# Reproducible Earth-surface modeling with Landlab on HydroShare

Christina J. Bandaragoda[1], Anthony Castronova[2], Jimmy Phuong[3], Ronda Strauch[1], Erkan Istanbulluoglu[1], Sai Siddhartha Nudurupati[1], David Tarboton[4], Dandong Yin[5], Shaowen Wang[5], Katherine Barnhart[6], Gregory E. Tucker[6], Eric W. H. Hutton[7], Daniel E. J. Hobley[8], Nicole M. Gasparini[9], Jordan M. Adams[9]

EarthCube All-Hands Meeting 2017
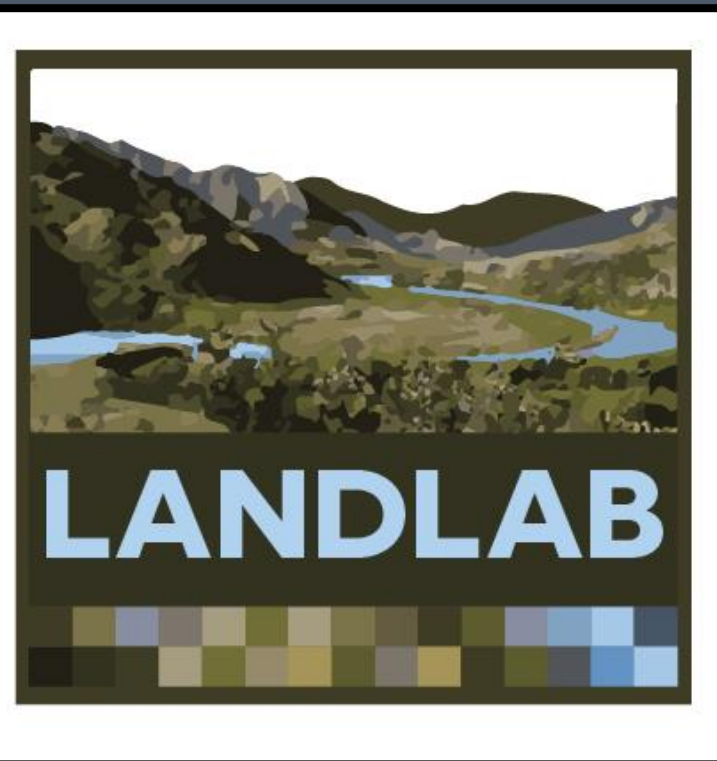
INTRODUCTION

- Modeling earth-surface processes – *complicated*
- Hardware, software, experimental design and data resources – *expensive*
- Community confidence and trust – *hard to earn*

In Panel 1, we show the typical model development from an individual PC. Panel 2 gives a conceptual diagram of how the Landlab, HydroShare, CyberGIS collaboration built an interoperable system to support reproducibility of models developed in the Landlab modeling toolkit. In Panel 3, we give a detailed example from a landslide modeling research study in the North Cascades National Park (NOCA) to show how the data architecture and cyberinfrastructure found in HydroShare, powered by supercomputing resources, enables collaboration and modeling developments. Other examples are listed in the Support Your Observatory section below.

