

Open Source Science ^[1]

The [Novartis Institutes for BioMedical Research \(NIBR\)](#) ^[2] is pioneering new informatics tools for drug discovery. We believe in the power of open-sourced, global collaboration for the greater good. Join us to help patients worldwide.

Software

Habitat



Habitat is a simple and yet powerful self-contained object storage management system. Based on Amazon Web Services, it is capable of virtually unlimited storage. Instead of a large centralized management system, Habitat can be used as a local repository for a single application or it can be shared and used with many clients.

Habitat is best used for situations where the client producers and consumers of the files do not require a file system protocol interface and can use http(s) to access the store.

[GitHub Project](#) ^[3] | [Download Habitat](#) ^[4]

YADA



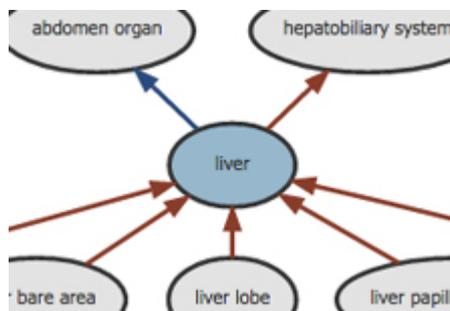
Access any data, at any source, in any format, from any environment, using just a URL, with just one-time configuration.

Get data from multiple sources, in different formats, merge the results into one with uniform column names, on-the-fly, using one URL.

Its raisons d'être are to enable efficient, non-redundant development of data-dependent applications and utilities, data source querying, data analysis, processing pipelines, extract, transform, and load (ETL) processes, etc. YADA does all this while preserving total decoupling between data access and other aspects of application architecture such as user interface.

[GitHub Project](#) ^[5] | [Download YADA](#) ^[6]

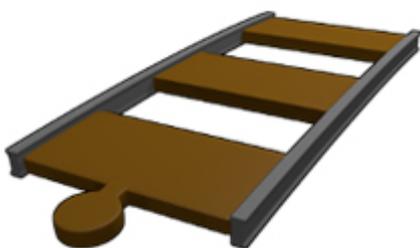
OntoBrowser



The OntoBrowser tool was developed to manage ontologies and code lists. The primary goal of the tool is to provide an online collaborative solution for expert curators to map code list terms (sourced from multiple systems/databases) to preferred ontology terms. Other key features include visualisation of ontologies in hierarchical/graph format, advanced search capabilities, peer review/approval workflow and web service access to data.

[GitHub Project](#) ^[7] | [Download OntoBrowser](#) ^[8]

Railroadtracks



Railroadtracks is a Python toolkit to handle graphs of dependent tasks such as the ones found in bioinformatics pipelines.

It was created for comparing RNA-Seq pipelines and found its use in other situations, such as

writing a flexible system for the QC of NGS data.

[GitHub project](#) ^[9] | [Download Railroadtracks](#) ^[10] | [Documentation \(PDF 0.5 MB\)](#) ^[11]

Yet Another Pipeline



YAP is an extensible parallel framework, written in Python using [OpenMPI](#) ^[12] libraries. It allows researchers to quickly build high throughput big data pipelines without extensive knowledge of parallel programming. The user interacts with the framework through simple configuration files to capture analysis parameters and user directed metadata, enabling reproducible research. Using YAP, analysts have been able to achieve a significant speed up of up to 36x in [RNASeq](#) ^[13] workflow execution time.

YAP has been designed to be scalable and flexible. We have implemented YAP with a focus on next-generation sequencing (NGS), to meet the large data processing challenges at NIBR. However, the framework can be easily adapted for any kind of analysis. It can be executed on your local Linux workstations or large HPC cluster systems. The framework achieves efficiency by implementing optimal data handling mechanisms such as, parallel data distribution, avoiding file I/O using data streams and named pipes.

[GitHub project](#) ^[14] | [Download Yet Another Pipeline](#) ^[15]

RDKit



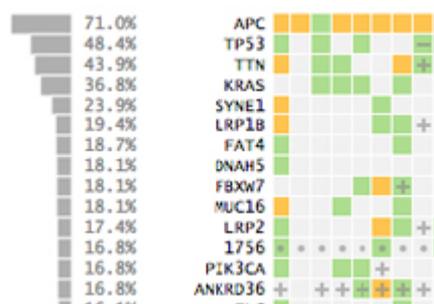
The RDKit is a collection of cheminformatics and machine-learning software written in C++ and Python. The core algorithms and data structures are written in C++. Wrappers are provided to use the toolkit from either Python, Java, or C#. Additionally, the RDKit distribution includes a PostgreSQL-based cartridge that allows molecules to be stored in a relational database and retrieved via substructure and similarity searches.

Please see the [RDKit Documentation](#) ^[16] for more information on installation, usage,

cookbooks, and lots more.

[GitHub project](#) [17] | [Download RDKit](#) [18]

GridVar



GridVar is a jQuery plugin that visualizes multi-dimensional datasets as layers organized in a row-column format. At each cell (i.e., rectangle at the intersection of a row and column), GridVar displays your data as a background color (like a color/heat map) and/or a glyph (shape). This enables different characteristics of your dataset to be layered on top of each other. For more information on usage, required libraries, and other developer information, please see our documentation on GitHub.

[GitHub project](#) [19] | [Download GridVar](#) [20]

Source URL: <https://www.novartis.com/our-science/open-source-science>

Links

- [1] <https://www.novartis.com/our-science/open-source-science>
- [2] <https://www.novartis.com/our-work/research-development/novartis-institutes-biomedical-research>
- [3] <https://github.com/Novartis/habitat>
- [4] <https://github.com/Novartis/habitat/releases/latest>
- [5] <https://github.com/Novartis/YADA>
- [6] <https://github.com/Novartis/YADA/releases/latest>
- [7] <https://github.com/Novartis/ontobrowser>
- [8] <https://github.com/Novartis/ontobrowser/releases/latest>
- [9] <https://github.com/Novartis/railroadtracks>
- [10] <https://github.com/Novartis/railroadtracks/releases/latest>
- [11] <https://www.novartis.com/sites/www.novartis.com/files/railroadtracks.pdf>
- [12] <http://www.open-mpi.org/>
- [13] <https://en.wikipedia.org/wiki/RNA-Seq>
- [14] <https://github.com/Novartis/yap>
- [15] <https://github.com/Novartis/yap/releases/latest>
- [16] <http://www.rdkit.org/docs/index.html>
- [17] <https://github.com/Novartis/rdkit>
- [18] <https://github.com/Novartis/rdkit/releases/latest>
- [19] <https://github.com/Novartis/gridvar>
- [20] <https://github.com/Novartis/gridvar/releases/latest>