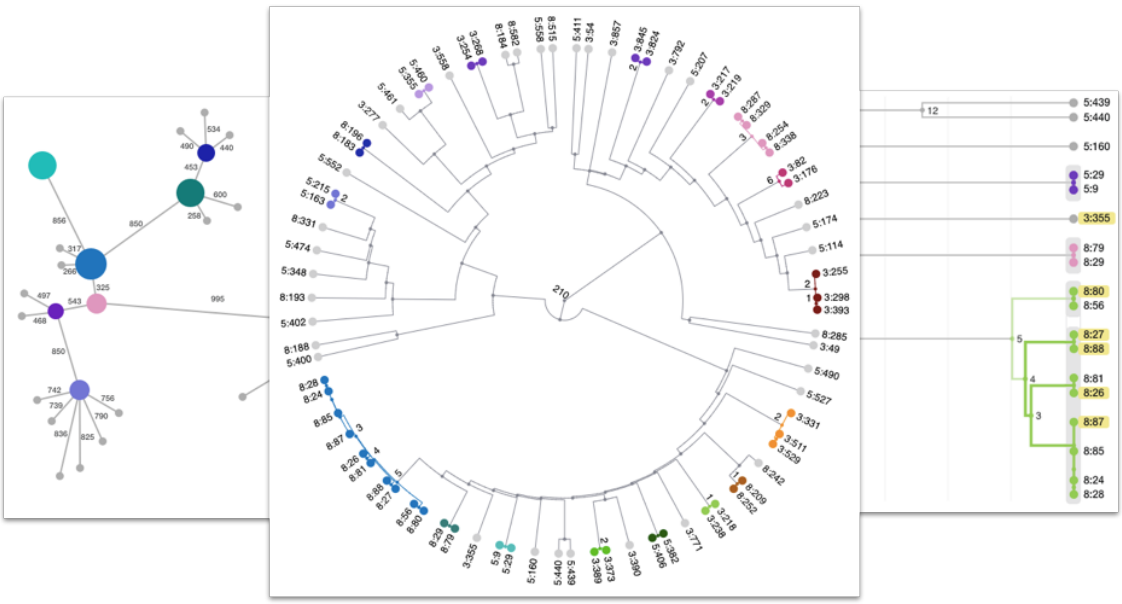
CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG



Improved Infection Control Reporting through Interactive Data Visualisation

Master's Thesis in Interaction Design and Technologies

MATILDA HORPPU

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2019

MASTER'S THESIS 2019:NN

Improved Infection Control Reporting through Interactive Data Visualisation

Matilda Horppu

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Improved Infection Control Reporting through Interactive Data Visualisation
MATILDA HORPPU

© MATILDA HORPPU, 2019.

Examiner: STAFFAN BJÖRK

Master's Thesis 2019:NN
Department of Computer Science and Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Final screens from the visualisation.

Department of Computer Science and Engineering
Gothenburg, Sweden 2019

Improved Infection Control Reporting through Interactive Data Visualisation
MATILDA HORPPU
Department of Computer Science and Engineering
Chalmers University of Technology

Abstract

1928 Diagnostics is a company in Gothenburg that strives to preserve the power of antibiotics by providing hospitals with a tool they can use in their work with infection control. The platform they have developed analyses NGS data and presents the results to the user in different ways. Today, the main users of the platform are microbiologists that use it for identifying disease outbreaks, among other things. An important task that the microbiologists are faced with is to report this result to the infection control unit at the hospital so that they can take action if an outbreak has occurred.

Currently, the reporting step in the infection control workflow is problematic since there is a large knowledge gap between the microbiologists and the infection control unit which puts high requirements on how the tests results are communicated. Furthermore, the microbiologists have no tool that includes all information they need to include in the reports they are creating which lead to a lot of extra work. Therefore, this master's thesis will focus on how interactive information visualisation can be used to design a tool that makes the reporting easier for the microbiologists and improves the communication with the infection control unit.

Before the concept development phase a literature study of the area and user study were carried out in order to build a good foundation for the project. An iterative design process was applied to the execution phase, where each iteration consisted of idea generation, prototyping suited to the goal of the iteration, user testing and an analysis of the results.

The project resulted in a set of design guidelines that should be considered a starting point for further research. The Guidelines were developed alongside a prototype that can be seen as an example of how the guidelines could be applied in a design.

Keywords: interaction design, information visualisation, infection control, NGS.

Acknowledgements

I would like to give a big thanks to the fantastic team at 1928 Diagnostics for supporting me throughout the project. Your input throughout the design process, both in focus groups and user tests, has been very valuable for the outcome of this project.

Secondly, I want to thank the two persons that participated in the initial user interviews and later the evaluation of the design. The information I received from you gave me a necessary understanding for how the infection control work is done and it also helped me to refine the final guidelines. I also would like to thank everyone that answered the user questionnaire.

Finally, I would like to thank my supervisor Marco Fratarcangeli at Chalmers University of Technology for helping me define the project scope and giving me guidance in my research.

Matilda Horppu, Gothenburg, May 2019

Contents

1 Introduction **1**
- 1.1 Antibiotic Resistance . . . 1
- 1.2 Genome Sequencing and Infection Control . . . 1
- 1.3 Stakeholders . . . 2
- 1.4 Research Problem . . . 2
  - 1.4.1 Research Question . . . 3
- 1.5 Limitations . . . 3

2 Background **5**
- 2.1 Current Visualisation . . . 5
  - 2.1.1 Understanding the Information . . . 5
  - 2.1.2 Visual Mappings . . . 6
  - 2.1.3 Interactions . . . 7
- 2.2 Related Work . . . 7
  - 2.2.1 Phylogenetic Tree Visualisation . . . 7
  - 2.2.2 Minimum Spanning Tree Visualisation . . . 11
  - 2.2.3 Evidence-Based Design of a Clinical Report . . . 12

3 Theory **15**
- 3.1 Wicked Problems in Design . . . 15
- 3.2 Information Visualisation . . . 15
  - 3.2.1 Visual Perception . . . 15
  - 3.2.2 The Designer-Reader-Data Trinity . . . 16
  - 3.2.3 Information Visualisation Principles . . . 17
- 3.3 Interface Design Principles . . . 18
  - 3.3.1 Affordance, Signifiers and Pliancy . . . 18
  - 3.3.2 Colour . . . 19
  - 3.3.3 Gestalt laws . . . 19

4 Methodology **21**
- 4.1 An Iterative Design Process . . . 21
- 4.2 User-Centered Design in Information Visualisation . . . 22
- 4.3 Methods . . . 23
  - 4.3.1 Empathise Methods . . . 23
  - 4.3.2 Define Methods . . . 23
  - 4.3.3 Ideate Methods . . . 24

4.3.4 Prototype Methods 24
4.3.5 Test Methods 25

5 Planning **27**
5.1 Process 27
5.2 Tools 27
5.3 Time Plan 28

6 Execution **29**
6.1 Pre study 29
6.1.1 Literature study 29
6.1.2 Review of Reports Used for Infection Control 29
6.2 Empathise and Define 30
6.2.1 Interviews and Observations 30
6.2.2 Personas 31
6.2.3 Task analysis 33
6.2.4 First Version of Guidelines 35
6.3 First Iteration 36
6.3.1 Ideation and Concept Development 36
6.3.2 Paper Prototype 40
6.3.3 Updated task analysis 40
6.3.4 User test 41
6.3.5 Focus Group 42
6.3.6 Results 42
6.3.6.1 User Tests 43
6.3.6.2 Focus Group 44
6.3.7 Revised Guidelines 46
6.4 Second iteration 47
6.4.1 Ideation and Prototyping 47
6.4.1.1 Phylogenetic Trees 47
6.4.1.2 Minimum Spanning Tree 50
6.4.2 User Questionnaire 51
6.4.3 Result 51
6.4.4 Revised Guidelines 54
6.5 Third iteration 55
6.5.1 Ideation and Prototyping 55
6.5.2 Focus Group 61
6.5.3 Expert Interview 61
6.5.4 User Tests 61
6.5.5 Questionnaire 61
6.5.6 Results 62
6.5.6.1 Focus Group 62
6.5.6.2 Expert interview 62
6.5.6.3 User tests 63
6.5.6.4 Questionnaire 64
6.5.7 Updated Guidelines 64

7 Results **67**
7.1 Final Guidelines . . . 67
7.1.1 Focus on tasks . . . 67
7.1.2 Adapt for Recipient . . . 68
7.1.3 Immediate Answers . . . 68
7.1.4 Flexibility . . . 69
7.1.5 Domain Familiarity . . . 70
7.1.6 Continuity . . . 70
7.1.7 Emphasise interaction possibilities . . . 70
7.2 Final Prototype . . . 71
7.2.1 Levels and Zooming in Diagrams . . . 71
7.2.2 Cluster and Outbreak Detection . . . 73
7.2.3 Colour Coding . . . 75
7.2.4 Highlighting . . . 76
7.2.5 Actions . . . 77
7.2.6 Annotations . . . 78
7.2.7 Time Line . . . 78
7.2.8 Infection Route Finder . . . 79
7.2.9 Reporting . . . 80

8 Discussion **83**
8.1 Results . . . 83
8.1.1 Design Guidelines . . . 83
8.1.2 Final Prototype . . . 84
8.2 Execution . . . 85
8.3 Ethical Considerations . . . 87
8.4 Future Work . . . 88

9 Conclusion **89**

A Tasks for the Paper Prototype User Test **I**

B Questionnaire Results **III**

C Questionnaire for the Infection Control Unit **IX**

D Questionnaire for the Infection Control Unit **XI**
D.1 Section 1 . . . XI
D.2 Section 2 . . . XII
D.3 Section 3 . . . XIII
D.4 Section 4 . . . XIV

1
Introduction

This first chapter gives an introduction to the domain in which the project was conducted and the problem that it is aiming to solve. The research question of the thesis will be sated as well as the limitations to the project scope.

