HydroShare: A Platform for Collaborative Data and Model Sharing in Hydrology

David G Tarboton¹, Ray Idaszak², Jeffery S Horsburgh¹, Daniel P Ames³, Jonathan L Goodall⁴, Alva Couch⁵, Pabitra Dash¹, Michael Stealey², Hong Yi², Tian Gan¹, Tanu Malik⁶, Christina Bandaragoda⁷, HydroShare Development Team⁸

¹Utah State University, ²RENCI, University of North Carolina at Chapel Hill, ³Brigham Young University, ⁴University of Virginia, ⁵Tufts University, ⁶DePaul University, ⁷University of Washington, ⁸https://www.hydroshare.org/group/2

HydroShare is an online, collaboration system for sharing hydrologic data, analytical tools, and models. It supports the sharing of and collaboration around “resources” which are defined by standardized content types for data formats and models commonly used in hydrology. Currently, with HydroShare you can: share your data and models with colleagues; manage who has access to the content that you share; share, access, visualize, and manipulate a broad set of hydrologic data types and models; publish data and models and obtain a citable digital object identifier (DOI); aggregate your resources into collections; discover and access data and models published by others; use the web services application programming interface (API) to programmatically access resources; and use integrated web applications to visualize, analyze and run models on data in HydroShare. Composite resources allow multiple file types from a study to be combined together, providing, as a single resource, an aggregation of all the data elements associated with a model or study. Hydroshare’s composite resource construct can be used to support software that enables transparency and reproducibility, and thereby enhance trust in the research findings.

Toward this, as part of the EarthCube GeoTrust project we are investigating how the composite resource construct can be extended to support transparency and reproducibility. The EarthCube GeoTrust project is creating “geounits” which are self-contained packages of computational experiments that can be guaranteed to repeat or reproduce regardless of deployment issues. Since geounits provide a complete description of all the data elements with an instance (run) of a computational experiment, including input files, parameter files, the model executable, associated libraries, and output files produced, they can be mapped to a specialization of HydroShare’s composite resource type. This has a direct effect of transforming HydroShare into a repository of geounits, and making published and cited experiments not only accessible but also reproducible, thereby enhancing trust in them. Tools that create geounits use HydroShare’s REST API to load them into HydroShare, where they can then be shared with other users and downloaded for reproduction of the computational experiment, or further research with additional or alternate data. This presentation will describe the functionality and architecture of HydroShare that enables the creation of geounits comprising: (1) resource storage, (2) resource exploration, and (3) actions on resources by web applications. HydroShare’s components are loosely coupled and interact through APIs, which enhances robustness, as components can be upgraded and advanced relatively independently. The full power of this paradigm is the extensibility it supports, in that anybody can develop a web application that interacts with resources stored in HydroShare. We welcome discussion of the opportunities this enables for interoperability with other EarthCube tools, to the benefit of the geoscience research community.