



# #1541450: CC\*DNI DIBBS: Merging Science and Cyberinfrastructure Pathways: The Whole Tale

Bertram Ludäscher<sup>1</sup>, Kyle Chard<sup>2</sup>, Niall Gaffney<sup>3</sup>, Matthew B. Jones<sup>4</sup>, Jaroslaw Nabrzyski<sup>5</sup>, Victoria Stodden<sup>1</sup>, Matthew Turk<sup>1</sup>, and Kandace Turner<sup>6</sup>

<sup>1</sup> School of Information Sciences, University of Illinois at Urbana-Champaign, <sup>2</sup> University of Chicago, <sup>3</sup> Texas Advanced Computing Center, University of Texas at Austin, <sup>4</sup> University of California Santa Barbara, <sup>5</sup> University of Notre Dame, <sup>6</sup> University of Illinois at Urbana-Champaign

## Challenges for Data Researchers

Workflow for data research is **fragmented**

- Chains of emails
- A cloud of cloud storage (Dropbox for data, Google Sheets for metadata, Evernote for notes)
- Programs are where they are (“no we cannot retire that laptop it’s the one used by my postdoc 5 years ago”)
- Results cannot be reproduced
- Software is used as a tool of discovery in nearly all research today.



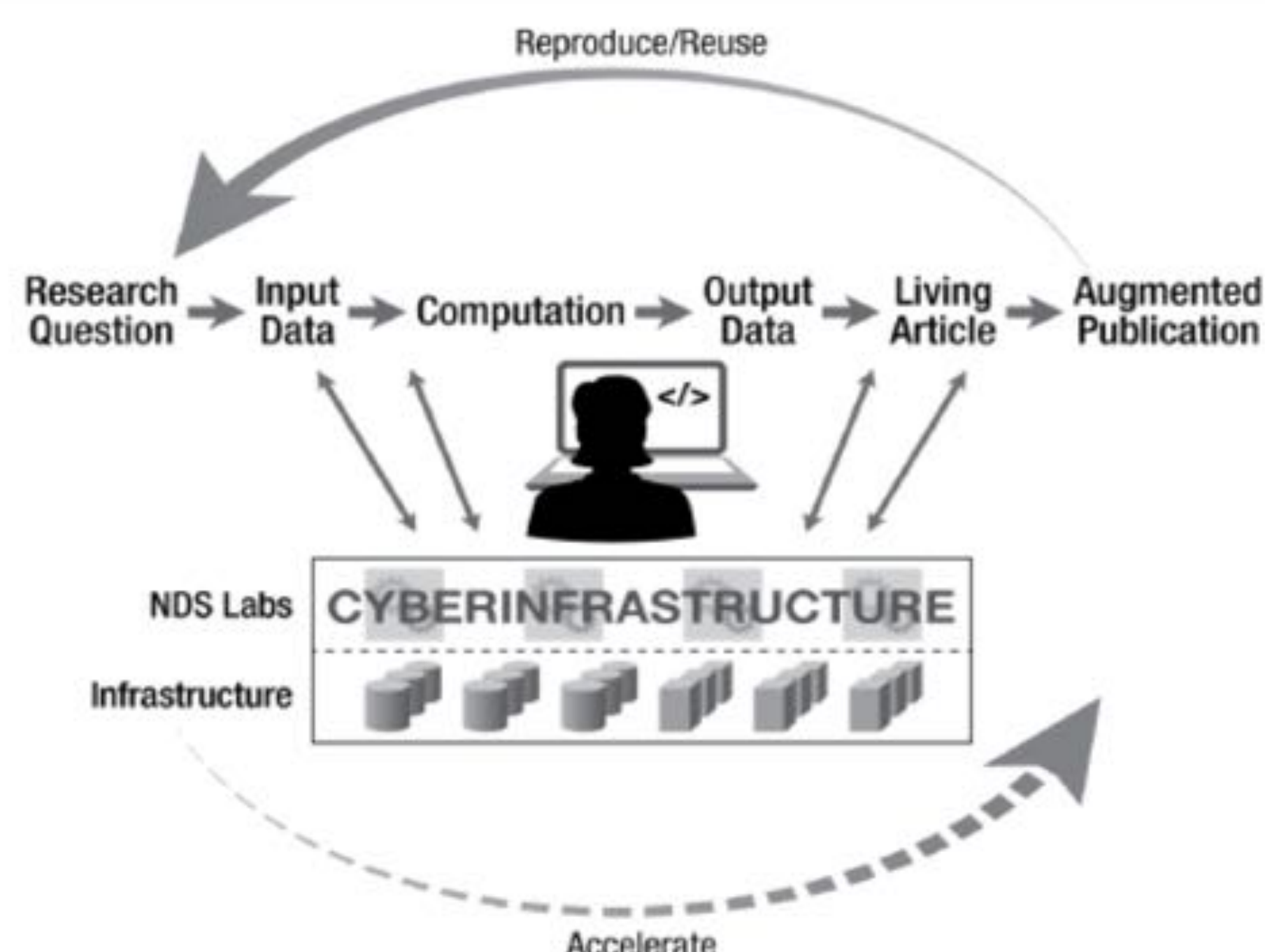
- Software is an integral and inseparable component in which most research takes place. Computational research is embedded in a social structure which includes many stakeholders.