1.1 Antibiotic Resistance

Antibiotic resistance is one of the biggest threats to modern healthcare globally today [1]. When bacteria develop resistance towards antibiotics, infections become increasingly hard to treat, especially if they become resistant to several antibiotics [2]. Even infections that we have been able to treat in the past might be very difficult to treat in the near future. It is always problematic if a person is infected by multi-resistant bacteria, but it is even more problematic if a person in a hospital is infected since it spreads to other patients quite easily.

Infections that patients get during their hospital stay are called hospital acquired infections (HAI) and each year hundreds of millions of patients are affected [3]. Hospitals continuously work towards limiting the occurrence and spreading of HAIs since they have several negative impacts on the patients and the hospital. The most serious consequence of HAIs is the unnecessary suffering for the patients, in some cases the HAI is even the reason why a patient dies. When there is an outbreak of a HAI a consequence can be that the hospital is forced to close an entire ward temporarily. That will lead to less patients being able to receive care and in some countries also reduce the profit for the hospital.

1.2 Genome Sequencing and Infection Control

Next Generation Sequencing (NGS) has revolutionised the technology of genome sequencing by making the process faster and cheaper [4]. Consequently, hospitals are able to start using the technology more frequently, for example for making the infection control work more efficient. By sequencing the genome of bacteria found in patients it is possible to see how closely related the bacteria are and thus if a patient has infected another patient or not [5]. This method of detecting outbreaks provides more scientific evidence, and in higher resolution, than more traditional lab-based methods. Therefore the hospital can make better informed decisions in infection control matters that can lead to various actions, for example

isolate patients and only close down wards when necessary. This method also has the potential to change how hospitals work with infection control and allows for more rapid discoveries of infection outbreaks and more efficient treatment.

1928 Diagnostics is a company operating in Gothenburg, Sweden, that has developed a web based platform that can be used for analysing data from NGS of several different bacteria [6]. The platform is used in hospitals as a tool for exploring the results from the analysis and finally provide the hospital with information that can inform decision making within infection control. In practice, this means compiling a report that is sent to the infection control unit that works towards preventing infections from spreading within the hospital. 1928 Diagnostics' goal is to create a user friendly product that is a natural part of the hospital's infection control workflow. Since the current presentation of the results is fairly static and require specialised skills to understand, they think that the platform lacks certain qualities to fulfill this goal.

1.3 Stakeholders

The main stakeholders in this project are 1928 Diagnostics, 1928 Diagnostics users and the recipients of the report used in infection control that is created based on the information acquired from 1928 Diagnostics' platform.

The users of the platform are mainly microbiologists that work in labs at hospitals. The microbiologist performs the NGS on bacteria from patients and then uses the 1928 Diagnostics platform to analyse the data. When the analysis is done the user can see the results in the platform. Based on the results the microbiologist compiles a report that is sent to the infection control unit at the hospital.

The infection control unit consists mainly of nurses and their task is to limit the spreading of infections within the hospital. They work closely together with the different wards at the hospital and has a mainly advisory role, for example they educate the staff in hygiene routines and give advice on actions when an outbreak has occurred. The advice is partly based on the information in the report with the NGS results that they receive from the microbiologists at the lab.

1.4 Research Problem

1928 Diagnostics has discovered that the reporting of the analysis results is surrounded by frustration from both the microbiologists who create the reports and the infection control unit that receives it. The microbiologists are frustrated because they have to add complementary information to the results that is not available in the 1928 platform which leads to extra work when creating the report. This feels like an extra time consuming and big task since the information needs to be collected from several different sources. 1928 Diagnostics wishes to solve this problem by including everything that the microbiologists need for the reporting in

their platform. Since using NGS technology within infection control is a relatively new thing it has to be investigated how the microbiologists work when analysing the results and how the platform can be designed to support those tasks and work flow. One step in that direction could be to make the visualisation of the results more interactive.

Within the infection control unit the frustration comes from the lack of the specific skills needed to understand the test results in the format that they are presented in the platform today. Therefore they rely on the microbiologists to provide a quite substantial explanation of the results. This is also the reason why the microbiologists need to include information from different sources and thus be able to present the information in different ways. By designing the visualisation of the results in a more easily interpreted way it could both make it easier for the infection control unit to understand the information and the microbiologists to communicate and explain it.

1.4.1 Research Question

The research problem described above resulted in the research question below which will be answered by providing a set of design guidelines.

"What should be considered when designing an interactive visualisation of NGS data used for facilitating the creation of reports for infection control in hospitals?"

1.5 Limitations

The project will only consider the part of the application that presents the results of the analysis, which includes the interactive visualisation. The steps the user has to walk through in order to enter the visualisation is outside the scope. It will be assumed that the user can choose a set of samples that will be included in the visualisation. Most users of the platform today use it on a desktop computer. Therefore, the project will focus on designing the interface only for this device. The final design of the static report will not be investigated thoroughly but a quick suggestion will be made.

There are different techniques to analyse NGS data but this project will be limited to cgMLST analysis since it was the only one implemented in the platform, that involves interactive visualisation, when the project began. This analysis technique is applied to several different species and the amount of information available in the result varies between them. Therefore the visualisation will be based on the analysis of the species *Staphylococcus aureus* since its results includes the largest amount of information.

2
Background

This chapter provides information that is useful to have in order to understand certain domain specific aspects of this project as well as related work that has been done in this area.

2.1 Current Visualisation

The current visualisation in the platform presents the analysis results of the NGS data to the user. This section will explain briefly what the information in the visualisation means and how the user can explore it.

2.1.1 Understanding the Information

When performing a genome sequencing the output is a text file with DNA sequences. The DNA sequences are then analysed using bioinformatic methods to find shorter segments of DNA, called genes. The analysis that is performed on the sequenced genome in the 1928 platform is called cgMLST (core genome Multilocus Sequence Typing)[7]. The organisms that are analysed all have a predetermined core genome, which is a set of genes that are likely to be present in all individuals of that species. Those genes are identified as a first step of the analysis.

The function of a gene is to produce a specific protein that is used within the organism. However, the gene that produces a protein for a specific purpose can be different in different organisms, these different genes are called alleles. This means than one organism can have one allele of a gene while another can have a different allele of the same gene. The next step in the cgMLST analysis is to identify what alleles are present in the organisms core genome. When having this information, the core genome from different organisms can be compared in order to see how much they differ in terms of what alleles of the genes they have. The more alleles that are the same in the two organisms the closer related they are to each other, and the number of alleles that differ is used as a measure of distance between them. Based on that information it is possible to make predictions of whether patients have infected each other with a disease and thus causing an outbreak at the hospital. All samples that are within a certain distance of each other are part of a an outbreak cluster.

By analysing the sequenced genome it is also possible to predict what antibiotics a bacteria is resistant to and also what toxins it has. This analysis also identifies different genes and alleles but is done separately to the cgMLST.

2.1.2 Visual Mappings

The genetic distances between organisms are commonly visualised using phylogenetic trees. The phylogenetic tree in the 1928 platform is constructed using a method called UPGMA (unweighted pair group method with arithmetic mean) and is therefore also called a UPGMA tree. The leaf nodes in the horizontal phylogenetic tree in 1928 Diagnostics' platform, in figure 2.1, are coloured blue and represent samples. The numbers beside the grey internal nodes of the tree is the distance between the bacteria in two samples or two groups of samples depending on if the next level consists of leaves or another internal node. Above the tree is a horizontal ruler that indicates the distance measure of the nodes.

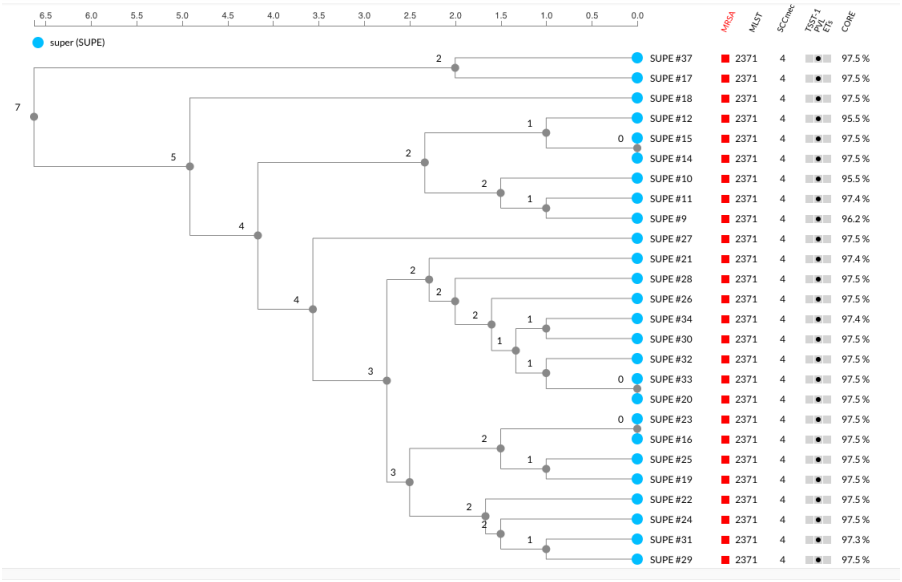


Figure 2.1: UPGMA tree from the 1928 platform displaying information from NGS.

