

EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



# Data Management for Quantitative Biology

## Project Paper

### **Project 1: Data Modelling and Visualization**

*SS 2015*



Sebastian Goerges  
Benjamin Schroeder  
Nils-Oliver Schliebs

## Abstract

....

## 1 Background

Data management in life sciences connects different fields and people with different scientific backgrounds. One challenge, especially when data needs to be collected and presented, is to keep track of the metadata that is important for analysis and reproducibility of an experiment, while exposing only informative and easy to understand data and metadata to users that are not concerned with computation or data management. Hence, for a given data set, the project task was to define what parts of data should be shown to end users (biologists) and to implement a web-based Graphical User Interface, using Java, to visualize given data. Users should also be able to add some annotation on their own and download a summarizing report of the results. In addition, the user interface should be intuitive and well documented to guide even first-time users easily through the task of data annotation.

## 2 Material and Methods

### 2.1 Data

The main data structure is provided in form of TSV-files. The file *projects.tsv* contains multiple projects together with several attributes, describing the project. Projects themselves can contain multiple different experiments (*experiments.tsv*), which in turn contain two different types of samples (*QCOFF.tsv* and *QMOUSE.tsv*). *QCOFF* is storing samples of a coffee diversity project, whereas *QMOUSE* consists of samples for a mouse knockout project. The set of samples can be structured hierarchically, with patients/organisms at the top (Entity samples) from which tissue/cell samples are derived. Experiments or samples can contain datasets, stored in *datasets.tsv*. Datasets link to one or more files. These files are for instance further descriptions in plain text, images, quality control HTML documents or FASTQ files. All structures of the system contain unique identifiers to connect the set of TSV-files together in a logical way.

## 3 Results

...

## 4 Discussion

...