

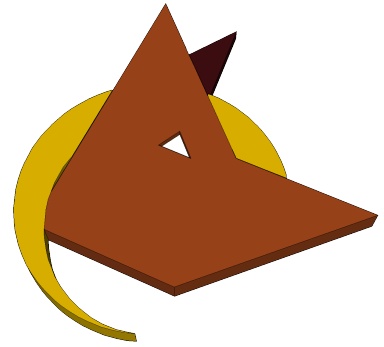
DIBBs: Brown Dog

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 needed datasets. DIBBs Brown Dog aims to alleviate some 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 (leveraging [Polyglot](#)) and content based extractions (leveraging [Clowder](#)) 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.

- **Documentation**
 - [Introduction](#)
 - [BDFiddle](#)
 - [BD CLI](#)
 - [BD Windows Client](#)
 - [API](#)
 - [Preparing Tools](#)
 - [Adding Tools](#)
 - [Project Supported Transformations](#)
- **Workshops & Tutorials**
 - [Brown Dog Tutorial - Beta Release](#)
 - [NCHC Tutorial 2017](#)
 - [XSEDE Tutorial 2016](#)
 - [Early User Workshop 2015](#)
 - [Kickoff Tutorials 2013 and 2014](#)
- **Project Information and Resources**
 - [Components & Source Code](#)
 - [Data Management Plans](#)
 - [Description](#)
 - [Definitions](#)
 - [Events, Outreach, and Media](#)
 - [Project Team](#)
 - [Publications](#)
 - [Resources](#)
- *Discussions / Under Development*

Blog Posts

- Blog: [NCSA Brown Dog and Box Skills Speed up Astronomical Research](#) created by [Shannon Bradley](#)
Apr 12, 2019
[BrownDog](#)
- Blog: [The Predictive Ecosystem Analyzer - PEcAn](#) created by [Kenton McHenry](#)
Mar 21, 2018
[BrownDog](#)
- Blog: [Using Machine Learning to Understand Public Preference Toward Landscape Design](#) created by [Kenton McHenry](#)
Feb 21, 2018
[BrownDog](#)
- Blog: [Brown Dog Tutorials](#) created by [Kenton McHenry](#)
Feb 01, 2018
[BrownDog](#)
- Blog: [What is Brown Dog Video](#) created by [Kenton McHenry](#)
Nov 06, 2015
[BrownDog](#)
- Blog: [IEEE Big Data](#) created by [Kenton McHenry](#)
Nov 06, 2015
[BrownDog](#)
- Blog: [Brown Dog Cheat Sheet](#) created by [Kenton McHenry](#)
Sep 16, 2015
[BrownDog](#)
- Blog: [To Be Heard and Not Seen](#) created by [Kenton McHenry](#)



<https://bd-api.ncsa.illinois.edu/>

Recent Space Activity



[Kenton McHenry](#)
[Events, Outreach, and Media](#) updated Feb 19, 2020 [view change](#)
[Publications](#) updated Feb 19, 2020 [view change](#)



[Mark Fredricksen](#)
[Transformations Catalog](#) updated Jan 27, 2020 [view change](#)
[Setting up Windows VM for ArcGIS extractors](#) updated Jan 24, 2020 [view change](#)



[Sandeep Puthanveetil Satheesan](#)
[Setting up Windows VM for ArcGIS extractors](#) updated Jan 09, 2020 [view change](#)

Brown Dog Flyer / Cheat Sheet

Apr 27, 2015

[BrownDog](#)

- Blog: [Critical Zone](#) created by [Kenton McHenry](#)

Apr 25, 2015

[BrownDog](#)

- Blog: [PEcAn](#) created by [Kenton McHenry](#)

Apr 25, 2015

[BrownDog](#)

- Blog: [Green Infrastructure](#) created by [Kenton McHenry](#)

Apr 25, 2015

[BrownDog](#)

- Blog: [Brown Dog Clients: A Command Line Interface](#) created by [Kenton McHenry](#)

Mar 18, 2015

[BrownDog](#)

- Blog: [The Brown Dog Tools Catalog](#) created by [Kenton McHenry](#)

Mar 05, 2015

[BrownDog](#)

- Blog: [Ecological Model Data Conversions: PEcAn](#) created by [Kenton McHenry](#)

Mar 03, 2015

[BrownDog](#)

- Blog: [Brown Dog Clients: Chrome Extension](#) created by [Kenton McHenry](#)

Mar 03, 2015

[BrownDog](#)



BROWN DOG

browndog.ncsa.illinois.edu

The persistent growth of "big data" presents challenges that include not only large quantities but also large varieties of digital data.