Every leaf node has a label to the right with the name of the sample. When continuing reading to the right from a sample label there are five categories of information visible. The first square indicates if the bacteria found in the corresponding sample is resistant to a specific kind of antibiotics, the square is red when resistant and green if not. The second column consist of the MLST type of the bacteria which describes the collection of alleles the bacteria has in its core genome. The next column holds the result from another kind of typing that is commonly used within infection tracing in hospitals, followed by an array that indicates if the bacteria has any of three toxins. The dark grey dot indicates that that toxin is present in the bacteria. The percentage in the right most column indicates how

much of the core genome has been found and analysed. If the percentage is too low the quality of the analysis suffers or it can not be performed at all.

2.1.3 Interactions

The visualisation is divided upon two pages in the web application. The first page is a list of all the samples that the user has uploaded with the chosen bacteria and the second page displays the samples in a phylogenetic tree with the extra information described in section 2.1.2 on the side. In order to be able to show a UPGMA tree with the clustering information the user has to chose which samples to include from the list in the first page. This gives the user an opportunity to choose which samples to focus on based on the information provided in the table.

The platform provides a few ways for the user to explore the data. When the tree has been generated with the chosen samples the user can zoom into specific areas of the tree by clicking an internal node. To zoom out again the user has to click the back button in the browser. If the user wishes to change the sample collection displayed in the tree they have to go back to the list view and select new samples.

2.2 Related Work

There are a multitude of tools that can be used for visualising genomic data in different ways. Most commonly the relationships are displayed using phylogenetic trees or minimum spanning trees (MST). In this section it will be described how a few tools have used these two types of diagrams. A study in which they redesigned a clinical report of genomic data will also be described.

2.2.1 Phylogenetic Tree Visualisation

Phylogenetic trees can be designed in different ways depending on the amount of data and what it is used for. Two common designs are the horizontal and circular designs. Treelink is a tool for creating phylogenetic trees to be used when, for example, comparing strands of viruses or working with epidemiology. Their main diagrams are horizontal phylogenetic trees and are shown in figure 2.2. The purpose of this tool is to allow the user to generate trees that allows for exploring, comparing, and displaying data that is not directly included in the tree [8]. Therefore the user can add additional information to the leaves by linking the leaf labels to keys in external data sets. The added information is visible in a pop-up that appears when hovering over the nodes.

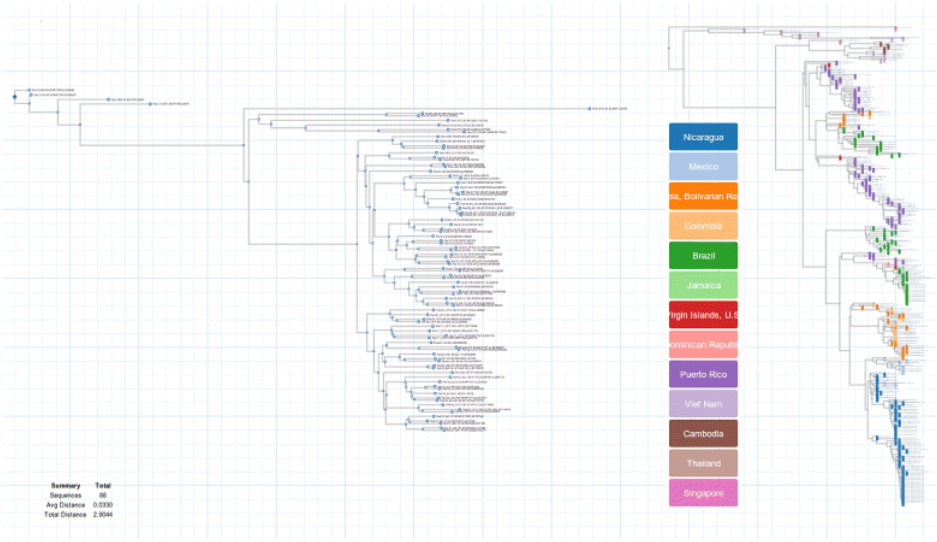


Figure 2.2: Left: Standard tree from Treelink. Right: Annotated version of the tree. From [8]. CC-BY

Another interesting feature in Treelink is that the user is able to colour code nodes and branches based on different attributes. This makes it easier for the user to detect patterns in the data. One property that is colour coded is clusters based on evolutionary distance. When colour coding clusters the entire branches are coloured as opposed to other attributes when only the nodes are coloured. Image 2.3 shows a horizontal phylogenetic tree with colour coded clusters.

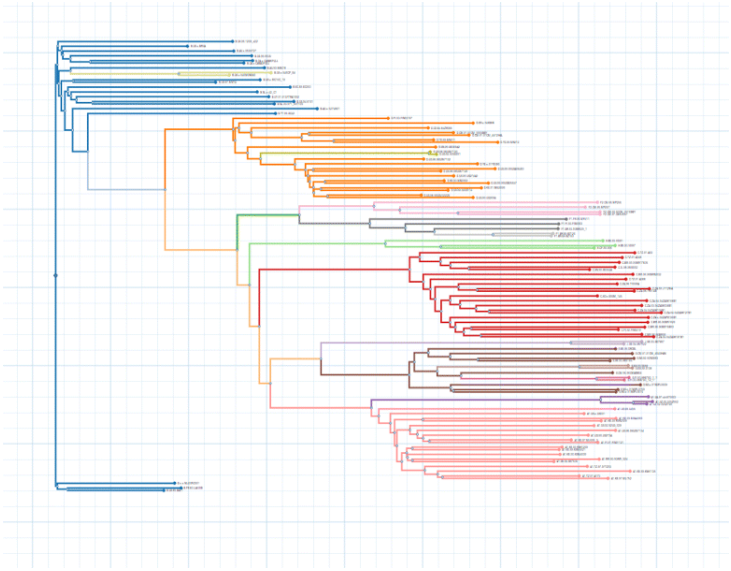


Figure 2.3: Clustering within a tree from Treelink. From [8]. CC-BY

As mentioned earlier a phylogenetic tree can also be drawn in a circular manner. One example can be found in the web application Microreact which is developed for visualising, exploring ans sharing large amounts of data from genomic sequencing within microbial genomics [9]. Except for the circular layout the user can also choose to display it with other layouts such as rectangular, diagonal and hierarchical. The visualisation in Microreact consists of several interconnected parts, namely the phylogenetic tree, a map showing the location data of each data point, a data view and a timeline that shows the temporal data of each data point. Two different map views are shown in figures 2.4 and 2.5 and the timeline can be seen in figure 2.5.

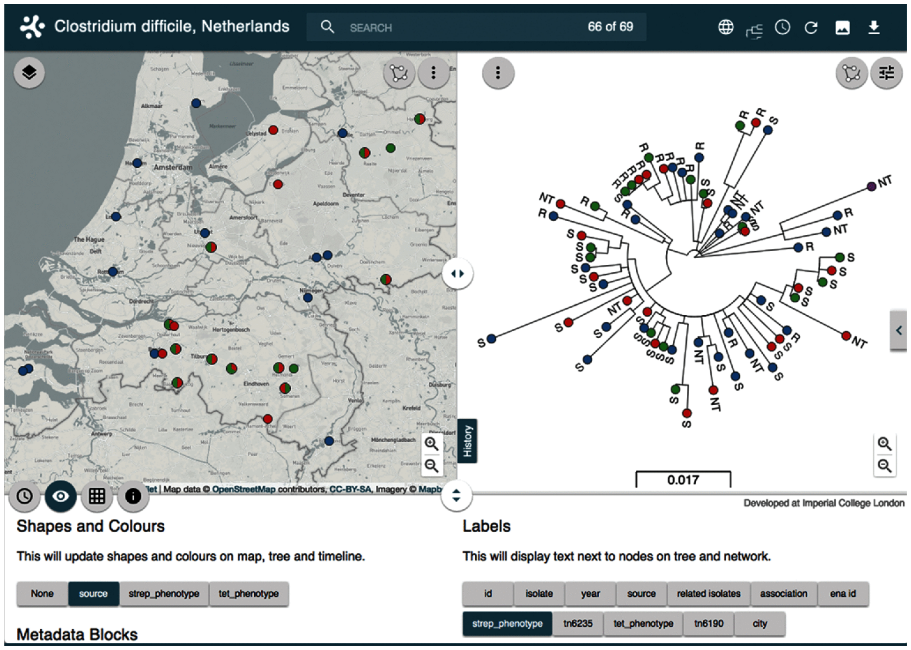


Figure 2.4: Map and circular tree view from Microreact. From [9]. CC-BY/cropped from original

Showing the data points on a map or a time line that is visible next to the phylogenetic tree provides the user with a bigger context which makes it easier to analyse the data. When analysing the data the user can highlight data points by clicking on them. Also, when the user highlights a node in the phylogenetic tree for example, the same data point will be highlighted in the other diagrams as well, in this case on the map and in the time line. Another featuree that facilitates the analysis process is that the user can right click on an inner node and choose to view the branch in more detail and how to display that branch. This is also reflected on the map and in the time line.

