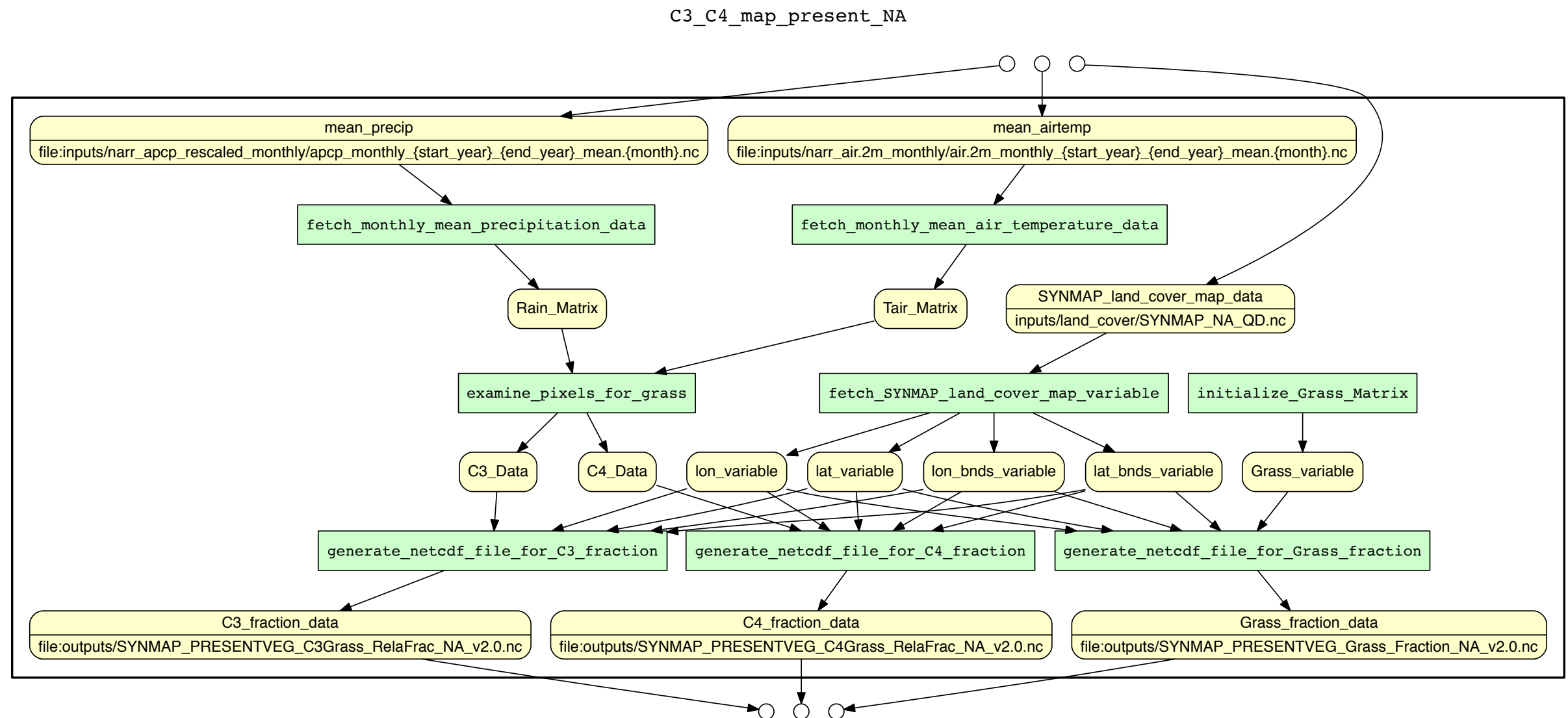
**Also:** A. Marinho, L. Murta, C. Werner, V. Braganholo, S. Serra da Cruz, E. Ogasawara, M. Mattoso. "ProvManager: A Provenance Management System for Scientific Workflows." *Concurrency and Computation: Practice and Experience* 24, no. 13 (2012): 1513–1530

<sup>1</sup>University of Illinois, Urbana-Champaign, <sup>2</sup>National Center for Ecological Analysis and Synthesis, UCSB, <sup>3</sup>Universidad Popular Autónoma del Estado de Puebla, Mexico, <sup>4</sup>School of Computing Science, Newcastle University, UK, <sup>5</sup>Woods Hole Research Center, Falmouth, MA, <sup>6</sup>University of Kansas, Lawrence, <sup>7</sup>Environmental Sciences Division, Oak Ridge National Lab, TN

2. **R and MatLab user libraries** to enable provenance recording during program execution

3. **YesWorkflow script annotation tool**

- language-agnostic
- Produces a dataflow rendering of process structure



#### 4. Visualisation of the data derivations through the script

Step-by-step user navigation of potentially complex provenance graphs.

The screenshot displays a user interface for navigating data provenance. On the left, a vertical sidebar contains 26 sources, each represented by a colored icon. A 'view more' link is visible below the sources. The main content area is divided into two panels. The top panel, titled 'Source data', shows details for a file named 'air.2m\_monthly\_2000\_2010\_mean.6.nc'. It includes a citation for Yaxing Wei (2016) and a description of the data processing. A 'View »' button is present. The bottom panel, titled '2 derivations', shows the data's lineage. It lists two derivations, each with a script icon (</>). The first derivation is labeled 'SENTVEG\_Grass\_Fraction\_NA\_v2.0.nc' and 'netCDF-3'. The second derivation is labeled 'SENTVEG\_Grass\_Fraction\_NA\_v2.0.nc' and 'netCDF-3'. A 'view more' link is also present at the bottom of the sidebar.

**Source data**

**air.2m\_monthly\_2000\_2010\_mean.6.nc**

Citation  
Yaxing Wei. 2016. MsTMIP: C3 C4 soil map processing: Run of C3\_C4\_map\_present\_NA\_with\_comments.m on 20160311T181011. MN Demo 2. metadata\_e859d2dd-c5e6-4ec6-892f-1b00bb6f8f65.xml.

[View »](#)

This data was used by `</>` C3\_C4\_map\_present\_NA\_with\_comments.m .

This data was used as an input to create C3\_C4\_map\_present\_NA\_workflow\_process\_diagram\_0.pdf, C3\_C4\_map\_present\_NA\_workflow\_data\_diagram\_0.pdf, C3\_C4\_map\_present\_NA\_workflow\_combined\_diagram\_0.pdf, SYNMAP\_PRESENTVEG\_C3Grass\_RelFrac\_NA\_v2.0.nc, SYNMAP\_PRESENTVEG\_C4Grass\_RelFrac\_NA\_v2.0.nc and the data you are currently viewing, SYNMAP\_PRESENTVEG\_Grass\_Fraction\_NA\_v2.0.nc (collapse).

[view more](#)

**2 derivations**

SENTVEG\_Grass\_Fraction\_NA\_v2.0.nc

netCDF-3

SENTVEG\_Grass\_Fraction\_NA\_v2.0.nc

netCDF-3

[view more](#)

[dbx-2.test.dataone.org/cn/v2/resolve/d87e1a6a-1a8-4-6-bba8-cb4ac2b1efb](#)

[dbx-2.test.dataone.org/cn/v2/resolve/15a312cb-83b9-44b6-b157-15a168507c38](#)