Brown Dog is a highly extensible set of services for:

- data format conversions
- metadata extraction from data contents
- indexing associated collections of data
- preserving & using data analysis/manipulation tools

Result of the data generated by sensors, social activities, and the humanities is smaller, unstructured, un-coupled and thus not easily stored. The vast quantity of "big data" data both past and present has the potential for great impact on future research in many disciplines.

In addition, science relies on digital data and software that evolve and disappear quickly as technology changes. We are entering a period when scientific results are no longer easily reproducible. Capturing legacy data and software is essential to maintaining the validity of large bodies of research. Without a consistent and uniform tool set for all the data, or at least associated metadata, such actions become prohibitively difficult.

Brown Dog addresses these and similar considerations through the development of a set of services for data format conversion, metadata extraction from existing data, and indexing of associated collections of data. Brown Dog provides the framework for an extensible suite of new and existing tools. Researchers using the extensible Brown Dog system will be able to apply the whole suite of tools to data collections in order to find and "convert" previously inaccessible data.

The Data Access Proxy (DAP) and the Data Tiling Service (DTS), focusing on format conversion and content based annotation/analysis, respectively, help relevant conversion and extraction operations within arbitrary software, manages their employment in an elastic manner and manages job execution from behind a deliberately compact REST API.

A number of client libraries and applications are being constructed to further reduce the overhead of accessing the provided functionality. Data Streams in Javascript, Python, R, MATLAB and C++/Fortran/C are available, provide wrappers to other database applications. At the heart of the tool services is their extensibility, which allows researchers to create extensions, or other database applications, as a conversion or extraction tool. This leverages the functionality they provide and promotes the tool's interoperability. Both services support a variety of scripting languages to wrap tools for inclusion in Brown Dog (e.g. Python, MATLAB, R, Java, JavaScript, etc.).

Brown Dog aims to support data conversion and extraction/analysis results from a broad range of conversions. Current efforts focus on biology, ecology, civil engineering, hydrology, and social science with goals to add functionality to other disciplines (astronomy, material science, the humanities, as well as for the general public).

How to include in data management plans
http://browndog.ncsa.illinois.edu/datamanagement

This data analysis/manipulation framework developed here will be pushed into the NSF DIBBs. Brown Dog (ACI-1201582) project on data management relies on the DTS and DAP services providing automatic data annotation/analysis and format conversions as broadly usable internal resources. Brown Dog aims to help provide services and tools to aid in the curation, accessioning, and indexing of data as well as the primary scientific software that might be developed for that purpose. As Brown Dog grows in community, the capabilities of these tools will be improved, will take part in an ecosystem of other data management tools, and will be integratable by others within the scientific community, perhaps in very different fields, as well as by the general public.

This research and development has been funded through National Science Foundation Cooperative Agreement ACI-1201582.

BROWN DOG

USER CHEAT SHEET

THE DAP REST API FOR CONVERSIONS
http://browndog.ncsa.illinois.edu

GET /api/v1/conversions/inputs	List all input formats that can be received
GET /api/v1/conversions/outputs	List all output formats that can be accepted
GET /api/v1/conversions/inputs?input=format	List all input formats that can reach the specified input format
GET /api/v1/conversions/outputs?output=format	List all output formats that can reach the specified output format
GET /api/v1/conversions/inputs?output=format	Convert the specified file to the requested output format
POST /api/v1/conversions/inputs?output=format	Convert the specified file to the requested output format
GET /api/v1/conversions/software	List all available conversion software
GET /api/v1/conversions/servers	List all currently available Software Servers

THE DTS REST API FOR EXTRACTATIONS
http://browndog.ncsa.illinois.edu

GET /api/v1/extract/extractables	List the legal file formats supported by currently running extractors
POST /api/v1/extract/extractable	Submit a file for extraction of metadata using the file's URL and returns a file id
POST /api/v1/extract/extractable	Submit a file for extraction of metadata and returns a file id
GET /api/v1/extract/extractable?id=extractable	Retrieve the status of all extraction processing for the file id
GET /api/v1/extract/extractable	Retrieve, including processing, and content based capabilities, and other desired products extracted for the specified file
GET /api/v1/extract/extractors	List currently available extractors
GET /api/v1/extract/extractor?id=extractable	List details of currently available extractors
GET /api/v1/extract/servers	List all currently available extractor servers

This research and development has been funded through National Science Foundation Cooperative Agreement ACI-1201582.