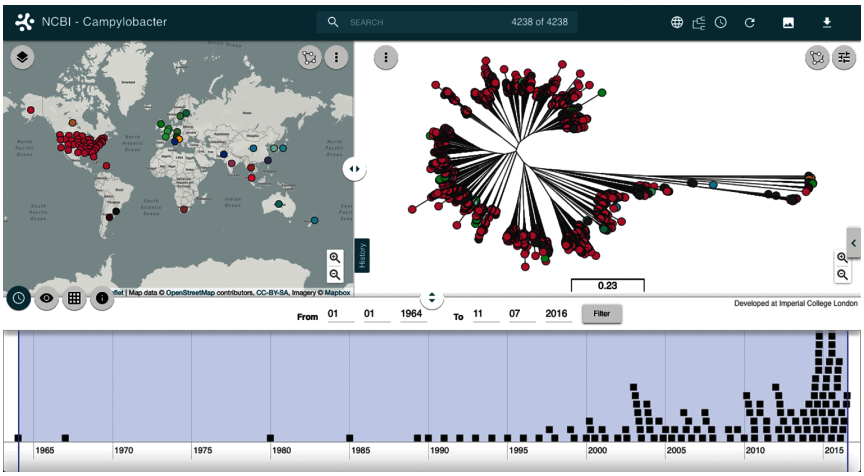


Figure 2.5: Microreact user interface with time line included. From [9]. CC-BY

Tree of Life is another web application that displays its phylogenetic trees in a circular layout. The diagram in figure 2.6 shows one of their circular phylogenetic trees and also that it is possible to add labels to the nodes and work with colour coding similar to Treelink. A circular layout makes it possible to get an overview of a quite large amount of data which is the reason Microreact has it as its default and figure 2.6 is also a good example of that property.

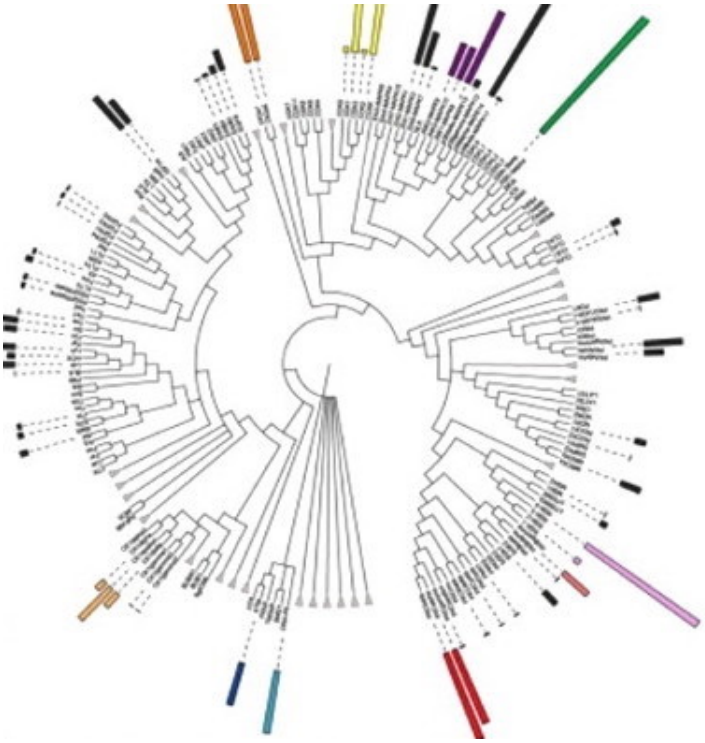
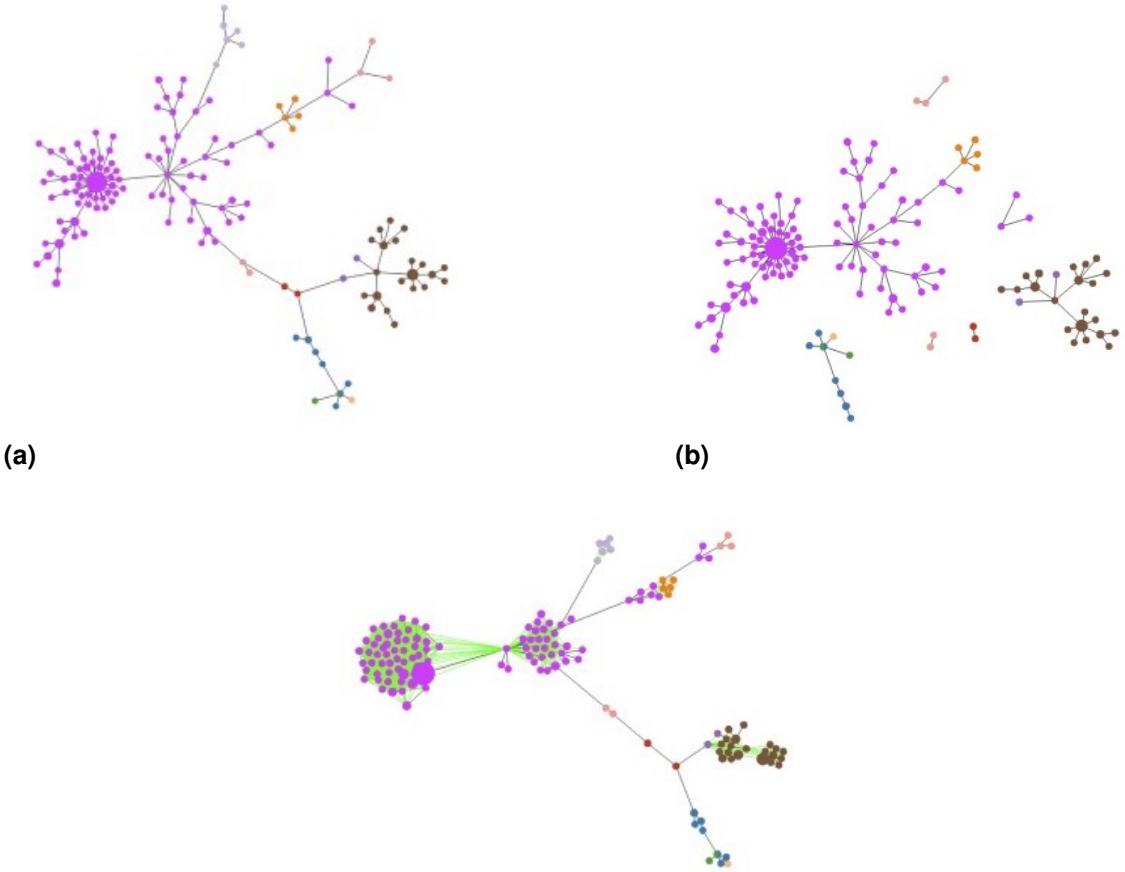


Figure 2.6: Circular phylogenetic tree from Tree of Life. From [10]

2.2.2 Minimum Spanning Tree Visualisation

The benefit of using a MST to visualise genomic data is that it can accommodate a large number of data points [11][12]. It is also scalable and reproducible which makes it suitable for epidemiological investigations and population studies of bacterial pathogens, and thus also for infection control [12]. Phyloviz [12] and GrapeTree [11] have both made interactive visualisations of MSTs for genomic data but the features and the way the user is able to interact with the visualisations differ between them and might be suitable for different purposes.

In Phyloviz the user is able to customize the visualisation in different ways. It is possible to colour code and add labels to the nodes based on different attributes and it is also possible to change the size of the nodes and length of the lines between them [12]. If the colour coded attribute is a distribution within the nodes the nodes represent it with a piechart. The colour coding is visible in all images in figure 2.7.

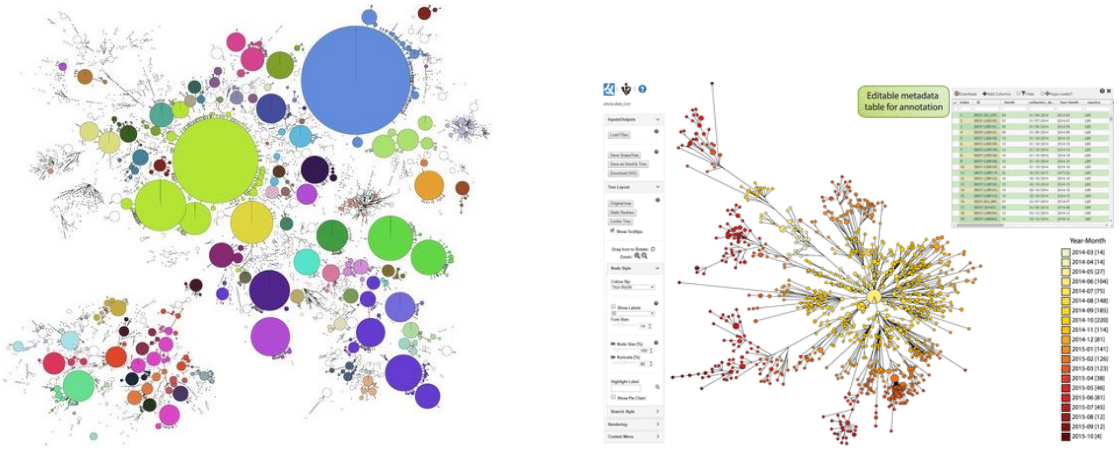


(a) **(b)** **(c)**

Figure 2.7: Minimum spanning trees from Phyloviz demonstrating its different features. From [12] CC BY 4.0

Two features that makes Phyloviz differ from GrapeTree are called NLV (N Locus Variant) and tree cut-off [12]. When using the NLV feature the user can choose a threshold using a slider resulting in all nodes within the length of that threshold value is displayed with all the edges between them instead of the MST. The resulting image would look similar to the graph in image 2.7c. The tree cut-off feature is used in a similar way by using a threshold value. In this case the threshold is used for deciding which edges to remove in order to create the separate graphs. All edges with a length longer than the threshold value are removes from the original MST and the result is similar to the graph in image 2.7b.

