Infrastructure for Lowering the Barrier to Computational Modeling of the Earth Surface: Run Landlab on HydroShare

Infrastructure Team: Christina J. Bandaragoda[1], Anthony Castronova[2], Jimmy Phuong[3], Erkan Istanbulluoglu[1], Ronda Strauch[1], Sai Siddhartha Nudurupati[1],
HydroShare-Cyber GIS Team: David Tarboton[4], Dandong Yin[5], Shaowen Wang[5],
Landlab Team: 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]

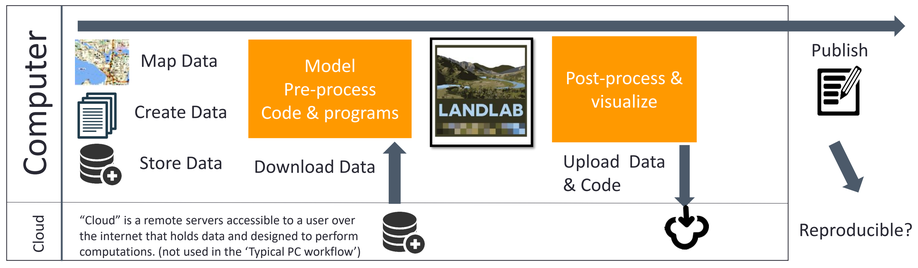
CUAHSI Hydroinformatics 2017

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 to compare to a reproducible model workflow in Panel 2. Panel 3 gives a conceptual diagram of how the Landlab [1], HydroShare, CyberGIS collaboration built an interoperable system to support reproducibility of models developed in the Landlab modeling toolkit. In Panel 4, 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

1. TYPICAL PC MODELING WORKFLOW



Reproducibility Barriers:

Collaborative Research: All team members (scientists, interdisciplinary experts, stakeholders) need to be able to access and interact with the model, not just the code developer/modeler.

Education & Curriculum: Training the newest scientists should focus on teaching science, and not require extensive computer science knowledge to make use of the latest tools.

Technology Practices for Publication: Evolving software versions, hardware requirements, computer science methods, and code quality limit the ability to replicate, reuse model applications.

Lowering the Barriers: Developing models from a personal computer (PC) requires installing a suite of specialized software tools and access to computational hardware to map, store, and prepare model inputs--reproducing a study often depends on the ability to reproduce the software environment. The Landlab modeling toolkit uses Python scripts driven by Jupyter Notebooks to perform processing steps, with expert users generally using the command line terminal. With model experiments (data and code) uploaded along with publications, our workflow (Panel 2) makes it easier to replicate, reuse, and build on models to advance science.

INFRASTRUCTURE FOR HYDROLOGY++

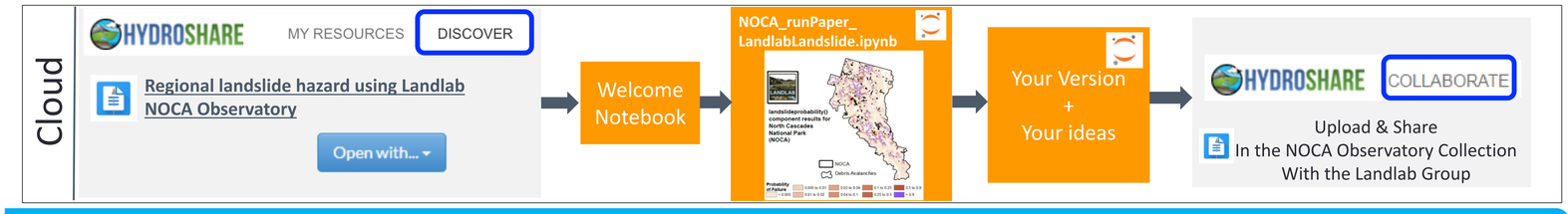
Knowledge infrastructure is an emerging intellectual framework to understand how people are creating, sharing and distributing knowledge -- which has dramatically changed and is continually transformed by Internet technologies. Knowledge infrastructures are most simply defined as "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" [2;3]. Infrastructures are not engineered or fully coherent processes. Rather, they are best understood as ecologies or complex adaptive systems. Here, we consider cyberinfrastructure "as an internet based system of tools that has evolved as a technological and sociological solution to the problem of efficiently connecting research laboratories, data, computers, and people with the goal of enabling derivation of novel scientific theories and knowledge [4]. Hydrologic and other earth surface modelers using Landlab + HydroShare + CyberGIS (Panels 2-4) are contributing to a shared knowledge infrastructure when they develop and couple models, publish data and code, and share their research through this system.

