# Background

A common problem in many fields of science is that it since there are so many articles being published, it can be hard to see the big picture and mentally combine results from different studies. Let’s imagine a case where a researcher is writing a review article on the effect of X to Y. They find multiple articles on that topic, but only have access to selected results that are presented on the article, either only as a visualization or as an embedded table in the article. In the best case, the complete result files are available for download as a supplementary material. Unluckily, the format of the results can vary and browsing through the results of multiple related studies can prove time-consuming.

Sharing and harmonizing results from multiple studies can also prove problematic for researchers generating these results. Imagine a researcher who has recently completed multiple studies on similar topics. They now want the combined results to be easily accessible for anyone, to maximize the impact of their work. Should they go through their research and summarize the results in a review paper? What about the most recent manuscript that will be submitted in a few months? Should they wait for that paper and delay the review or publish the review without the most recent results?

To address these challenges, we have developed Ninni, a web-based application for representing and visualizing study results. Ninni has its own dedicated database with a harmonized data format, where results from multiple different studies can be saved. The results stored in the database can be accessed through a web interface. Users can combine results from multiple studies, visualize and download them.

# Implementation

Ninni was developed in R (version 3.5.3) with the Shiny package.

The data displayed by Ninni is stored in a PostgreSQL database. Data is imported with a Python script.

The source code, detailed structure of the database and instructions for importing data are found on GitHub. Ninni is licensed under the terms of the MIT license.

Online version of Ninni with results from studies of the University of Eastern Finland is available at bioinformatics.uef.fi/ninni. Local version of the software can be used as a standalone Shiny application. Instructions for setting up a local copy of the software are also found on GitHub.

# Results

Ninni is specifically designed to represent results from studies where some form of association between variables and an outcome is studied. These associations can be for example the effect of drugs to the risk of arrhythmia or the differential expression of genes in healthy and diseased individuals. The associations can also represent the effect of interaction of variables to an outcome, e.g. the effect of an interaction of drugs to the risk of arrhythmia.

## Data storage

Researchers need to store their data in a specified format to be able to import them into Ninni’s database. The associations should always have a measurable effect size. The effect size can be measured in odds ratio, fold change or correlation. One should also report 95 % confidence intervals of the effect size, sample size and the statistical significance of the effect, measured as both unadjusted p-value and false discovery rate adjusted p-value. In addition to this obligatory basic information, any number of extra columns can be added, which provides flexibility for researchers. More detailed information on importing data is available on GitHub.

The data format used by Ninni requires the following fields:

* Variable labels
* Effect size
* Upper and lower bounds of 95% confidence intervals for effect size
* Sample size for the association
* P-value
* FDR-adjusted p-value

Table 1: The data format for studies with association of variables to an outcome. Example data in this table represents the effects of drugs on ????. The variable labels are ATC codes for drugs.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **VARIABLE1\_LABEL** | **EFFECT** | **EFFECT\_L95** | **EFFECT\_U95** | **N** | **P** | **P\_FDR** |
| A01AB04 | 0.02217 | 2.75E-74 | 1.79E+70 | 2 | 0.964 | 0.973 |
| H01AC01 | 1.579 | 1.13 | 2.20 | 19113 | 0.00696 | 0.022 |
| C02DC01 | 1.652 | 0.61 | 4.40 | 2087 | 0.316 | 0.598 |

Table 2: Data format for studies with associations of two variables. The example data shows correlations between metabolites measured with LC-MS metabolomics. The data is anonymized data from a metabolomics study at University of Eastern Finland.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| VARIABLE1\_LABEL | VARIABLE2\_LABEL | EFFECT | EFFECT\_L95 | EFFECT\_U95 | N | P | P\_FDR |
| Metabolite\_2 | Metabolite\_1 | 0.6537 | 0.5777 | 0.718477 | 258 | 7.78E-33 | 2.73E-31 |
| Metabolite\_3 | Metabolite\_1 | 0.0964 | -0.025 | 0.21605 | 258 | 0.122 | 0.227 |
| Metabolite\_4 | Metabolite\_1 | -0.212 | -0.325 | -0.0923 | 258 | 0.000607 | 0.0023 |

## Browsing results

Users can browse Ninni’s database and retrieve data in different ways via the web application. It is possible to choose the datasets by their name or search for all datasets related to a topic, e.g. drug interactions. Users can also search for certain variables, such as drugs or genes in the complete database or in a limited number of datasets. This way users get all the available information related to that variable.

After retrieving data from the database, users can filter the results based on any column e.g. effect size, p value or sample size. Users can also limit the results to certain variables or exclude unwanted variables. The resulting data table will then be presented to the user. The data can also be downloaded as a .csv file.

## Visualizations

In addition to the data table, Ninni provides a range of tools for visualizing study results. For data with less than NNNNN (to be tested) associations, all the visualizations are interactive plots with tooltip boxes that appear over points on mouse hover. These tooltip boxes provide basic information of the specific point: name(s) of the variable(s), effect size, sample size and adjusted p-value. Interactivity makes finding relevant information straightforward. Unfortunately, for large amounts of data interactive visualizations have proven too slow to compute. All plots can be downloaded as PNG or PDF files. If the dataset size is small enough for interactive plots, the interactive plot can be downloaded as an HTML file that can be opened in a browser without connection to Ninni.

Data containing interactions of variables can be visualized using heat maps with an option to apply hierarchical clustering to order the variables. Symmetrical heat maps can also be visualized as lower triangular heatmaps, where the duplicated entries are removed so that only the lower triangular is plotted.

To visualize the spread of effect size and statistical significance, Ninni provides volcano plots where points can be colored based on limits on both effect size and p-value. Users can also test distribution assumptions using Q-Q plots for both effect sizes and p-values. The distribution of statistically significant associations in different parts of the data can be visualized with a so-called signed Manhattan plot. In a traditional Manhattan plot, the –log10 transformation of p-values is used as the y-axis. In our variant, the values on the y-axis are multiplied by the sign of the effect (or sign of log2-transformed effect if more sensible). This adds extra information to the plot, since the viewer can separate between positive and negative associations.

When retrieving data from Ninni’s database it is possible to retrieve effects of differing types, such as correlation and fold change. In this case, visualization of the data is disabled as plotting different effect types together could yield misleading plots. The data can still be viewed in a table and downloaded.

When visualizing results from multiple studies users need to make sure that the graphs remain sensible and they can provide valid information.

**Conclusion**

Ninni offers a great way of publishing study results. The results can be easily accessed through a web interface and visualized on any platform without installing any software. This tool will hopefully promote open science and make results of scientific studies more easily accessible.