GrapeTree has features that are slightly more adapted to very big data sets. In order to visualise a large amount of data in a way that still is easy for the user to interpret they collapse closely related nodes into one bigger node, as shown in figure 2.8a. The user can set a threshold using a slider that decides how closely related nodes have to be in order to be collapsed. GrapeTree offers more features that the user can use in order to customize the appearance of the diagram to make it easier to interpret. Some of them include scaling branches, moving nodes and deciding what property the size of a collapsed node corresponds to. Similar to Phylotree it is also possible to add labels and colour coding to the nodes based on metadata, shown in the images in figure 2.8.



(a) **(b)**

Figure 2.8: Minimim spanning trees from Grapetree. From [11] CC BY 4.0

2.2.3 Evidence-Based Design of a Clinical Report

Crisan et al. have redesigned a clinical report communicating tuberculosis genomic test results using a human-centred approach in order to adapt it for the work flow of the staff at a hospital [13]. The report is a static visualization of information used for diagnostics, treatment and surveillance of infection-spreading. Therefore the study focused heavily on figuring out exactly which information that should be included for these different purposes. Since the doctors that use the report for diagnosing patients need to be able to absorb the information very quickly the amount of information needs to be limited to what is necessary.

During the design phase they used the Design Study Methodology which provides a methodological framework for designing visualisations [14]. The framework consists of nine stages that are grouped together into three phases, namely precondition, core analysis and reflection, but only the use of the three stages in the core analysis stage are described in detail. This phase include discovery, design, and implementation.

In the discovery stage qualitative expert interviews and an online questionnaire were conducted in order to link specific data types to different tasks. This information was then used in the design phase for producing prototypes for different ways of visualising the different data types. The prototypes were designed in an iterative process in cooperation with the University of British Columbia's Information Visualization research group. The prototypes were then evaluated with the help of another online questionnaire by including images of both entire reports and smaller elements in the reports. Based on the results from the second questionnaire and medical test reporting requirements, the final report was designed.

From the interviews and questionnaire it was found that a phylogenetic tree was preferred for visualising the information used for surveillance, i.e. genomic relatedness between samples and their membership in clusters. Apart from including this information it was little consensus among the participants around what data they would want to include for surveillance tasks. From a questionnaire it was found that many participants wished to have more information in connection to the tree, but this is an area that was not focused on in this study and needs more investigation.

The study resulted in a set of guidelines that can be used when designing visualisations for microbial genomic data. The guidelines, divided into experimental guidelines and design guidelines, are listed below:

Experimental guidelines:

- Design around tasks
- Compare isolated components, and not just whole systems
- Compare against a control whenever possible

Design guidelines:

- Structure information such that it mimics a stakeholder's workflow
- Use emphasis carefully
- Present dense information in a careful and structured manner
- Use words precisely
- If using images, do so judiciously

3
Theory

This chapter gives a theoretical introduction to the field of information visualisation and principles that are often adopted in the design process. In addition to the visualisation principles the designer also need to have general interface design practice in mind. A collection of design principles that could be especially important to consider when designing an interface for data visualisation is also described in this chapter.

3.1 Wicked Problems in Design

In contrast to tame problems, which have a clear definition and a solution, wicked problems are hard to define and it is impossible to know if a solution has solved the problem [15]. A wicked problem can be defines as "a form of large-scale social or cultural problem that is difficult to solve because of incomplete, contradictory, and changing requirements" [16]. Design problems are generally wicked since "design has no special subject matter of its own apart from what a designer conceives it to be" [15]. Since the vague design problem statements give the designer a lot of room for interpretation it might contribute to the lack of disciplinary consolidation in the area of research through design [17].

3.2 Information Visualisation

Originally the term visualization referred to the ability to construct an image in the mind. Now the most common interpretation of the word is a graphical representation of a data set that can be used as a tool for decision making. [18]. Knowledge about visual perception and how the human brain processes an image forms a foundation for designing information visualisations. That knowledge has been used for developing different frameworks for information visualisation design. This section provides a brief explanation of visual perception as well as different frameworks.

3.2.1 Visual Perception

The human brain is remarkably good at searching for and detecting patterns in what we see [19]. By studying the science of visual perception it can help designers to make design decisions when designing data visualizations that takes advantage of this ability in order to show patterns in data [18].

A simplified model of visual perception consists of three stages that each focus on solving specific tasks [18]. In the first stage the image seen by the eye is processed to detect low-level properties of different parts of the image simultaneously. Properties that are detected during this stage include orientation of edges, colors, and texture. The second stage, pattern perception, divides the image into regions and simple patterns based on, for example, continuous contours, color and texture. The last stage of visual perception makes visual search, based on queries stored in the long-term memory of the task, possible by holding certain element in the visual working memory. A diagram over the three stages of visual perception is shown in figure 3.1.

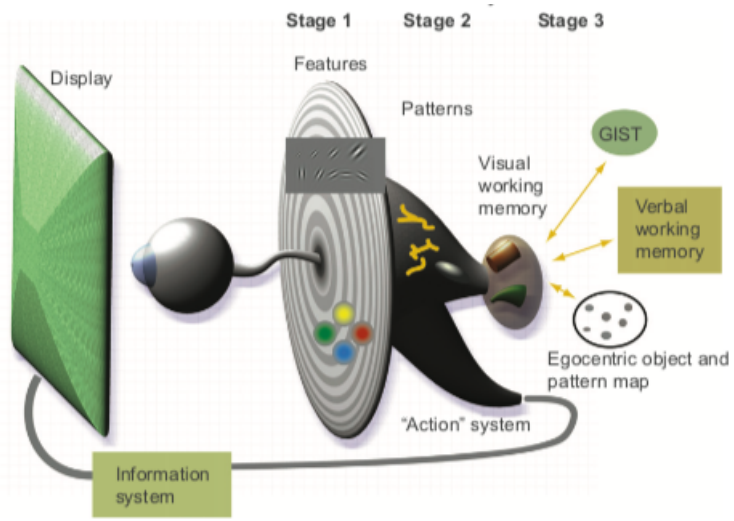


Figure 3.1: A diagram over the three stages in visual perception. From [18]

3.2.2 The Designer-Reader-Data Trinity

Visualisations can be categorised into two overarching categories, exploratory visualisations and explanatory visualisations [20]. When the designer already knows what information lies within the data they will create an explanatory visualisation in order to explain this to the reader. When designing explanatory visualizations it is possible to make design choices in order to highlight what you want the visualisation to communicate and also adapt this to the reader. Visualisations that are designed to convey a specific message but still allow the reader to explore the data are categorised as hybrids between explanatory and exploratory visualisations.

There are three main categories of explanatory visualisations, which are based on the designer-reader-data trinity, namely *informative*, *persuasive*, and *visual art* [20]. The relationships between these entities are shown in figure 3.2. An informative visualisation aims to present the data in a neutral way in order to educate the reader. In this kind of visualisations the information in large data sets is distilled into a more comprehensible form. Persuasive visualizations, on the other hand, are used by the designer to change the readers mind about something.

They usually highlights specific parts of the information available in the data. Visualisations from the last category, visual art, can be difficult for the reader to decode since they often are created by the designer to visualise data in a purely beautiful and interesting way.

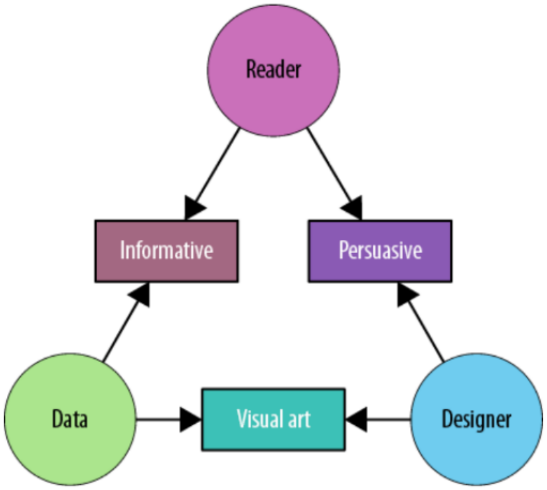


Figure 3.2: The designer-reader-data trinity. From [20]

3.2.3 Information Visualisation Principles

Ben Shneiderman has introduced the visual information seeking mantra "overview first, zoom and filter, then details on demand" for guiding the design process of information visualisations, which follows the structure of how the human brain processes visual information as described in section 3.2.1 [21]. Based on the mantra he also constructed the *Type by Task Taxonomy* which defines seven tasks of high abstraction that a user is likely to perform within a visualisation. The tasks are *overview*, *zoom*, *filter*, *details-on-demand*, *relate*, *history* and *extract*. Listed in this order the tasks can be seen as a user flow through a visualisation application.

