

# iRefIndex: a meta resource providing access to a large collection of protein-protein interaction data

Alexander Botzki<sup>1,2</sup>, James Collier<sup>1</sup>, Mainak Guha Roy<sup>1</sup>, and Shoshana Wodak<sup>2</sup>

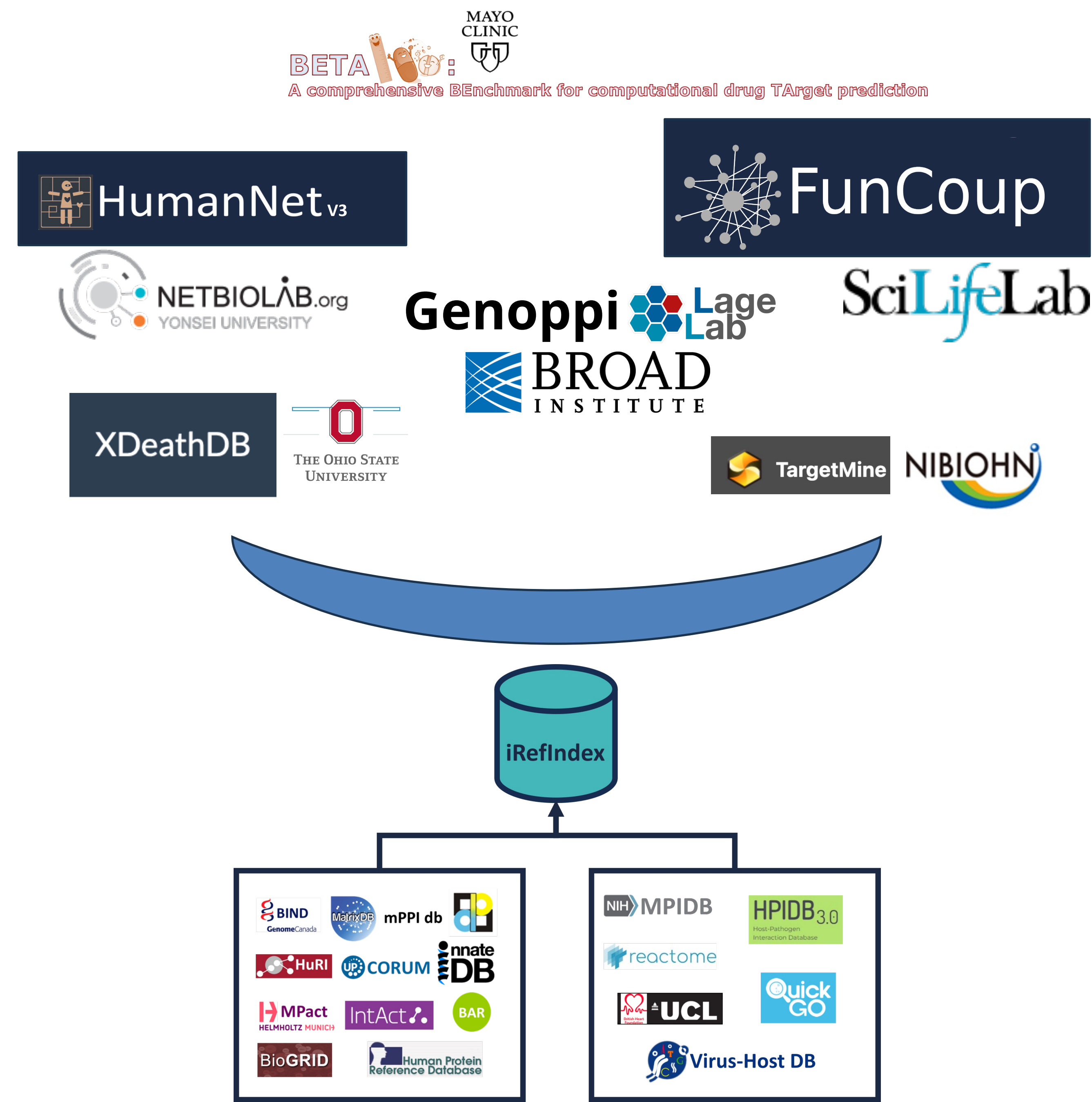
1 VIB Technologies, Ghent, Belgium 2 ELIXIR Belgium, Ghent, Belgium



