# Requirements Document

The Many Faces of Microbial Communities Project

Team Name: ViewCrobe Software

Members: Thomas Albertine, Michael Phelps

Contents

[1 Requirements Document 1](#_Toc434431039)

[2.1 Introduction 2](#_Toc434431040)

[2.1.1 Purpose 2](#_Toc434431041)

[2.1.2 Scope 2](#_Toc434431042)

[2.1.3 Definitions, Acronyms, and Abbreviations 2](#_Toc434431043)

[2.1.4 References 3](#_Toc434431044)

[3 Bibliography 3](#_Toc434431045)

[3.1 Overall Description 3](#_Toc434431046)

[3.1.1 Product Perspective 3](#_Toc434431047)

[3.1.2 Product Functions 3](#_Toc434431048)

[3.1.3 User Characteristics 4](#_Toc434431049)

[3.1.4 Assumptions and Dependencies 4](#_Toc434431050)

[3.2 Specific Requirements 4](#_Toc434431051)

[3.2.1 Sample Subset 4](#_Toc434431052)

[3.2.2 Organism Subset 4](#_Toc434431053)

[3.2.3 Model Parameter Assignment 4](#_Toc434431054)

[3.2.4 BIOM Format Support 5](#_Toc434431055)

[3.2.5 Transforming Data 5](#_Toc434431056)

[3.2.6 Producing a Model 5](#_Toc434431057)

[3.2.7 Drawing Models 5](#_Toc434431058)

[3.2.8 User Controls 5](#_Toc434431059)

[3.3 Additional “Stretch” Goals 5](#_Toc434431060)

[3.3.1 Provide a Web UI 5](#_Toc434431061)

[3.3.2 Export Images for Later Comparison 5](#_Toc434431062)

[3.3.3 Save and Load Visualization Profiles 5](#_Toc434431063)

## Introduction

### Purpose

The theory behind the project is that, since there is a specific portion of the human brain dedicated to processing faces, visualizing microbial data as faces allows users to better compare the complex data. Therefore, the purpose of the project is to provide that functionality in a tool usable primarily by microbiologists, but secondarily by researchers in other disciplines.

### Scope

FaceView will be a user-friendly tool for providing visualizations of microbial population data by representing data as 3d models of humans. Given a supported data file, FaceView will allow the user to select a subset of samples from the file from which to generate models based on a subset of the various populations in the file. After these models are generated, these models will be draw to the screen for comparison by users.

In order to enhance user-friendliness, sample subsets will be relatively small for ease of comparison, while population subsets will be significantly larger to allow for more comprehensive comparisons.

### Definitions, Acronyms, and Abbreviations

* Model—In the context of this project, “model” refers to a 3 dimensional representation of a human based on a subset of the input data.
* Sample—In the context of an input file, a sample is a subset of the data representing one population. As an example, a sample may contain the number of colonies and which species they are associated with in a petri dish
* Common Camera Angles—A set of configurations for the observer of a 3d scene that are used often. Such a set would include a front view, a profile view, and a three-quarter view at least.
* Visualization Profiles—Data files may include many samples and many organism populations. It may contain more populations than there are supported models or model parameters. Therefore it may become necessary to ignore some samples and some organism populations. Furthermore, the user may want to reassign the model parameter that corresponds to a given organism’s population to better emphasize differences in that organism among the samples. However, manually modifying which model parameters correspond to which organisms and which samples should be ignored could be time-consuming to do each time. Therefore it may be expedient to create a profile of these settings, which could be saved and loaded later
* Model Parameters—Values that can be passed in when generating a model in order to influence the outcome. For example, nose length might be a model parameter
* Data Parameters—Values that are read from the data. These can be associated with model parameters to make the visualization reflect the data. For example, the population of E. Coli in a set of samples, which may be tied to the nose length in the finished model.

### References

# Bibliography

|  |  |
| --- | --- |
| [1] | "The Biom File Format," 21 October 2015. [Online]. Available: http://biom-format.org/documentation/biom\_format.html. [Accessed 3 November 2015]. |
| [2] | "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information," *Psychological Review,* vol. 63, pp. 81-97, 1956. |

## Overall Description

### Product Perspective

This project interacts with several other components. It reads and interprets supported data file formats. The minimal acceptable set of supported formats is only the BIOM format [1], as it is sufficient for use, but expanding the set of supported file formats is a stretch goal.