1. TYPICAL PC MODELING WORKFLOW
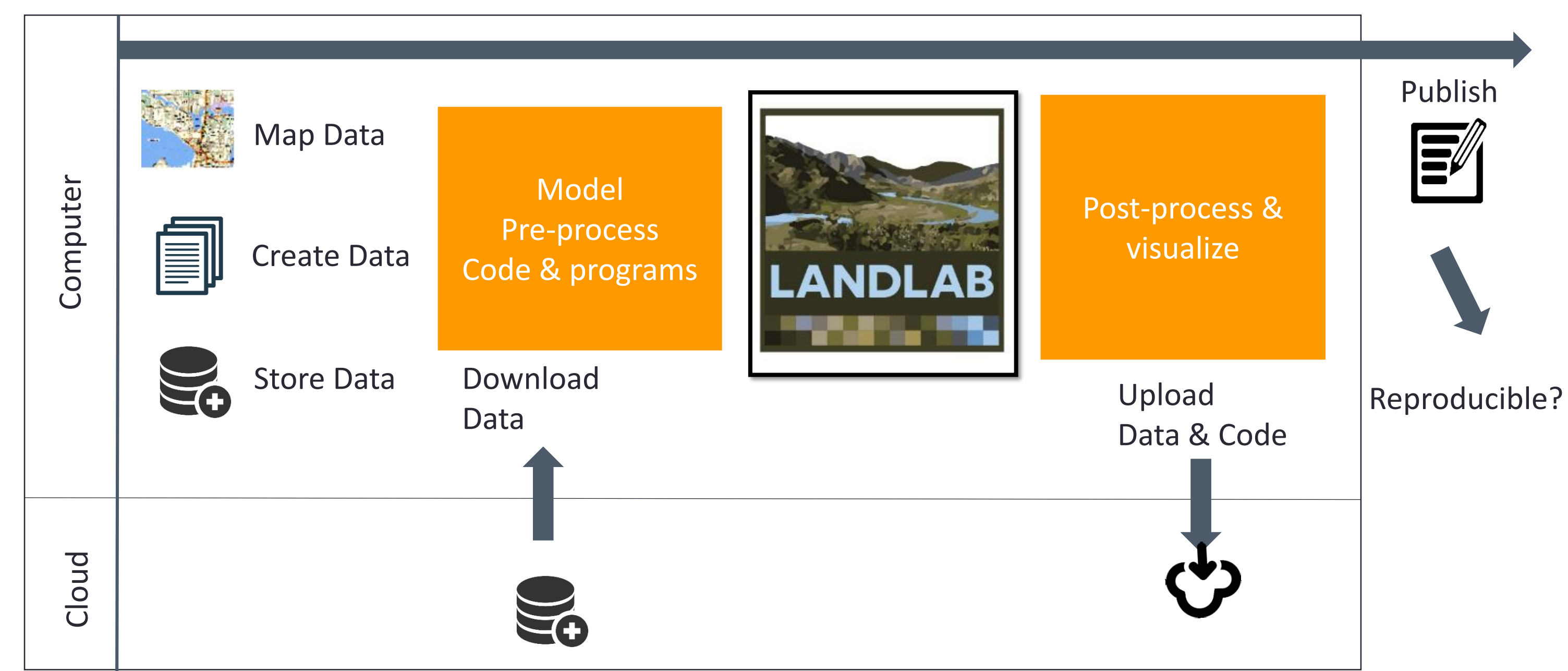
Developing models from a personal computer requires installing a suite of specialized software include tools and hardware to map, store, and prepare model inputs. Here we consider "Cloud" to be a remote servers accessible to a user over the internet that holds data. These servers may be designed to perform computations, and now with HydroShare, this function is now accessible for hydrologic modeling. The Landlab modeling toolkit uses Python scripts driven by Jupyter Notebooks to perform much of the pre and post processing steps. The Landlab components are generally called using command line terminal. Model experiments (data and code) and uploaded along with publications. The ability to reproduce a study often depends on the ability to reproduce the software environment used to develop the research.
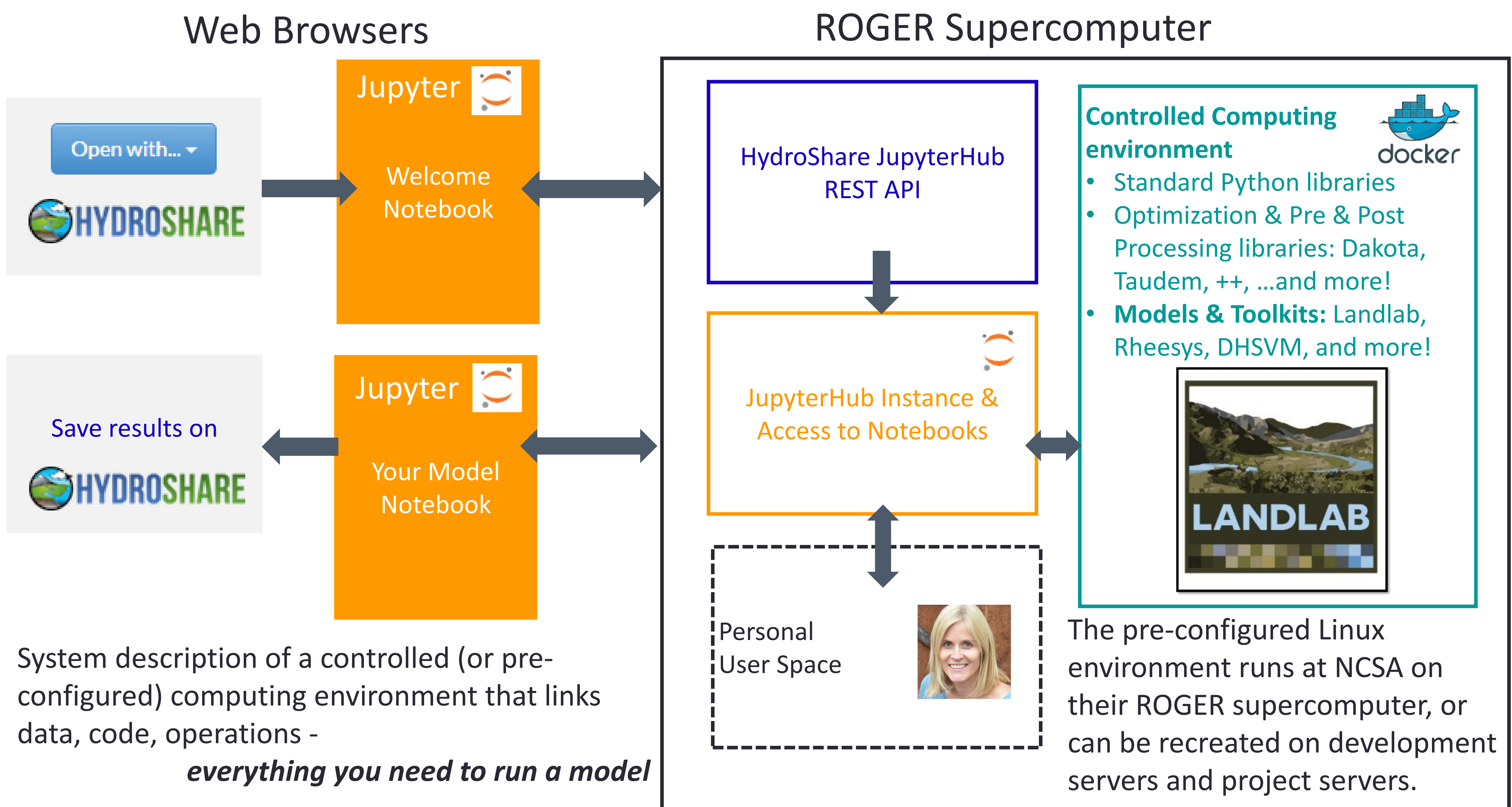
SUPPORT YOUR OBSERVATORY