According to the visual information seeking mantra the first thing a user looks for in a visualisation is a good overview of the data set. This is quite easy to achieve for small data sets but poses problems for large data sets that can be difficult to solve. To show every data point of a large data set with its own visual entity results in abundant visual noise that makes it difficult for the reader to get the desired overview of the data and understand what it is trying to say. By grouping the data points and create a visual entity that represents the entire group, Elmkvist and Fekete suggests that it is possible to design a visualization that provides a good overview of the entire data set [22]. They call visual entities that represent a group of data points *visual aggregates* and the purpose of them is to convey information of the underlying data while reducing the number of visual entities in the visualisation. The information that a visual aggregate conveys can be, for example, the number of data points it represents, the average value of the data points or a distribution.

3.3 Interface Design Principles

The area of graphical user interface design has many principles and guidelines that the designer can benefit from following. In this section a number of principles that could be especially relevant to designing visualizations will be explained.

3.3.1 Affordance, Signifiers and Pliancy

Affordance is a concept commonly referred to in the design community and is considered both for designing physical objects and digital interfaces. Donald Norman introduced the concept to the design community in his book *The Design of Everyday Things* and defined it as "the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used" [23]. However, in a later edition of the book he revised this definition to focus more on the relationship between the user and the object and that what actions with an object are perceived as possible depends on the user's own abilities and knowledge [24].

For design it is important to remember that affordances do not communicate what the user should do with an object, but what the user perceives as possible to do with it. Norman introduces signifiers as a concept for defining the properties that communicate how the user should interact with an object [24]. He also suggests that signifiers actually are more important in design than affordances, since a usage can be perceived without actually being intended. This type of affordances are called perceived affordances and is a concept also discussed by Alan Cooper in *About Face* [25].

Cooper highlights that affordances in a digital user interface can easily fool the user since there is no natural connection between what is seen on the screen and the functionality that lies behind it [25]. Therefore it is important to fulfil the users expectations of the affordances of an object. For example, if an object is perceived as a clickable button it should be visibly clear that is is being pushed when the user clicks on it. In order to describe this the term pliancy is introduced, which is similar to Norman's signifiers. An object is pliant if it reacts to input and the user can manipulate it. In order to fulfil the expectations of affordances it is important for the designer to visually communicate the pliancy to the user. This can be done by:

- Static hinting - communication by static rendering of the object itself
- Dynamic hinting - temporary changes in appearance of an object when hovering over it
- Pliant response hinting - changes in appearance of an object when cursor is clicked but not released
- Cursor hinting - cursor changes appearance when hovering over an object

3.3.2 Colour

Colour is an important attribute of an interface, it plays a big role in if the user perceives it as pleasant or stressful for example. Different colours are also perceived as meaning different things, for example is green generally something good and red is a warning [25]. But these associations can differ between different cultures which should be considered when designing an interface, especially if the colour is used for communicating a specific message. If the colour of an element conflicts with the message that the element communicates cognitive interference can arise, which increases the time it takes for the user to understand the meaning of it [20].

When designing data visualizations colour can be a powerful tool for visual encoding of certain types of data [18]. It is especially useful for classifying visual symbols into separate categories. In order to assure that this quality is preserved the designer should not use too many colours, which implies that if there are too many categories to visualise more encoding properties should be used. A general rule of thumb is to use the six colours on the left hand side in figure 3.3 first when classifying using colour, and after that use the six colours on the right hand side in figure 3.3 When assigning colour to a data type it should be considered that colour is not naturally ordered, which means that there is no natural order of colours that everyone would understand, as numbers have for example [20].

Figure 3.3: Recommended colours for classifying visual symbols into categories. From [18]

The first things to consider when choosing colours are the users' goals, environment, the content, and the brand of the product [25]. When suitable colours for those purposes has been chosen value, hue, and saturation should be considered. These have different qualities that can be beneficial for certain things. Saturation quantitative, which means that a greater saturation is perceived as a higher value than a lower saturation, and higher contrasts between colours makes certain elements stand out.

3.3.3 Gestalt laws

The Gestalt laws is a set of laws that provide a clear description of many basic perception phenomena, specifically pattern perception [18]. These laws can serve as design principles for organising data in such a way that important patterns are easily perceived by the human brain. There are eight Gestalt laws but in this section only the six gestalt laws that do not consider motion perception will be described.

Proximity: The law of proximity is one of the most important for design [18]. It states that visual elements that are spatially close together are perceived as a group. A similar concept is spatial concentration which describes the phenomenon that the human brain group regions of similar element density. With this in mind, the designer should place visual elements that represent related information close together when designing visualizations.

Similarity: Another law that explains how the human brain perceives visual elements as groups is the law of similarity [18]. If elements have similar shapes or colours they are perceived as being part of the same group.

Connectedness: Connecting two or more elements with lines is a powerful way of visually express a relationship between them. This is the principle of the law of connectedness. Connectedness can be a stronger visual grouping principle than, for example, proximity of similarity [18].

Continuity: The law of continuity states that the human brain is more likely to "construct visual entities out of visual elements that are smooth and continuous, rather than ones that contain abrupt changes in direction" [18]. This principle can be used, for example, when it should be easy to identify sources and destinations of connecting lines.

Symmetry: The law of symmetry serves as a powerful organising principle and is useful for comparing two different data sets [18]. In order to fully take advantage of symmetry the important patterns need to be small since we are most sensitive to such symmetrical patterns [26]. When designing a visualization for comparing data sets it should be considered whether the patterns are small enough in terms of visual angle, in order for symmetry to be beneficial.

Closure: The law of closure describes that a closed contour usually is perceived as an object [18]. If a contour has a gap in it the human brain tends to close that contour and still perceive it as a whole object. This principle is useful for visualising sets and is commonly used for euler diagrams.

4

Methodology

The iterative model that the design process of this project was inspired by is described in this chapter. In order to adapt the design process to the area of information visualization the methods in the phases were chosen with the findings from the literature study in mind. This includes a user-centred and task oriented focus. The methods used in the project will also be described in short in this chapter.

4.1 An Iterative Design Process

The design thinking model is a solution focused approach to design that is especially useful when dealing with complex and ill-defined problems [27]. Such problems are, as described in section 3.1, wicked and are very common in the area of design. The Design Thinking model consists of five phases named Empathise, Define, Ideate, Prototype and Test, as shown in figure 4.1.

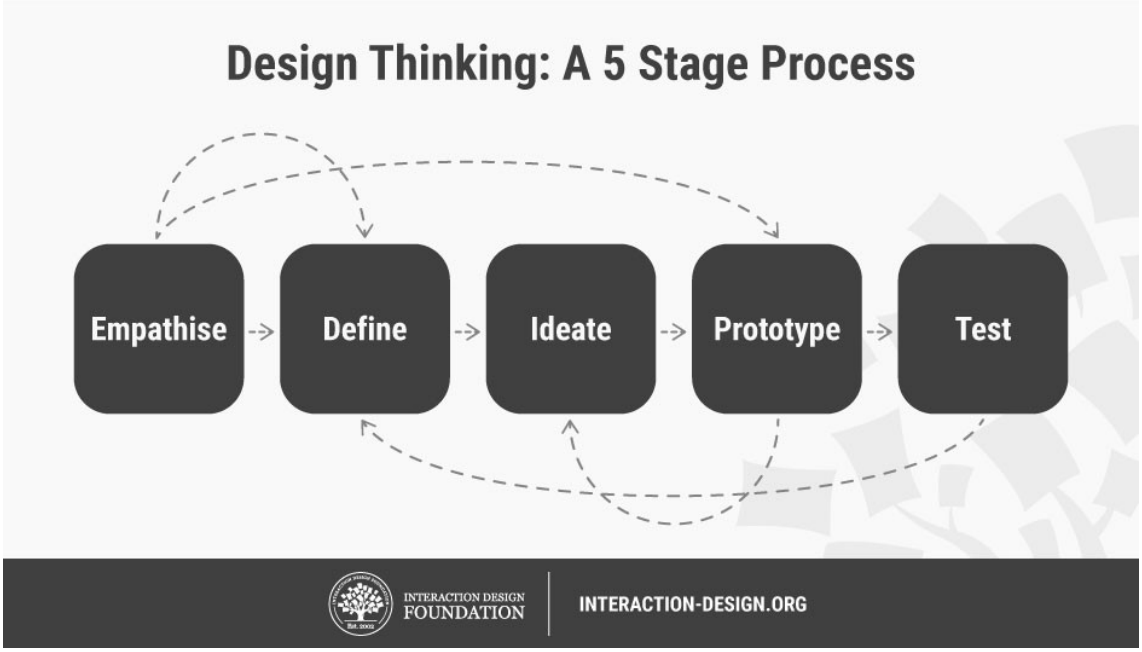


Figure 4.1: Diagram explaining the flow through the phases of the Design Thinking Process, from [27].

The Empathise stage aims to develop a deeper understanding of the problem to be solved and also the motivations and goals of the users. The insights from this stage will then be analysed in the Define stage and serve as a foundation for defining the core problems that need to be solved.

When an understanding of the users has been developed and the problem defined the next stage is Ideation. During this phase new solutions to the problem are brainstormed and also exploring new ways to look at the problem. The ideas from the Ideation stage are then brought to the Prototyping stage where they will be transformed into inexpensive prototypes. When a number of prototypes have been developed they are tested within the company to find things to improve upon. The Ideation and Prototyping phases can be iterated over a couple of times until reaching a prototype that is ready for the testing stage. In the testing stage a more rigorous testing is done, if possible with real users.

