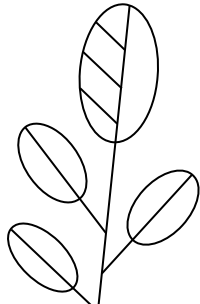# Describing Scientific Workflows with yt

Sam Walkow

Informatics PhD Student

University of Illinois
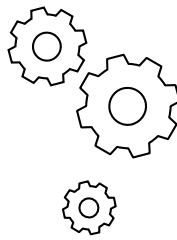
My Background

- Informatics PhD Student at the University of Illinois with a focus on **data visualization and analysis and open source software**
- Previous worked in Healthcare and Research IT
- BS and BA in Psychology and Sociology
- Email:
- Twitter: @samwalkow

# Supporting Open Source Software Development

- **How can we make open source more sustainable**
  - How do we get more people involved?
  - How do we keep people involved?
- **What are the barriers to computational work and open source work?**
  - Lowering barriers for researchers that do computational work
  - How can open source barriers be worked with?
- **Using yt as a use case**
  - An open source python package for volumetric data analysis and visualization
  - Developed for computational astrophysics
- **Cognitive Load**
  - Not in line with how users work or think
  - The Learning Cliff: Computational Workflows
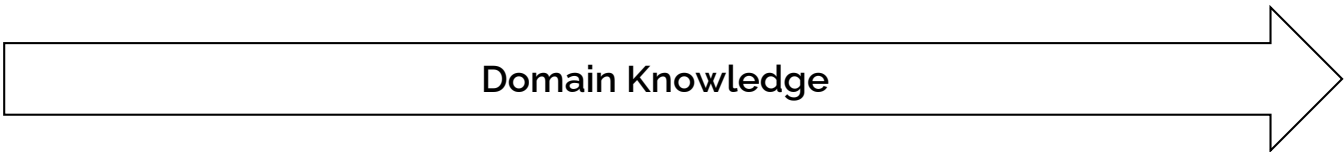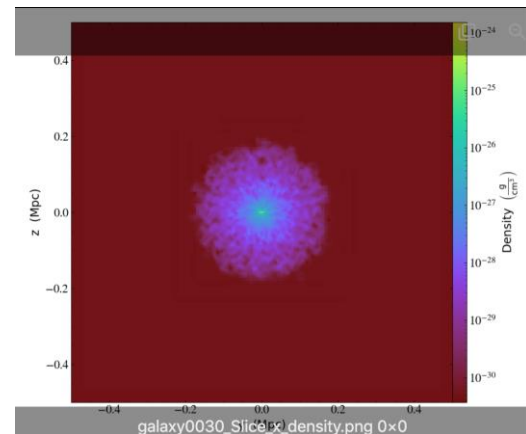
The Learning Cliff
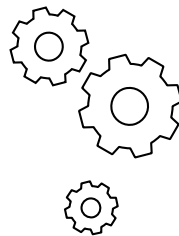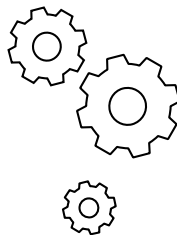
Researcher
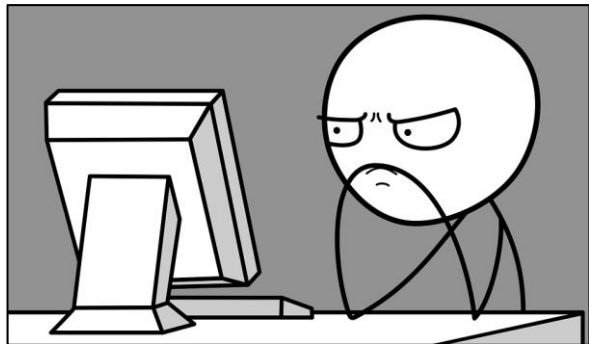
Cognitive Load

Output

Domain Knowledge

yt

Working with open source libraries to identify common methods and functions

Work towards harmonizing tools by understanding the overlap between functionality and user workflows

```
class SlicePlot(ytBaseModel)
    ds: Optional[Dataset] =
    FieldNames: FieldNames
```

```json
"SlicePlot": {
  "Axis": "x",
  "FieldNames": {
    "field": "density"
  },
  "Dataset": {
    "FileName": "./Isol
  }
}
```
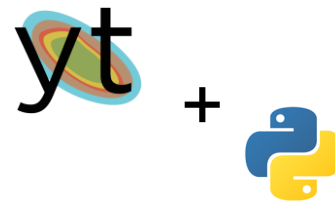
Create an interface to visualize and analyze natural science data that does not require knowledge of python or yt to use
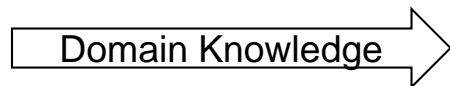
The Analysis Schema

Human + Machine Data = Analysis Schema

Data in the form of open source code bases: methods, functions, and the 'ways of thinking' encoded in them
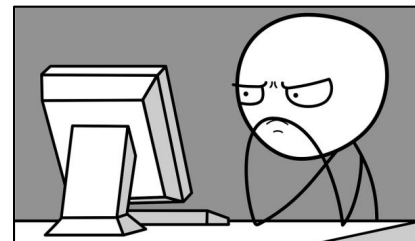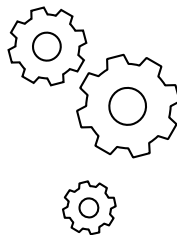
Data from users, researchers and open source communities who use and develop open source tools

Domain Knowledge

Data on the analysis schema such as navigation, accessibility, and impact on user cognitive load
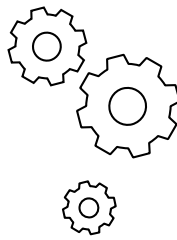
Future Directions

- **A way to add context to data analysis**
  - Data from ontologies
  - Human data from open source communities
- **A way to translate between users, developers and communities**
  - Getting more users involved who will hopefully become developers
  - Making specific programs available to more users across domains

Thank you!
swalkow2@illinois.edu

And thank you to:

Dr. Matthew Turk, Assistant Professor in School of Information Sciences at the University of Illinois

Dr. Chris Havlin, Research Scientist in School of Information Sciences at the University of Illinois