What is **Whole Tale**?

- NSF-funded **Data Infrastructure Building Blocks (DIBBs)** project
- **Platform** to create, publish, and execute **tales**
- Simplify process of creating & verifying **reproducible** computational artifacts
- [https://dashboard.wholetale.org](https://dashboard.wholetale.org)
Why Whole Tale?

- Increased reliance on **computation across domains**
  - new skill requirements for researchers
- **Open Science** changing norms and expectations
  - increased emphasis on **sharing data & code**
  - ... with **transparency** and **reproducibility** in mind!
  - => from sharing data to sharing **research objects**
  - **FAIR** principles
Whole Tale: Enables Computational Science
Whole Tale & the Elements of a … Reproducible Computational Research Platform

<table>
<thead>
<tr>
<th>Develop</th>
<th>Analyze</th>
<th>Share</th>
<th>Package</th>
<th>Reproduce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy-to-access cloud-based computational environments</td>
<td>Transparent access to research data</td>
<td>Collaborate and share with others</td>
<td>Export or publish executable research objects</td>
<td>Re-execute Review Verify Re-use</td>
</tr>
</tbody>
</table>

Coming soon
Whole Tale Roles and Stakeholders

- Reviewers, Curators
- Scientific Software + Data
- Analysis
- Publish & Re-use
- Verify
- Badging, Verification
- Editors, Publishers
- Repositories
- Researchers, Grad Students
Develop & Analyze with Whole Tale

- Easy to access cloud-based environments
  - Your laptop in the cloud
- Popular tools
  - + ... extensible!
- Work with data & code in transparent (provenance-enabled) ways
  - Automatic data citation
  - Automatic computational provenance capture (coming soon)
Package & Reproduce with Whole Tale

- Executable Research Objects
- Publish or export to research archives
- Compatible with new norms for reproducibility and transparency
- For verification and re-use
Whole Tale and DataONE

- **Discover & access data** from any DataONE repository
- **Analyze** data in Whole Tale
- **Package & publish** tales to Metacat-based repositories
- **Provenance** support
What exactly is (in) a Tale?

- **Tale**: Research object
  - data, code, narrative, compute environment
- Executable
- Transparent
- Publishable
- Verifiable
- Remixable
- Standards-based
Whole Tale Platform Overview

- **Authenticate** using your institutional identity
- **Access** commonly-used computational environments
- Easily **customize** your environment (via repo2docker)
- Reference and access externally **registered data**

**Research & Quantitative Computational Environments**

**External Data Sources**

- **Code + Narrative**

**Publish Tale**

- 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

"Analyze in WT" or register data by URL or digital object identifier:

Create a Tale, entering a name and selecting interactive environment:

A container is launched based on selected environment with an empty workspace and external data mounted read-only:

Customize environment adding special packages/software dependencies:

Create/upload code and scripts:

Create/upload code and scripts

Schema:author
Schema:name
Schema:category
Pav:createdBy
Schema:license

Export the Tale in compressed BagIt-RO format to run locally for verification:

Enter descriptive metadata including authors, title, description, and illustration image:

Execute code/scripts to generate results/outputs:

Re-execute in Whole Tale

Publish the tale to a supported repository, generating a persistent identifier:
Demo: Analyzing Seal Migration Patterns

A research team is preparing to publish a manuscript describing a computational model for estimating animal movement paths from telemetry data:

- **Telemetry data** published in Research Workspace
- **Analysis and visualization** in RStudio
- Existing routines stored in **local R files**
- Analysis requires **specialized R packages**
- **Publish results** for the community in **DataONE**

**Live Demo** or **Demo Video**

Key features

Supported environments

- Extension to Binder's `repo2docker`
  - Jupyter, JupyterLab
  - RStudio (based on Rocker Project)
  - OpenRefine

- Coming soon:
  - Matlab, Stata
Key features

Supported data repositories

- **Register data** from supported research data repositories
- Referenced data is **cited**
  - Ideally eventually contributing to citation counts
- **Publish tales** back to research repositories

Coming Soon:
Key features

Export to BagIt-RO

- BagIt: archival format
- Re-runnable in WT
- BagIt-RO
  - Open archival format
  - Research Object support
  - Extended for Big Data
Key features
Export and **Run Locally**

- Natural outcome of Tale **export** and **repo2docker**
- Download a zip file (BagIt-RO)
- run-local.sh
  - Build image (**repo2docker**)  
  - Fetch external data (**bdbag**)  
  - Execute (**Docker**)
Coming soon

- Publish to Zenodo, Dataverse
- Tapis/Agave data sources
- Sharing/collaboration
- Create tale from Git repository
- Image preservation
- System provenance capture
- Better user experience
Thank you! Questions?

Bertram Ludäscher
ludaesch@illinois.edu

Craig Willis
willis8@illinois.edu
References


Whole Tale Collaboration (PI Team)

- **U Illinois** (NCSA) Bertram Ludäscher, Victoria Stodden, Matt Turk
  - overall lead (co-operative agreement)
  - reproducibility; provenance; open source software development; outreach
- **U Chicago** (Globus) Kyle Chard
  - data transfer & storage; compute; infrastructure
- **UC Santa Barbara** (NCEAS) Matt Jones
  - (meta-)data publishing; provenance; repositories
- **U Texas, Austin** (TACC) Niall Gaffney
  - compute; HTC; “big tale”; Science Gateways
- **U Notre Dame** (CRC) Jarek Nabrzyski
  - UX design; UI design
The Whole Team

- Adam Brinckman (Notre Dame, former Dev)
- Bertram Ludäscher (UIUC, PI)
- Bryce Mecum (UCSB, former Dev)
- Craig Willis (UIUC, Dev, tech project manager)
- Damian Perez (Notre Dame, former Dev)
- Ian Taylor (Notre Dame, SP, Dev)
- Jarek Nabrzyski (Notre Dame, co-PI)
- Joe Stubbs (U Texas, Dev)
- Kacper Kowalik (UIUC, Dev, Senior Architect)
- Kandace Turner (UIUC, former project mgr)
- Kristina Davis (Notre Dame, UI, UX)
- Kyle Chard (U Chicago, co-PI)

- MT Campbell (UIUC, project manager)
- Matt Jones (UCSB, co-PI)
- Matt Turk (UIUC, co-PI)
- Michael Lambert (UIUC, Dev)
- Mihael Hategan (U Chicago, Dev)
- Niall Gaffney (U Texas, co-PI)
- Rachel Volentine (UTK, UX)
- Sebastian Wyngaard (Notre Dame, Dev)
- Sivakumar Kulasekaran (U Texas, former Dev)
- Thomas Thelen (UCSB, Dev)
- Timothy McPhillips (UIUC, Dev)
- Victoria Stodden (UIUC, co-PI)

+ WT Summer Interns (7); WT/RDA Fellows (4+4); WG Leads (5); other collaborators