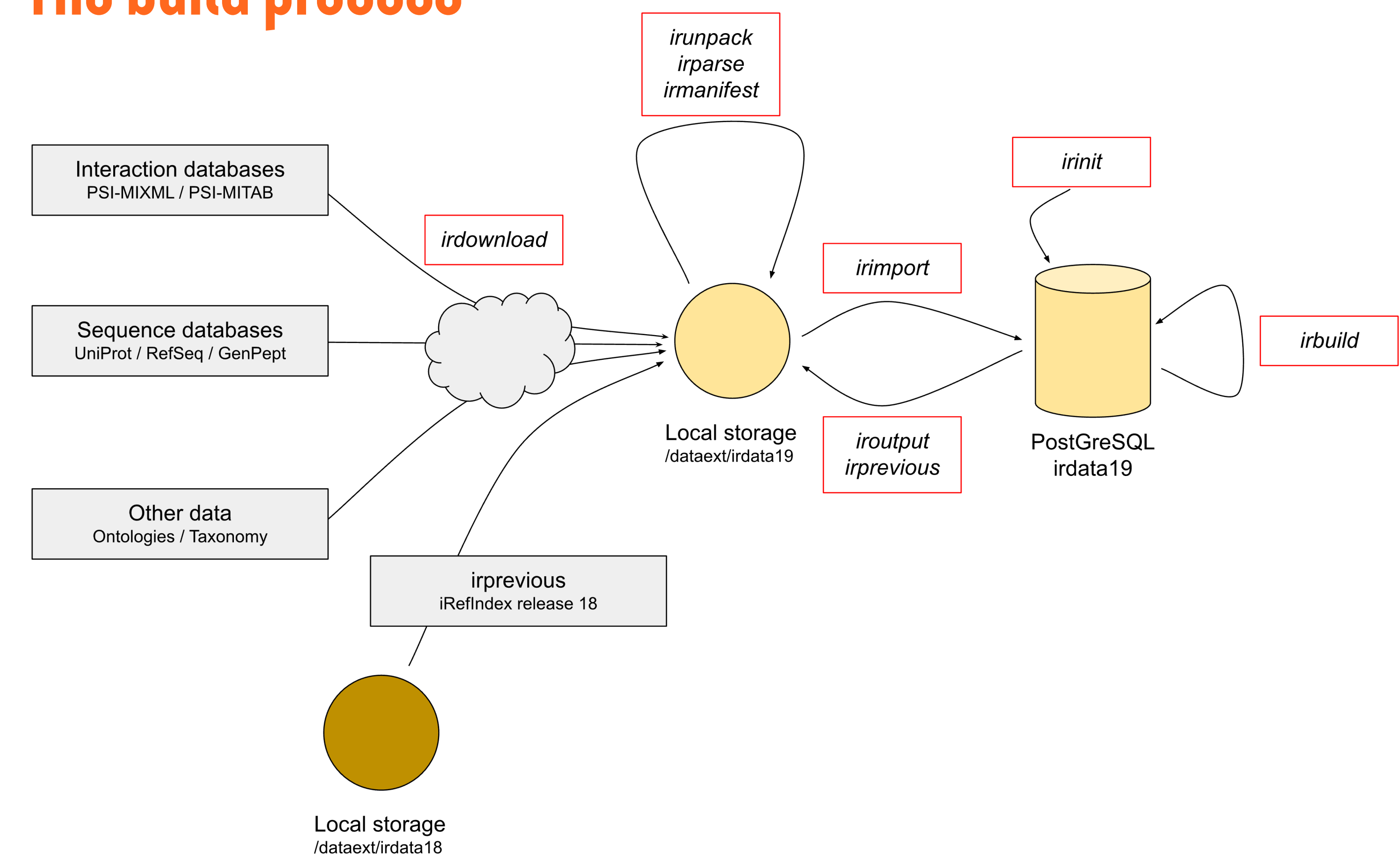
## What is it?

iRefIndex is a meta resource based on matching protein sequence data, providing access to a large collection of protein-protein interaction data aggregated from major primary public databases.

