Contributors Landing Page

Brown Dog (BD) data transformation services, Data Access Proxy (DAP) and Data Tilling Service (DTS), provide data format conversions and metadata extractions, respectively, through a set of consistent and programmable REST APIs. These services make up a framework that is extensible by design. DAP, built on top of the Polyglot framework and DTS, built on top of the Clowder framework, support contributions of new tools from users across various research domains and make the tools available for others to use. A BD tool is a Brown Dog script or program that wraps any piece of code, software, library or web service and exposes their function, either data format conversion, or metadata extraction and analysis. Each BD tool is designed to fit into the framework of either Polyglot (data format conversions) or Clowder (metadata extractions), and is made available through the DAP or DTS API.

Within the Polyglot framework, each tool is known as a *converter* to emphasize its data format conversion capabilities. A *converter* is a wrapper script, written in any scripting language, such as bash, R, Python. A *converter* resides within a *software server*, which also hosts third-party software or libraries which provide the actual format conversion. The *software server* orchestrates the operation of the *converter* tool which runs the third-party software, handling input and output files.

Similarly, a tool within Clowder framework is known as an *extractor* to emphasize the extraction of metadata from files. An *extractor* is a program that extracts metadata from within a file. It analyses the file's contents and tags it according to some specific classification or criteria, etc. It resides in a distributed environment, such as a cloud server, as an extraction service. It listens to a message queue for extraction requests from DTS API. The extraction process is triggered based on specific file types (MIME Types) and the metadata extracted are then made available through the DTS API.

In the BD framework, *extractor* and *converter* are the extensible units, and can be contributed by researchers across different research domains. To enable contributions from the research community, a Tools Catalogue was designed and implemented where a user can register and share their *extractor* or *converter* with other users.

How can I contribute to BD data transformation services?

1. Identify your tool capabilities - extractions/analysis or conversions.
2. Decide if your tool is best fit as converter within DAP or as extractor within DTS. Look into some existing examples in BD Tools catalogue (http://browndog.ncsa.illinois.edu/tools).
3. Download bd-template project from the NCSA opensource repository (http://opensource.ncsa.illinois.edu/bitbucket/projects/BD/repos/bd-templates/)
4. Follow the README instructions within bd-converter-templates and bd-extractor-templates, to set up the BD development environment for converter and extractor, respectively.
5. Test the bd-templates with the test scripts provided within bd-templates.
6. Follow the README instructions to write a new extractor/converter with your tool, thereby converting it to BD tool.
7. Once tested in the local development environment, push the bd-tool code to the NCSA open-source repository(or github**??**).
8. Create am account in Tools Catalogue (http://browndog.ncsa.illinois.edu/tools) and register the new BD tool to it by clicking "contribute". This will involve you filling up forms providing description, link to the code repository, a dockerfile, input /output file. Submit it for admin approval.
9. The admin goes through your tool, test it and approves it.
10. On approval, your tool is visible to others and you can share the tool across different domain.

Why should I contribute?

Researchers often develop new tools for their research, in order to extract useful information from unstructured or semi-structured data. A lot of effort goes into developing new tools and such efforts are often unacknowledged. Instead, the value of the research is measured in terms of publications and analysis. In addition, similar tool development efforts are repeated by multiple researchers within the same domain of science. Towards acknowledging such tool development effort, we built the BD Tools Catalogue, where members of different research communities can contribute and share their tools within BD framework. In the BD Tools Catalogue you get proper credit for your effort in creating the new tool.

Can I contribute if my tool uses a third-party proprietary software?

Yes, you can submit your tool, even if it uses third-party proprietary software. However, before it is made available as Data Transformation Services, we will need to review the license agreement. Our current focus is on using open source software to build Brown Dog tools.