Additionally, we include a UI component to facilitate user interaction. This contains primarily two components.

* A system of menus that allow users to select a subset of samples and organisms to include, and to configure which data parameters correspond to which model parameters.
* A window in which the produced models will be drawn. At this point, the user would be able to rotate, pan, and zoom around the models in order to better compare their features.

We recognize that certain additional features, while not absolutely necessary for a minimal functional product, would enhance usability, and therefore we add, as stretch goals, shortcuts to certain common camera angles, some way to save and load visualization profiles, and a way to export comparison images. Via our UI, a user from our target user group should be able to use our tool to produce their visualization in 2 minutes[[1]](#footnote-1) or less with approximately 2 minutes or less of initial instruction.



Figure 1: A Functional Diagram, demonstrating the interactions between the user and the application.

### Product Functions

The project should support the following functions

* The project should allow the user to select a subset of samples in a file to be drawn to the screen. Note that, in order to allow the user to easily identify similarities and differences in the drawn models, the number of models that can be drawn to the screen, and thus samples that can be represented at once, is limited to 6, since the maximum number of things a person can process at a given time is 7 ± 2 [2].
* Similarly, the project should allow the user to select a subset of organisms to associate with certain module parameters, as those are limited by the number of model parameters that can be modified.
  + Furthermore the user should be able to choose which supported model parameters correspond with which organisms in the data file. This allows the user to allocate more obvious model features to the organisms he or she is most interested in.
* The project should, using the data files and visualization profile, generate models and draw them to the screen.
* Once these models are drawn to the screen, the user should be able to rotate, pan, and zoom the views.

### User Characteristics

Our intended users are researchers who are analyzing population data, specifically microbial population data. Such users are highly educated, but do not want to have to use unnecessarily complex or difficult tools.

### Assumptions and Dependencies

We assume that Windows will be installed on the machine running the project.

## Specific Requirements

### Sample Subset

The project should allow users to select a subset of at most 6 samples to compare. This is to improve user-friendliness. Restricting the number of models shown may reduce distractions, allowing users to better compare certain samples.

### Organism Subset

The project should allow users to select a subset of organism appearing in the samples. This subset is limited by the number of model parameters which we can support, which should be at least 150.

### Model Parameter Assignment

The project should allow users to specify certain model parameters to correspond to certain measurements in the data file, such as the population of a certain microbe corresponding to eye shape. This is to improve user-friendliness by allowing users to match parameters used in an existing result, or to allow certain features to be made more obvious.

### BIOM Format Support

Initially, the project should support the BIOM format for data files. It is relatively human readable as it is essentially JSON, which makes it a good choice for a proof of concept. Expanding file format support is a stretch goal.

### Transforming Data

The project should transform the data from the data file into a format that can be used to generate the models. This allows us to produce a standard format for model generation, which allows us to expand supported file formats.

### Producing a Model

The project should produce a 3d human model based on the parameters from each selected sample. For each model generated in the batch, the same data parameter corresponds to the same model parameter. For example, if the population of a certain microbe in a certain sample corresponded to the straightness of the produced model’s nose for that sample, then the population of the same microbe in a second sample would correspond to the straightness of its model’s nose. This process should not take longer than 30 seconds.

### Drawing Models

The project should draw the produced models to the screen such that each of the models are viewable at the same time. This is to facilitate the user comparing models.

### User Controls

The project should allow users to rotate, pan, and zoom around the models in order to better compare them. All models should rotate, pan, and zoom.

## Additional “Stretch” Goals

### Provide a Web UI

In addition to the local version, this would provide a web UI which would invoke the component that generates models on the server. These models would then be sent to the client browser to be presented to the user. This stretch goal would necessitate that the project be portable enough to run on a Linux server.

### Export Images for Later Comparison

This feature would allow users to not only compare samples in the same file, but also to compare to files compared earlier. However, this would allow the user the opportunity to associate different data parameters with different model parameters between the two results, which could lead to misleading results.

### Save and Load Visualization Profiles

This feature would allow users to save and load visualization profiles, configurations of which samples and organisms from a given file are to be used when generating the model, and which organism corresponds to which feature.

1. This time does not include time spent configuring the visualization profile, since that may increase dramatically based on the size of the data file. [↑](#footnote-ref-1)