And yet we still only preserve the endpoint and not the journey towards discovery

## Whole Tale Vision & Approach

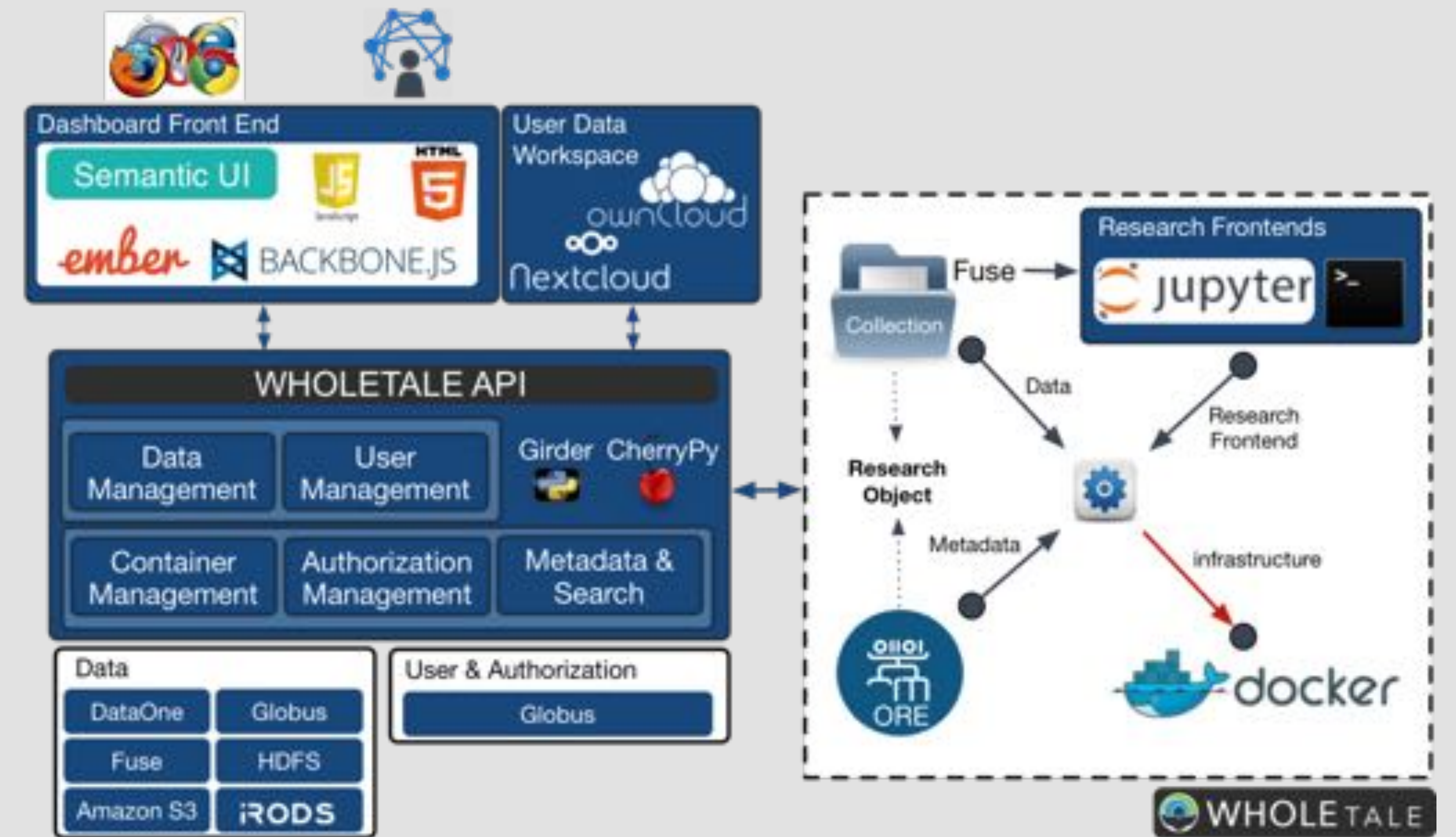
- Preserve all digital scholarly objects
  - Input, intermediary, and derived datasets
  - Software and software environment
  - Workflow process
  - Publication
- Capture computational steps / provide compute environment as they are used
- Provide unique identifiers to data/code/workflows associated with results
- Provide links to embedded in the publication for discoverability
- Integrate existing tools and technologies from other DIBBS and open source projects to create an integrated CI for data research
- Work with science working groups to build the infrastructure researchers want and need