SUMMARY OF FEATURES

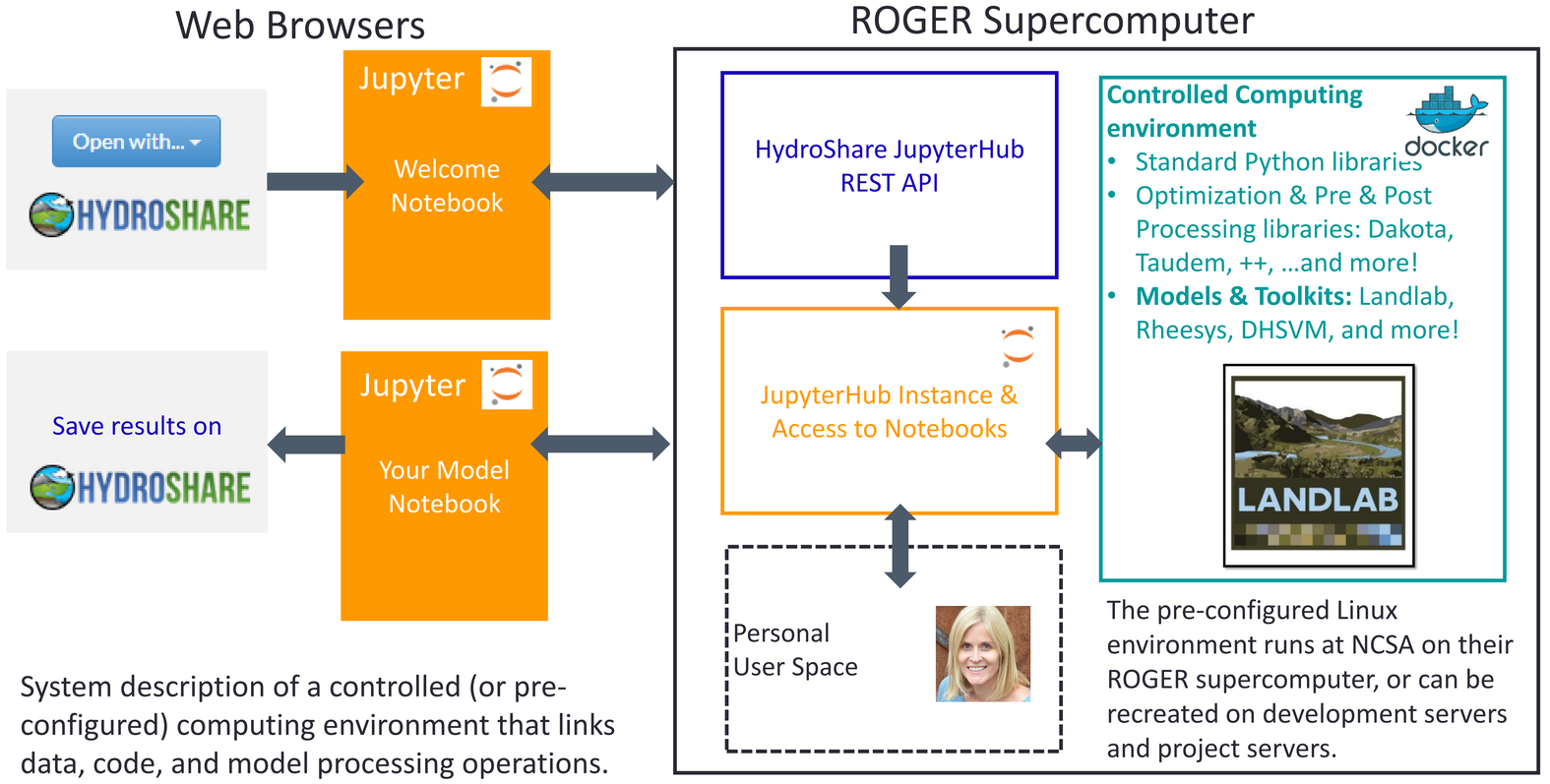
- Tools to make it faster and easier to develop new physical process components - ***Landlab***
- Interoperable process components combine to create new integrated models - ***Landlab***
- Infrastructure for collaboration, sharing and privacy settings, and DOI publishing - ***HydroShare***
- Cloud access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

2. REPRODUCIBLE MODEL WORKFLOW

Replicate, reuse, 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 Collection resource shared with modelers in the Landlab Group in HydroShare.