Use Observatory as a label in HydroShare resources and Jupyter Notebooks to identify that the research is 1) intended to communicate the State of Knowledge about a place, 2) developed with the intent for others to reproduce and advance, and 3) accessible for scientific research, management, and public audiences to build shared knowledge. These are examples:

- Publish a reproducible paper describing a Landlab landsliding model (see right).
- Contribute to studies of the Sevilleta National Wildlife Refuge with a Landlab Ecohydrology model, shared with the Long Term Ecological Research (LTER) Program and the public.
- Develop a Landlab sediment model for the Skagit watershed that couples public data and models with a hydrology model, a stream temperature model, and is shared with researchers and a broad consortium of stakeholders.

SUMMARY OF FEATURES

- Tools to make it faster and easier to develop new physical process components - *Landlab*
- Interoperable process components combine to create new integrated models - *Landlab*
- Cyber infrastructure with collaboration, sharing and privacy settings, Creative Commons license options, and DOI publishing - *HydroShare*
- Cloud access with high-speed processing from the CUAHSI CyberGIS HydroShare JupyterHub server at the National Center for Supercomputing Applications - *CyberGIS*

2. ARCHITECTURE FOR REPRODUCIBLE COMPUTING

Web Browsers

ROGER Supercomputer

Controlled Computing environment
- Standard Python libraries
- Optimization & Pre & Post Processing libraries: Dakota, Taudem, ++, ...and more!
- **Models & Toolkits:** Landlab, Rheesys, DHSVM, and more!

System description of a controlled (or pre-configured) computing environment that links data, code, operations -

everything you need to run a model

The pre-configured Linux environment runs at NCSA on their ROGER supercomputer, or can be recreated on development servers and project servers.

