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

PI: Bertram Ludäscher[1], **Co-Pis:** Victoria Stodden[1], Matthew Turk[1], Kyle Chard[2], Niall Gaffney[3], Matt Jones[4], Jaroslaw Nabrzyski[5]

Institutions: [1] University of Illinois at Urbana-Champaign, [2] University of Chicago, [3] University of Texas at Austin, [4] University of California, Santa Barbara , [5] University of Notre Dame

Award #: 1541450

Whole Tale: A Platform for Reproducible Research

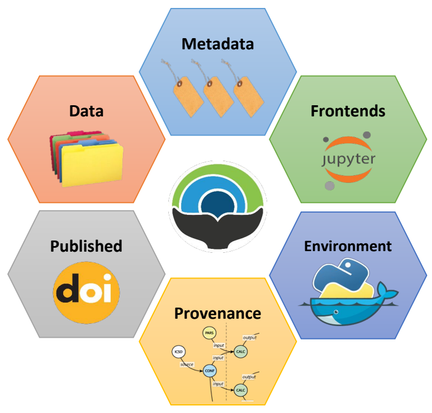
- **D**ata **I**nfrastructure **B**uilding **B**lock project (DIBBS)
- Platform to **create**, **publish**, and **execute** "**tales**"
- Simplifies the process of creating and verifying **reproducible** computational artifacts
- Integrates with **existing** research data infrastructure
- Whole **Tale** = Whole (end-to-end comp-sci) **story** for the **long tail** of science

Develop	*Analyze*	*Share*	*Package*	*Reproduce*
Easy-to-access **cloud-based** computational **environments**	**Transparent** access to research **data**	**Collaborate** and **share** with others	**Export** or **publish** executable **research objects**	**Re-execute** **Review** **Verify** **Re-use**

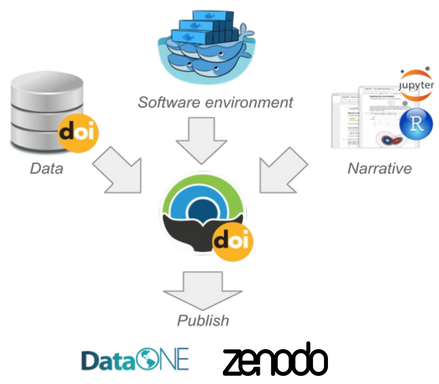
Coming soon

What's in a Tale?

- **Tale** = **executable** **research object** that captures **data**, **code** and complete **software environment** (along with a science narrative)
- Standards-based **format**: BagIt-RO, JSON-LD
- **Metadata** includes ***prospective*** (= workflow) and **retrospective** (= trace) **provenance**



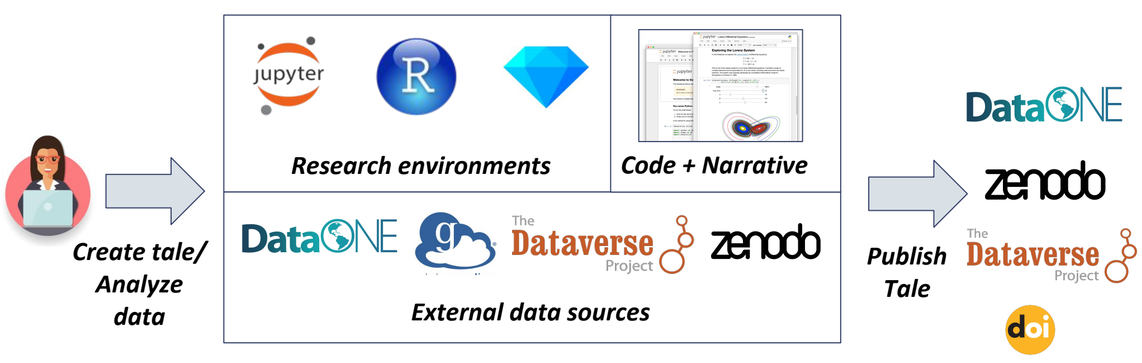
***Elements** of a Tale*



***Publishing** a Tale*

A Platform for Executable Research Objects

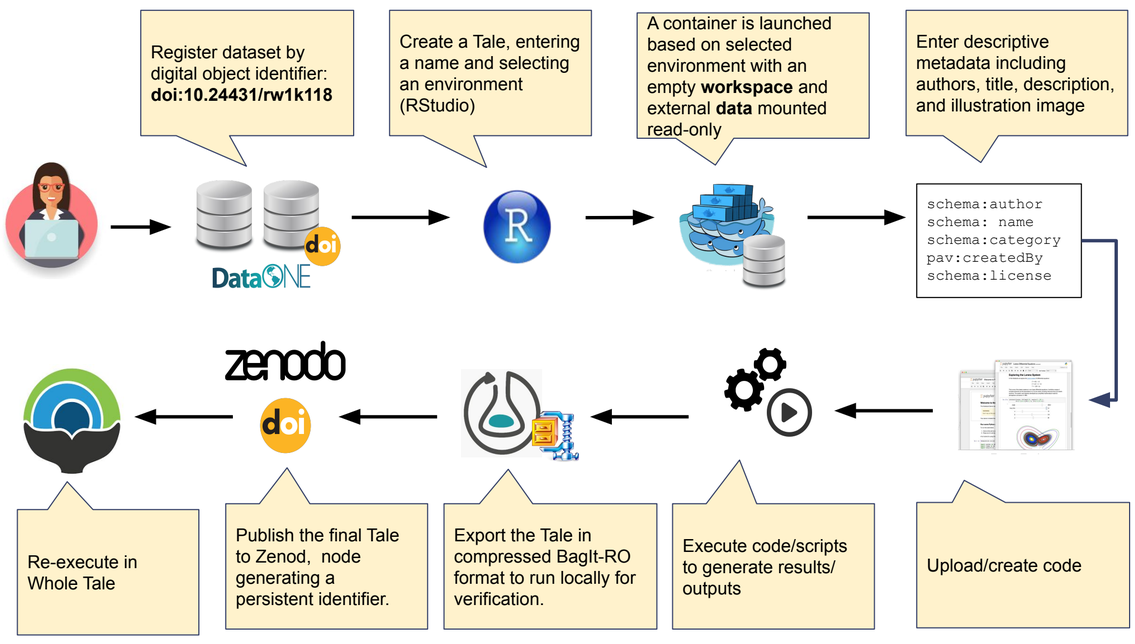
An extensible, scalable, open-source platform that **integrates** with **existing** research data infrastructure