## Whole Tale Overview

Whole Tale will leverage & contribute to **existing CI and tools** to support the **whole science story** (= *run-to-pub-cycle*), and provide access to big CI & HPC for **long tail** researchers.

➔ *Integrated tools to simplify use and promote best practices.*



## Successes and Milestones

- Selected key technologies and started integration
- Federated storage systems at TACC and NCSA
- Formed working groups and advisory board
- Started usability studies for different interfaces
- Deployed basic development systems
- Presenting concepts at key meetings looking for support and feedback from potential users
- Integrated Globus Authentication with key systems (common identity for diverse infrastructure)

## Science & CI Pathways design driven by Working Groups!

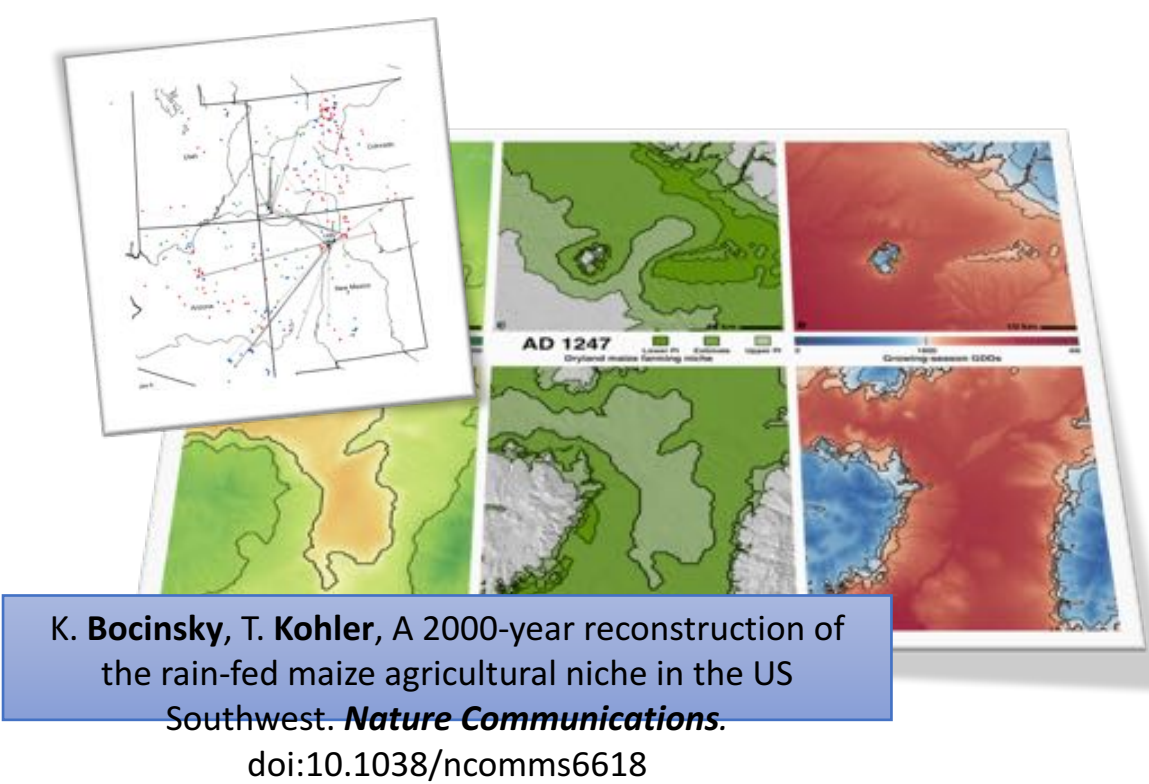


Working Groups to Provide Key Components

Working Groups Driving Use Cases and Adoption

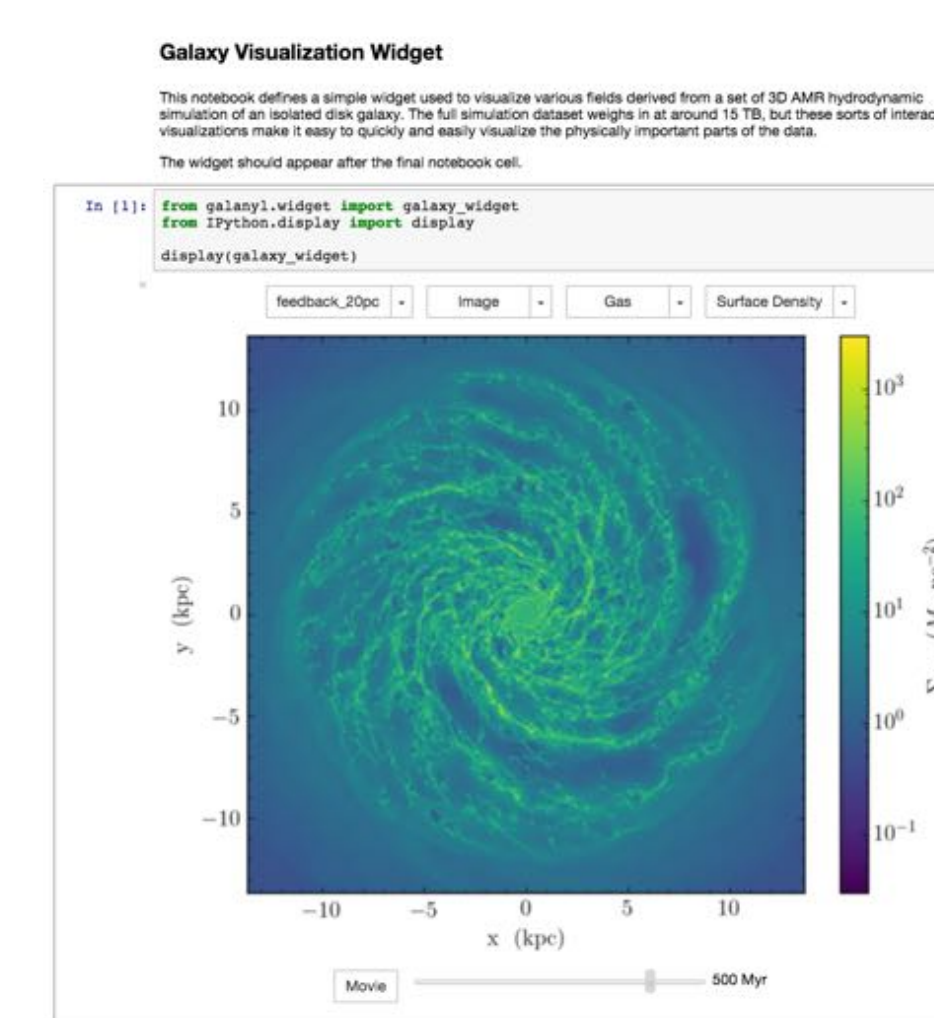
## Archaeology

Whole Tale will enable direct analysis and collaborative research on simulation outputs stored in compliant repositories via user-supplied Python scripts.



## Astronomy

Whole Tale will integrate with domain specific tools (e.g., YT) to provide advanced, customizable analysis and visualization, leveraging Jupyter to support arbitrary scripting.



## Preliminary Interface



Simple and familiar users interface integrating resources, data, collaborations, and applications.

## Further Information

[www.wholetale.org](http://www.wholetale.org)  
<https://github.com/whole-tale>