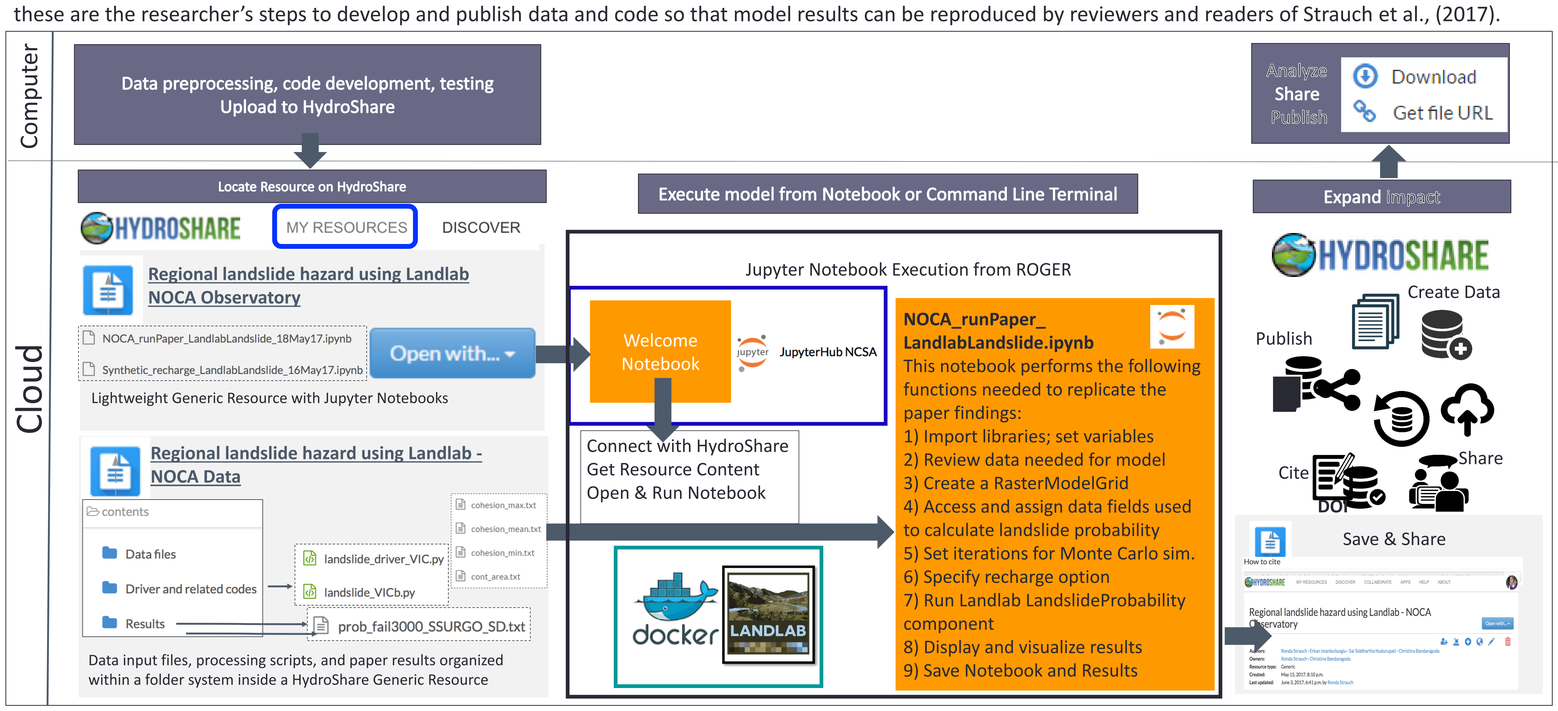
3. ARCHITECTURE FOR REPRODUCIBLE COMPUTING



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

4. 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).



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

References
1. Data-Driven Hobley, D.E.J., Adams, J.M., Nudurupati, S.S., Hutton, E.W.H, Gasparini, N.M., Istanbulluoglu, E., and Tucker, G.E. **Creative computing with Landlab: an open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics.** *Earth Surface Dynamics*, 2017, doi:10.5194/esurf-5-21-2017
2. Borgman, C.L. and S. Traweek (2012), The Transformation of Knowledge, Culture, and Practice in Science: A Knowledge Infrastructures Perspective.
3. Edwards, P.N. (2010), A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming. The MIT Press, Cambridge, MA.
4. Newman, H.B., Ellisman, M.H., and J.A. Orcutt, (2003), Data-Intensive e-Science Frontier Research in the Coming Decade". Communications. Association for Computing Machinery. **46** (11).
5. Strauch, R., Istanbulluoglu, E., Nudurupati, S. S., Bandaragoda, C., Gasparini, N. M., and Tucker, G. E.: A hydro-climatological approach to predicting