With growing and diverse collections of data becoming part of modern scientific workflows, many research projects today begin with a process of data wrangling, i.e. finding, manipulating, indexing, cleaning, and bringing together fragmented datasets. Brown Dog aims to alleviate much of the overhead and heterogeneity in the processes involved in this step which tends to otherwise hinder scientific progress and reproducibility. Through a REST API Brown Dog provides data transformations such as format conversions and content based extractions as a service which supports diverse usage through various clients and programming languages. Further, Brown Dog provides a venue to access and preserve data transformation tools, track provenance, track information loss, manage data movement, and process jobs in a scalable manner across a diverse set of computational resources. Overall, Brown Dog provides a low-level data infrastructure to interface with digital data contents and through its capabilities move software to being more agnostic to the format/structure of data, enabling the scientific community to focus more on their research, less on data wrangling, and allow researchers to more easily access datasets that would otherwise be inaccessible.

Innovations and Capabilities

- Distributed Data Transformation Service (DTS), analogous to a DNS for applications to build upon
- Publication and reuse of data wrangling tools across science

82 Tools in beta release spanning Ecology, Biology, Hydrology, Civil Engineering, Social Science, as well as general users.

http://browndog.ncsa.illinois.edu