Today, most software developed by scientists is never shared.

- There are repositories of model software (e.g., CSDMS), also code repositories (e.g., GitHub).
- However, software is rarely shared, particularly all the data pre-processing and visualization software.

**Problem:** Hard to build on other work, especially across disciplines

- Software captures valuable geoscience knowledge and should be shared
- “Scientists and engineers spend more than 60% of their time just preparing the data for model input or data-model comparison” (NASA A40)

### GeoSoft Project Goals

**An on-line community for sharing knowledge about geosciences software**

- **Intelligent assistance** to publish and describe new software: how to use it appropriately, what kinds of data, how it relates to other software, propagating provenance
- **Recommender system** and sophisticated search capabilities to find software that fits science needs
- **Interactive advice** on open source software, forming successful developer communities, and other software sharing topics

### Novel research contributions

- An intelligent user interface that organizes the interaction in terms of science tasks (e.g., understand assumptions, do research with the software, cite the software, etc)
- An ontology to represent metadata for scientific software (http://www.ontosoft.org)
- Automated crawlers that extract metadata for the software from the user’s web site (e.g., extracting license and other information from GitHub)
- Training modules to help scientists learn to share and describe software

### Open science

- Science products, including software, are of general societal importance [Holdren’s OSTP memo, 2013] [NSF 2015]
- Reproducible science requires open software
- Open source software development is more sustainable
- Scientists want to get credit for software much more than for data [Nature Metrics survey, 2010]

### Benefits

- **Open science:** Easy to disseminate models and software across disciplines
- **Accelerate research:** By reusing software
- **Reproducibility:** Easy to replicate results
- **Accessibility:** For non-programmers
- **Quality:** Best software that is well tested
- **Integration:** Well-described software is easier to integrate