

Presenters

Masomi Hatakeyama and Hubert Rehrauer

Functional Genomics Center, University of Zurich

Title

Sushi: An exquisite recipe for fully documented, reproducible and reusable NGS data analysis

Abstract

We present Sushi, an agile framework for web- and commandline-based data analysis that lets users build fully reproducible analysis workflows. All results and associated meta-information are fully defined on the file system in tabular clear text format that is independent of Sushi and ready for sharing and distribution.

Demo Summary

In the demo slot we will initially introduce the sushi framework using slides. Subsequently we will demo the analysis of an RNA-seq data set, which gives the user view on the system. Again on the life system we will show the system manager view and add an additional analysis App. We will end the live demo by showing the commandline interface that is targeted for power users, i.e. experienced bioinformaticians. At the end we will, again with slides, highlight the benefits of Sushi for the users and for the bioinformatics analysis.

The Sushi data analysis framework is driven by the goal of making bioinformatics analysis simple, portable and reproducible. We put simplicity first but achieve all three goals at multiple levels. First, users have a simple interface to run their analysis, since we expose only the most important options as user interface elements, and we give at the same time advanced users full control over the entire set of options of the underlying applications. Second, data sets are fully defined on the file system in a straightforward structure, so that even researchers who have not run the analysis and may even not be aware of Sushi can make use of the analysis results. Third, we store all meta-information in tabular clear text format which is at the same time human-readable and machine-parseable, which makes data sets self-explanatory to a human reader but also suitable as input for subsequent analysis packages. Forth, we don't impose any content of the meta-information so that every installation can define this information with respect to the needs of the researchers and to the requirements of the data analysis domains. Fifth, we make it straightforward to add new apps to the sushi framework, allowing the bioinformatician to fully define an app in a single file. The full definition includes user interface elements, documentation, input requirements, validity checks, and data processing commands. Data processing commands can be written in any language that is supported by the host system. The probably biggest benefit for the

bioinformatics analysis comes from the open design with respect to the structured but extendible storage of meta-information and with respect to the fact that generated data sets have no dependency on sushi and can be directly further reused.

Sushi is implemented in Ruby, which as a language lends itself readily to the above-mentioned goals. The web interface relies on Ruby-on-Rails which makes web development very efficient. We have chosen a modular design with respect to the data sets, the applications and the computing infrastructure. Sushi can make use of local compute resources as well as job submission systems that are connected to clusters or the cloud.

Material Requirements

Ideally there would be two beamers, so that slides and live demo can be shown simultaneously. Ideally we also have WIFI access to the sushi demo server.

Expected number of attendees

Sushi is relevant for bioinformaticians and researchers in the field of NGS. We expect at least 100 attendees per 30min slot.