3. COLLABORATIVE CLOUD APPROACH: e.g., REPRODUCIBLE LANDSLIDE MODEL

Develop a reproducible model: Landlab was used to study landslides in the North Cascades National Park. The work is available on HydroShare and these are the researcher's steps to develop and publish data and code so that model results can be reproduced by reviewers and readers of Strauch et al., (2017).
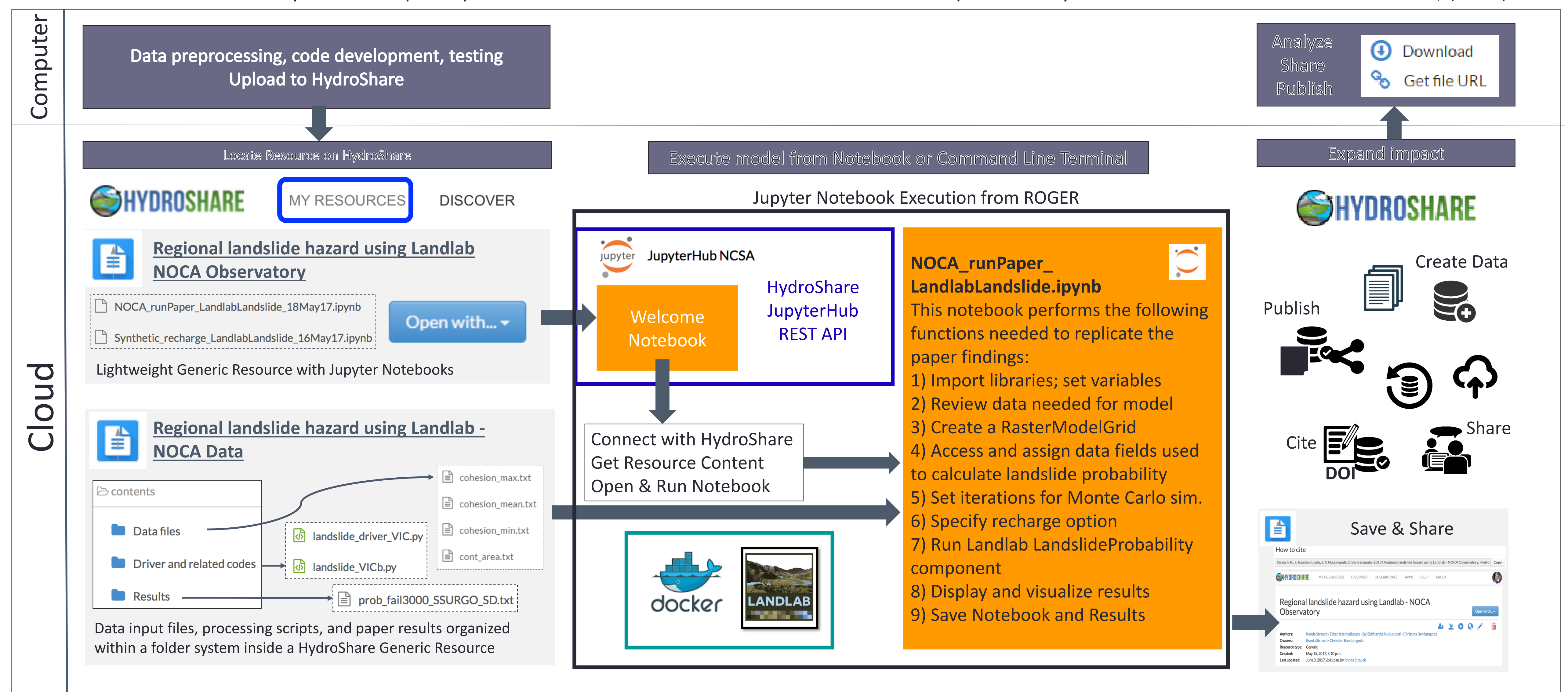
Reproduce and build on model results: Explore the data and model from a web browser. Create a personal version of the Notebook and model instance, and then Collaborate by adding new research to the NOCA Observatory resource shared with modelers in the Landlab Group in HydroShare.
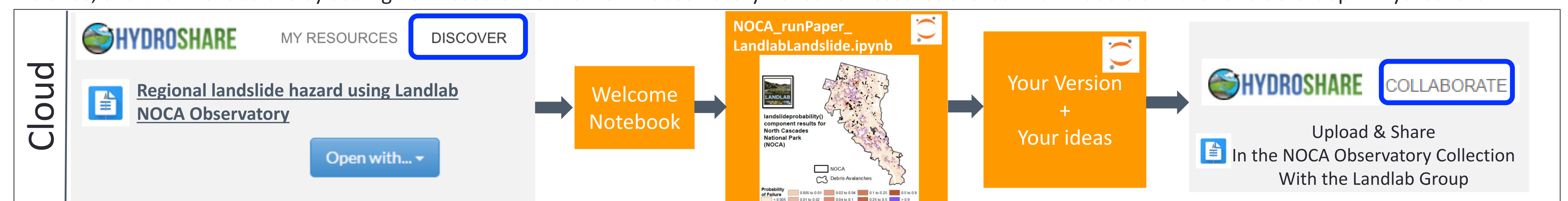
Contact: cband@uw.edu

Learn more about Landlab:

http://Landlab.github.io

1 Department of Civil and Environmental Engineering, University of Washington, Seattle, USA; 2 Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), USA; 3 Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, USA; 4 Department of Civil & Environmental Engineering, Utah State University, Logan, USA; 5 National Center for Supercomputing Applications (NCSA), University of Illinois, Urbana-Champagne, USA; 6 Department of Geological Sciences, University of Colorado, Boulder, USA; 7 Community Surface Dynamics Modeling System (CSDMS), University of Colorado, Boulder, USA; 8 Cardiff University, Cardiff, UK; 9 Department of Earth and Environmental Sciences, Tulane University, New Orleans, LA, USA.