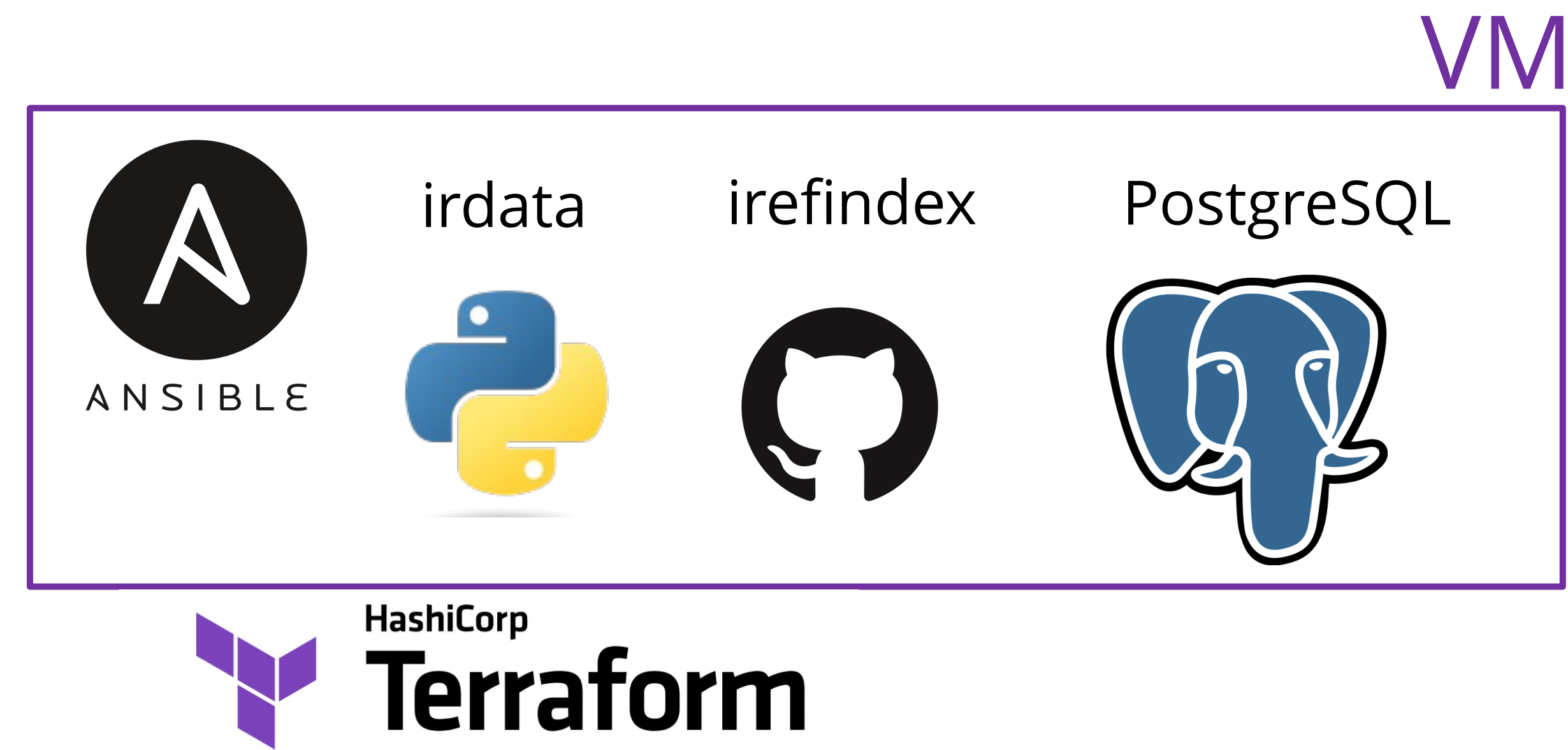
A user can retrieve a non redundant list of interactors with a query protein. iRefIndex is a unifying index that facilitates searching for these data whilst grouping together redundant interaction data and recording the methods used to perform this grouping. Users can integrate their own data in a way that ensures uniqueness for proteins with the identical sequence. Overall, iRefIndex is a valuable resource for life scientists to help them better understand the complex interactions and networks in living systems. Various resources use iRefIndex as one of the key data sources.