The model is designed so that it can be used in an iterative manner, which makes it possible to redefine the problem during the process. This quality is especially valuable when working with a design problem that in nature is hard to define. As shown in figure 4.1 it is possible to iterate over several of the phases. What to iterate over can be decided during the project when new needs have been discovered.

4.2 User-Centered Design in Information Visualisation

When practicing user-centered design the designer puts the user of the product in the center throughout the entire design process [28]. From the definitions of information visualisation and user-centered design it can be concluded that the design process for designing an information visualisation should be user-centered [29]. Mainly because the purpose of a visualisation often is to explain something to the user or provide a way for the user to explore the data themselves. In order to do this efficiently the designer must focus more on certain things than when designing other types of interfaces, for example the users' prior knowledge [30]. The users' prior knowledge in how to operate the device in question or the components in the user interface needs to be considered in any design process, but when designing an visualization another category of prior knowledge to consider is the domain specific knowledge.

Zhang et al. have constructed a framework for applying user-centered design in the design process for information viaualizations. The framework includes four stages, namely functional analysis, user analysis, task analysis and representational analysis. The goal of the functional analysis is to identify critical domain relationships and goals [29]. The structure of tasks and information flow is also determined here. The next phase, user analysis, intend to identify the characteristics of the users. An important goal of this phase is to understand the prior

knowledge of the users, but also in what context they are likely to use the visualization in terms of, for example, time available. The task analysis aims to identify procedures and actions to be carried out in order to achieve task goals. This is done to ensure that no unnecessary features are included in the visualization since that would contribute to extra cognitive load for the user. The last phase, representational analysis, aims to identifying the best way to display the information and the best information flow structure while considering different users and their needs.

4.3 Methods

The methods described in this section are categorised into the phases of the Design Thinking Model described in section 4.1. The selection of methods has also been inspired by the user-centered design process for information visualisations described in section 4.2 in order to tie the two frameworks together.

4.3.1 Empathise Methods

Semi-structured Interviews

The purpose of conducting semi-structured interviews is to collect qualitative data about the users. They follow a predetermined structure but allow for some flexibility in order to be able to shape the interview more as a conversation. Interviews are good for investigating users' general attitudes toward using a product and how they think about a problem rather than specific design choices [31]. Well designed questions can give valuable insights that can guide the define phase of the design process by, for example, providing information to the personas.

Observations

By observing the users in their everyday setting doing their everyday tasks the designer can get their own view of the users behaviour, rather than having to rely on the users descriptions [32]. Observations can be organised with different levels of participation of the researcher, with the "fly on the wall" method on one end of the spectrum and full participation on the other. For this study a place focused observation is most suitable, which means that the researcher observes the events and how people act at a specific place. The knowledge gained from the observations can be used for guiding interview questions and keep them relevant to what the researcher has found they need more information about.

4.3.2 Define Methods

Personas

A persona is a fictional character that is developed based on data acquired from user research [33]. By constructing one or more personas at the beginning of the project, it helps the designer to understand the users better and also to step out of their own values and opinions and look at the product from another point

of view. The goal-directed persona is most suitable for the task oriented design focus often implemented for information visualisations. The aim of creating a this type of persona is to investigate what workflow your users would prefer when working towards satisfying their goals they have with using your product.

Hierarchical Task analysis

Conducting a hierarchical task analysis can help the designer to understand the tasks the user needs to perform in order to reach their goals [34]. The analysis starts by defining all high level tasks for a specific persona, and for each of these tasks sub tasks are defined. Sub tasks can be defined at as many levels as needed in order to fully understand the high level task. By breaking down each task in this way the analysis provides a way to define design problems as well as evaluate the design at a later stage.

4.3.3 Ideate Methods

Brainstorming

Brainstorming in a group is a common way to produce ideas for solving a problem, but some studies suggest that individual brainstorming often result in more and better ideas since it is easier for a individual to stay focused on the task at hand [35]. When placing people in a group there are many psychological phenomena taking place, such as conforming to group opinions and the participants own social inhibitors, which is not present in individual brainstorming [36]. Regardless the type of brainstorming it is important to define the task carefully, so that it is specific enough to keep the participants focused but vague enough to promote creativity.

Sketching

Sketching can be defined as rapid freehand drawing used for expressing ideas and preliminary designs [37]. Due to the nature of the sketching process it is mainly used for exploring ideas and concepts rather than focusing on design details . By providing a visual dimension to ideation, and thus incorporating another sense into the thinking process, sketching improves creativity and makes it easier to build upon previous ideas. Therefore, it can be seen as part of the thinking process.

4.3.4 Prototype Methods

Low fidelity prototyping

Early in the design process it is important to focus on the cohesiveness of the interface and to accommodate the users needs. To test these qualities it can be helpful to create low fidelity prototypes so that the designer do not focus too much on the design on specific widgets and controls [**?**]. Paper prototyping is a common method for creating low fidelity prototypes, where the designer creates a paper version of the interface that can be used for early usability testing [38].

High fidelity prototyping

In contrast to low fidelity prototypes, high fidelity prototypes are used for communicating the final look and feel of the design [37]. This type of prototype is important for evaluating the usability of the design since all details and interactions are included. For this reason it can also be used for usability testing towards the end of the design process.

4.3.5 Test Methods

Usability testing with think aloud

It is important to evaluate whether a design provides a good solution to the problem it set out to solve. In a user-centered design process usability testing, with real users as the participants, will provide valuable insights about how easy it is to use the product and how the product is perceived by the users [28]. When the prototype is refined enough usability testing can identify major interaction problems and naming issues, but it is difficult to test usability beyond how easy it is to use the product the first time [**?**]. A good method for extracting as much information as possible from the usability tests is to ask the participants to think aloud when performing the tasks. This gives the designer an idea of exactly what is problematic in the interface.

5
Planning

The project was planned to span over approximately 19 weeks during the spring of 2019. This chapter will give an overview of the planned phases, tools used and a time plan.

5.1 Process

The planned process of the project included four phases which were based on the methodology framework described in section 4.1. The first phase included planning the project, background research and writing the planning report. Weeks 4-7 were dedicated to this phase. This amount of time would give a good foundation for the rest of the project. The next phase, empathise and define, involved user research and persona development which was planned to be done during weeks 8-10.

The third phase was the iterative concept development phase, which was divided into three iterations. Each iteration lasted for 2 to 3 weeks and weeks 11-17 were allocated to them. During this iterative process the guidelines were also planned to be developed.The last phase involves the report writing and presentation preparations. It is different from the other since it lasted throughout the entire project. The last five weeks were dedicated to this phase only.

5.2 Tools

Since one of the main topics of this project was data visualisation, a tool for prototyping diagrams in a good way was needed. For this purpose D3 was chosen [39]. D3 contains many building pieces that allow the user to shape their visualisation as desired. It was planned that, for prototyping the interface as a whole, a user interface design tool would be used, such as Sketch [40]. In order to convey the concept better to different stakeholders interactivity would be introduced to the prototype using InVision [41]. Other tools that were used in the project were Google Drive for collecting relevant files, Google Forms for questionnaires and LaTeX for writing the mandatory reports.

5.3 Time Plan

The phases described in section 5.1 above, were planned to follow the time plan in the gantt chart in figure 5.1.

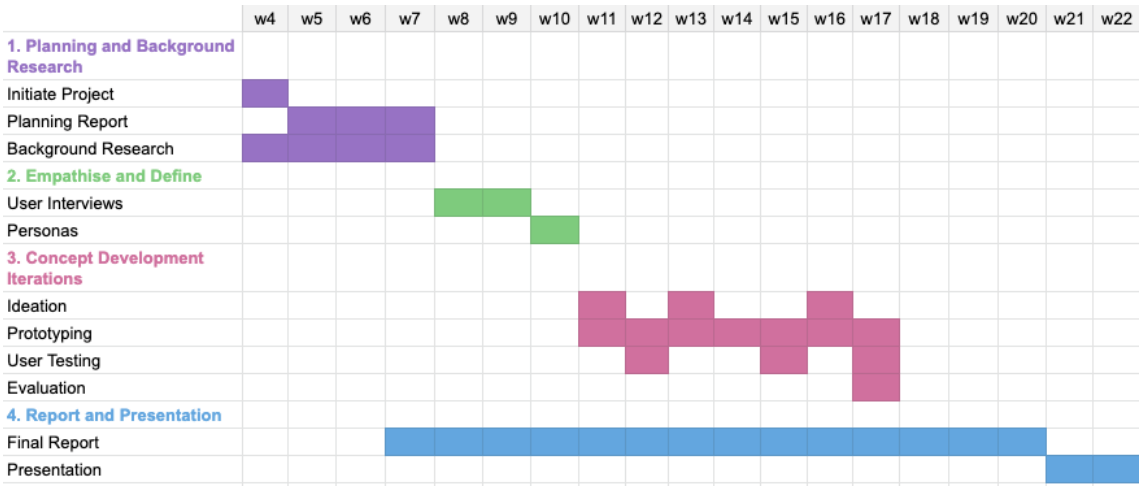


Figure 5.1: Gantt Chart

6
Execution

The project was divided into a pre study, an user empathising phase and three concept development iterations. The first iteration focused on concept ideation and a first draft of the concept. The second iteration focused on the design of the diagrams visualising the data. The third and last iteration refined the results from the previous two iterations and evaluated the concept. This chapter explains what was done in each of the development phases and also the results from each concept development iteration.

6.1 Pre study

The first part of the pre study of this project was a literature study that aimed to explore what has previously been done in the field. The literature study was followed by a review of infection control report examples to see what is usually included. This section will explain how the pre study was conducted and also the findings from studying the reports.