- Researchers **authenticate** using e.g. institutional identity (Globus Auth)
- **Access** commonly-used **computational environments**
- Easily **customize** environment (via repo2docker)
- Reference and access externally **registered data**
- Create or upload **your data** and **code**
- Add **metadata** (including **provenance** information)
- Submit code, data, and environment to **archival repository**
- Get a **persistent identifier**
- **Share** for **verification** and **re-use**

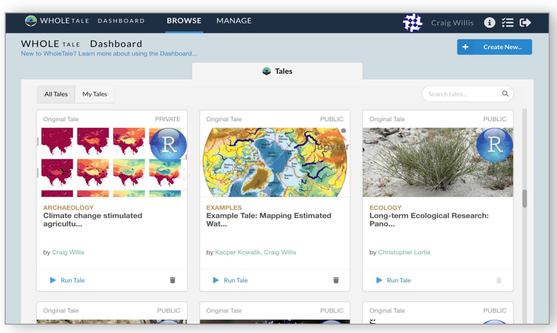
Tale Creation Workflow

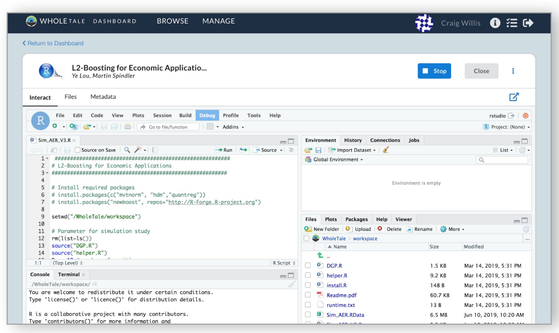
Use case: A researcher publishes a tale to **Zenodo** based on the analysis of data previously published in **DataONE** using the **RStudio** analysis environment.

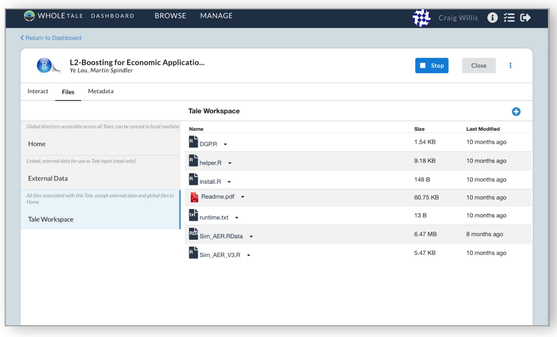


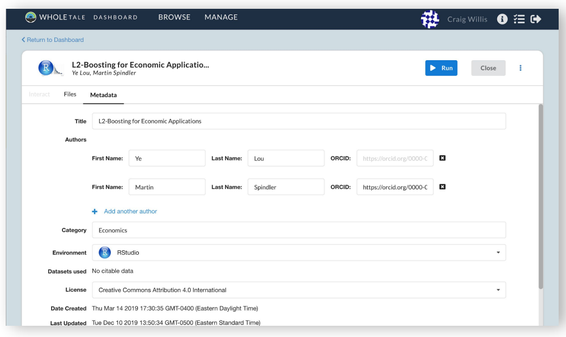
The resulting tale published to Zenodo can be downloaded and **re-executed locally** or **re-executed in the cloud** via the Whole Tale platform.

Whole Tale Dashboard Views



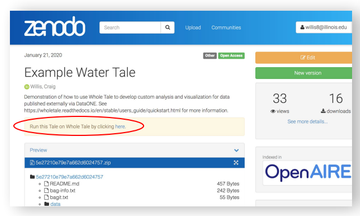


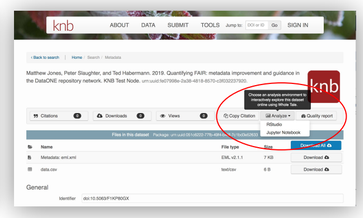


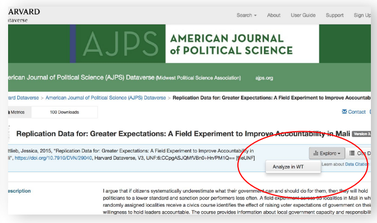


Key Features

- Cloud-based platform for interactive and exploratory data-driven and computational analysis using popular tools (Jupyter, RStudio, etc.)
- Analyze data from nearly 100 repositories in the DataONE and Dataverse networks plus Globus and Zenodo with automatic data citation.
- Publish reproducible research objects to DataONE members and Zenodo
- Create standards-based archival research artifacts that editors increasingly want for verification of computational research







Coming Soon

- Support for licensed environments including Matlab and Stata
- Automated capture of computational provenance information for improved transparency
- Integration with C²Metadata and Brown Dog (DIBBS) and PresQT (IMLS)
- Publishing to Dataverse network members

ILLINOIS · THE UNIVERSITY OF CHICAGO · UC SANTA BARBARA · TEXAS The University of Texas at Austin · UNIVERSITY OF NOTRE DAME