## The build process

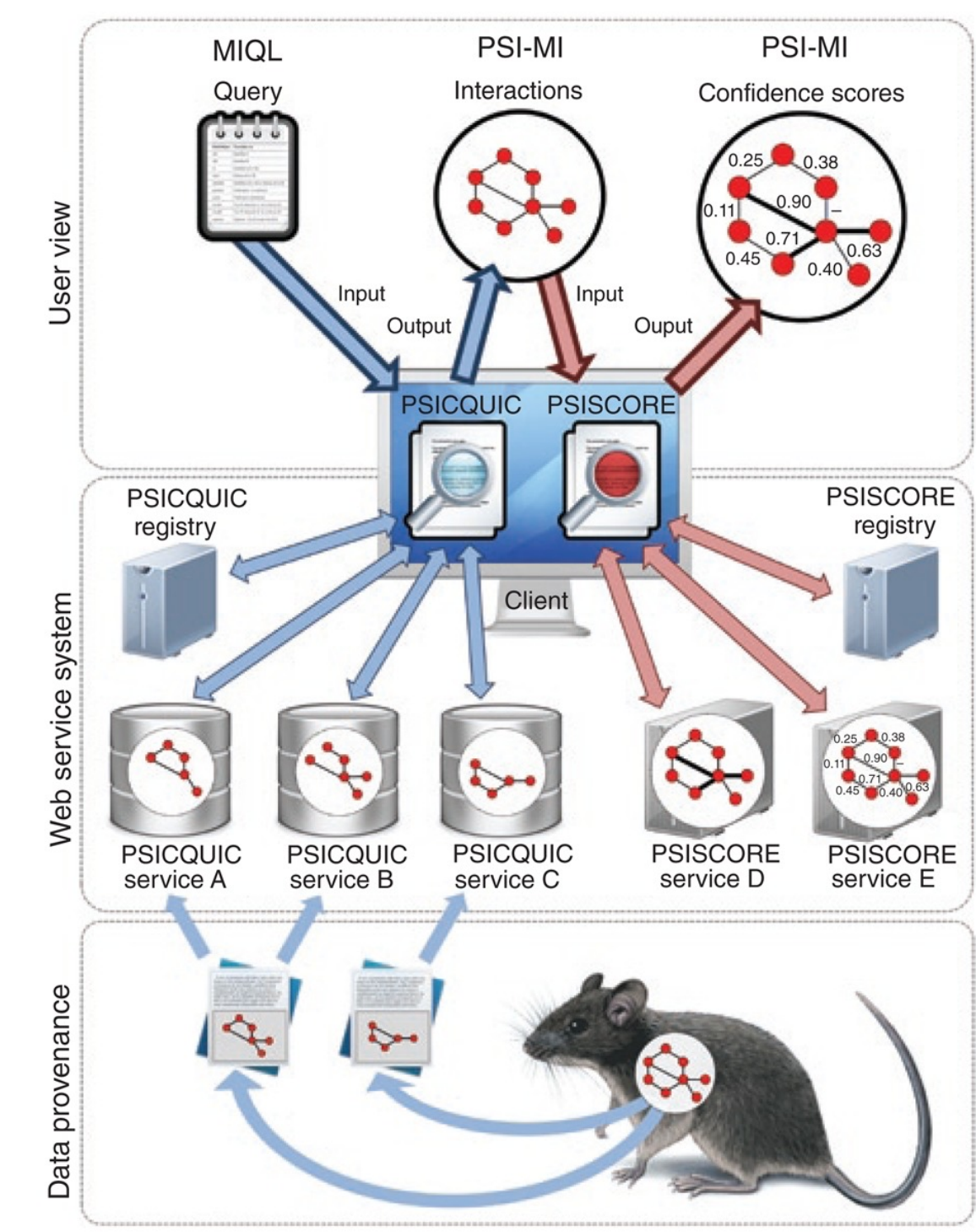


## The build infrastructure



The build infrastructure uses Terraform and Ansible to automate the provisioning, configuring, deployment, and tearing down of infrastructure on Google Cloud. Terraform is used to provision the virtual infrastructure resources like a VM including a PostgreSQL database, Ansible is used to automate configuration, dependency management, and deployment.

## Querying the iRefIndex dataset



Bioconductor  
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS  
**PSICQUIC**  
**Path2PPI**  
**cisPath**

The iRefIndex dataset can be queried using tools such as PSICQUIC, cisPath, and Path2PPI from BioConductor. PSICQUIC is based on a template from the IntAct group at EBI. cisPath is a package on BioConductor that allows users to search for the shortest paths between proteins in a network. Path2PPI is another package in BioConductor that provides access to protein-protein interaction data.