6.1.1 Literature study

Before starting the project a literature was carried out with the focus of finding the best methodology for designing an information visualisation application. Different ways of displaying phylogenetic trees and minimum spanning trees was also researched. The findings of the literature study is presented in chapter 2, Background, chapter 3, Theory, and chapter 4, Methodology.

Mainly the Google scholar and Chalmers Library databases were used for searching for literature. Examples of keywords used for search:

- Information Visualisation
- Design methodology + Information visualisation
- Design principles + Information visualisation
- Phylogenetic trees + Visualisation
- Genomic data + Visualisation

6.1.2 Review of Reports Used for Infection Control

Two example reports from hospitals and the report developed in the project explained in section **??** were included in the review. Two of the reports are focused around one patient while the third one reports the result of a group of samples.

For the reports that reports the result for one patient the patient information is included and the sample from the patient is compared to recently sequenced samples at the hospital. The relationships between the current sample and the previous samples are shown in different ways in the two reports, one uses a MST and the other uses a phylogenetic tree. In the report that includes a group of samples both of the diagram types are included. The phylogenetic tree is used for showing the relationships with a larger group of previously sequenced samples while the MST is used for showing a close up of the relationships of the samples within an outbreak.

Two of the reports included an interpretation of the results in text in order to explain the diagrams to the reader. The interpretation texts were included in the reports that also included a MST. In the reports that included a phylogenetic tree there was an explanation for the relatedness intervals, i.e. below a certain distance the samples are very likely to be part of the same outbreak and in an interval above that the samples could be part of the same outbreak. Other information in the reports include typing information, date, drug susceptibility, and method explanation.

6.2 Empathise and Define

Before starting to ideate and sketch on solutions, some time was spent on trying to understand the users and their goals. Based on information obtained through observations and interviews, personas were constructed for guiding the design process. In order to further define the design problem a hierarchical task analysis was performed. In this section the process and findings of this initial phase are described.

6.2.1 Interviews and Observations

1928 Diagnostics has conducted a couple of documented user interviews prior to this project in order to advice the product design in general. Therefore, the first step in the user research was to read transcripts and takeaways from these interviews. The interviews were carried out in a semi structured manner with a variety of user types. Since the purpose of these interviews was to collect knowledge about the whole platform they touched upon every part of the products user interface which resulted in a shortage of information regarding requirements on the visualisation the visualisation and reporting procedure.

After studying the existing user research made by 1928 Diagnostics the areas that lacked information important to this project was complemented by conducting two extra user interviews. The interviews were conducted at the clinical microbiology department at Jönköping hospital with one microbiologist and one nurse that works at the infection control unit. Questions to guide the interviews were constructed beforehand in order to make sure that all identified knowledge gaps

were filled. They aimed to be strategic, or whole-cycle, in order to keep the focus on why things are as they are and how the interviewees feel about that rather than only how things are [42]. The interviews were carried out in a semi structured manner to allow the interviewees to expand upon areas that are important to them.

In connection to the interviews at Jönköping hospital there was a tour of the lab explaining how they work and what kind of tasks are performed there. Seeing the environment in which the users work provided valuable insights to their way of thinking about their work. When walking through the lab it was apparent that the work there is very logistic in nature which means that there is a need for structure and routines. From the interviews it was explained that sometimes things out of the ordinary happens which forces the personnel to address that immediately, but mainly there are structured workflows that everyone follows.

6.2.2 Personas

From the user study two personas were identified to be important in this project, a microbiologist that is the main user of the visualisation and an infection control nurse that is the recipient of the report constructed by the microbiologist. Since the infection control nurse does not interact with the product directly this persona was defined as secondary. It was decided that it is important to include the infection control nurse since the resulting report must be designed so that they can understand the information. The primary persona is constructed based on the users experiences working with the existing product, whereas the secondary persona is more general and not connected to the product.

The primary persona, in figure 6.1, Alexandra Persson has a PhD in microbiology and works as a microbiologist in the hospital lab. Her work involves doing research and more advanced analyses such as NGS. When NGS has been requested by the infection control unit on one or a group of samples it is her job to perform the NGS and the following analysis of the result. When she has analysed and explored the data she wants to construct a report with her findings that the infection control unit can understand and use in their work. Since she has many work tasks she wants to perform this process as efficiently as possible to save time and the general attitude is that the 1928 platform can help her do that. Right now she finds it frustrating the the platform does not handle metadata, such as place and time for the samples, and that she needs to do additional work, such as inserting information from other software, in order to compile a good report.

The secondary persona, in figure 6.2, Alex Granqvist is a trained nurse and works at the infection control unit at the hospital. The main function of the infection control unit is to prevent diseases from spreading within the hospital which involves a large variety of tasks. Alex's main task is to support hospital wards and clinics in their work with infection control. Which he can do through educating the personnel and also take action when an outbreak has been detected by communicating with relevant persons and advice on what they should do in order to

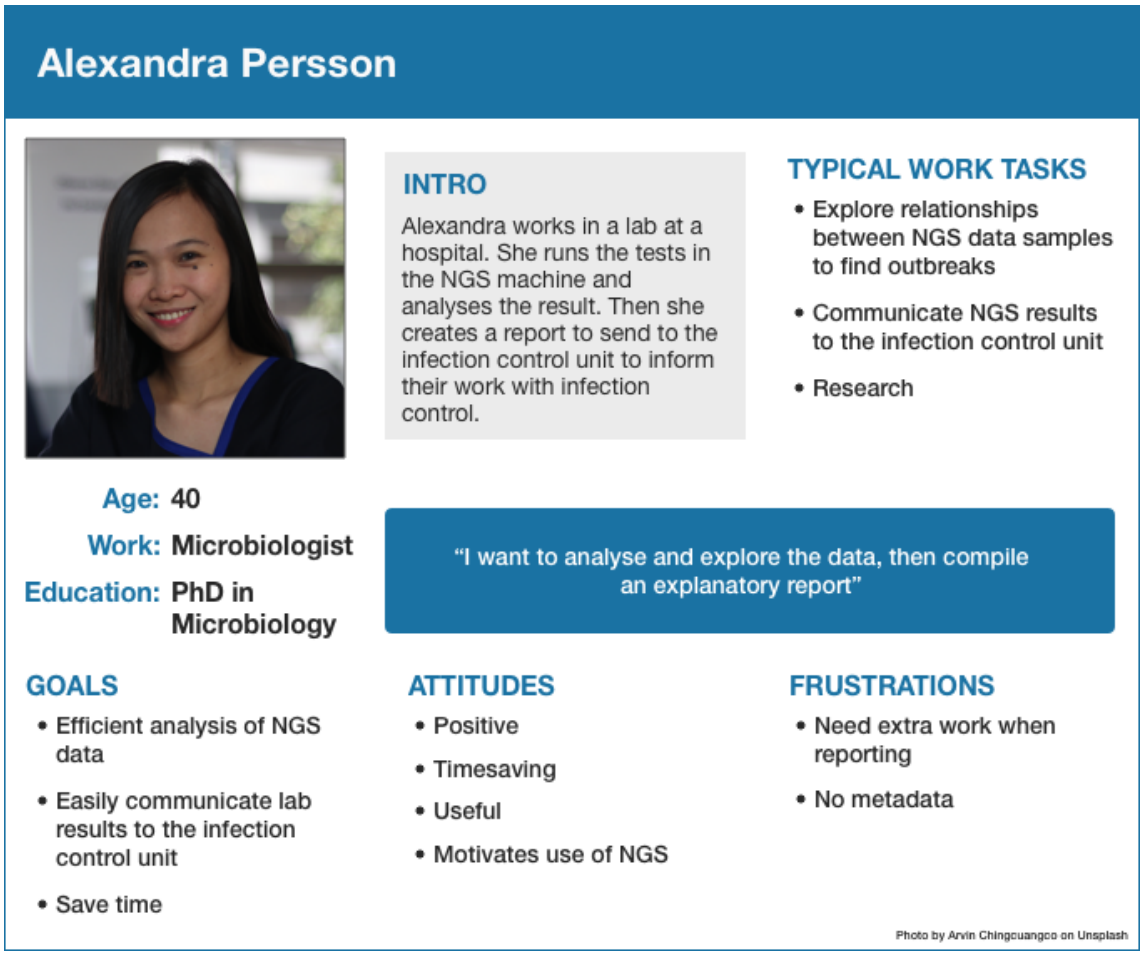


Figure 6.1: The primary persona Alexandra who works at a microbiologist at a hospital.

prevent further spreading. It is important to Alex that this communication is good and efficient and that he can base his advice on test results. Since NGS is an advanced technology Alex is frustrated that he does not understand the test results and he feels like he lacks time to learn the details of how to interpret them. Another frustration he has is that there are many guidelines set in place at the hospital that he has to follow which can prevent him from using the information in the best way possible. A consequence of this is that he is very process oriented and follows the same process every time an outbreak is detected and that he also is hesitant towards big changes in his work flow.

The relationship between the two personas greatly affect how the report should be designed and thus also the visualisation. From the user research it was found that this relationship is greatly affected by how the hospital has organised the work of the infection control unit. Two main organisation models were identified. In the first case the infection control unit works closely together with the lab, they are often located in the same place in the hospital. In this case there is a lot of personal contact between the lab personnel and the people in the infection control unit. This means that the person that performed the tests and analysed the results

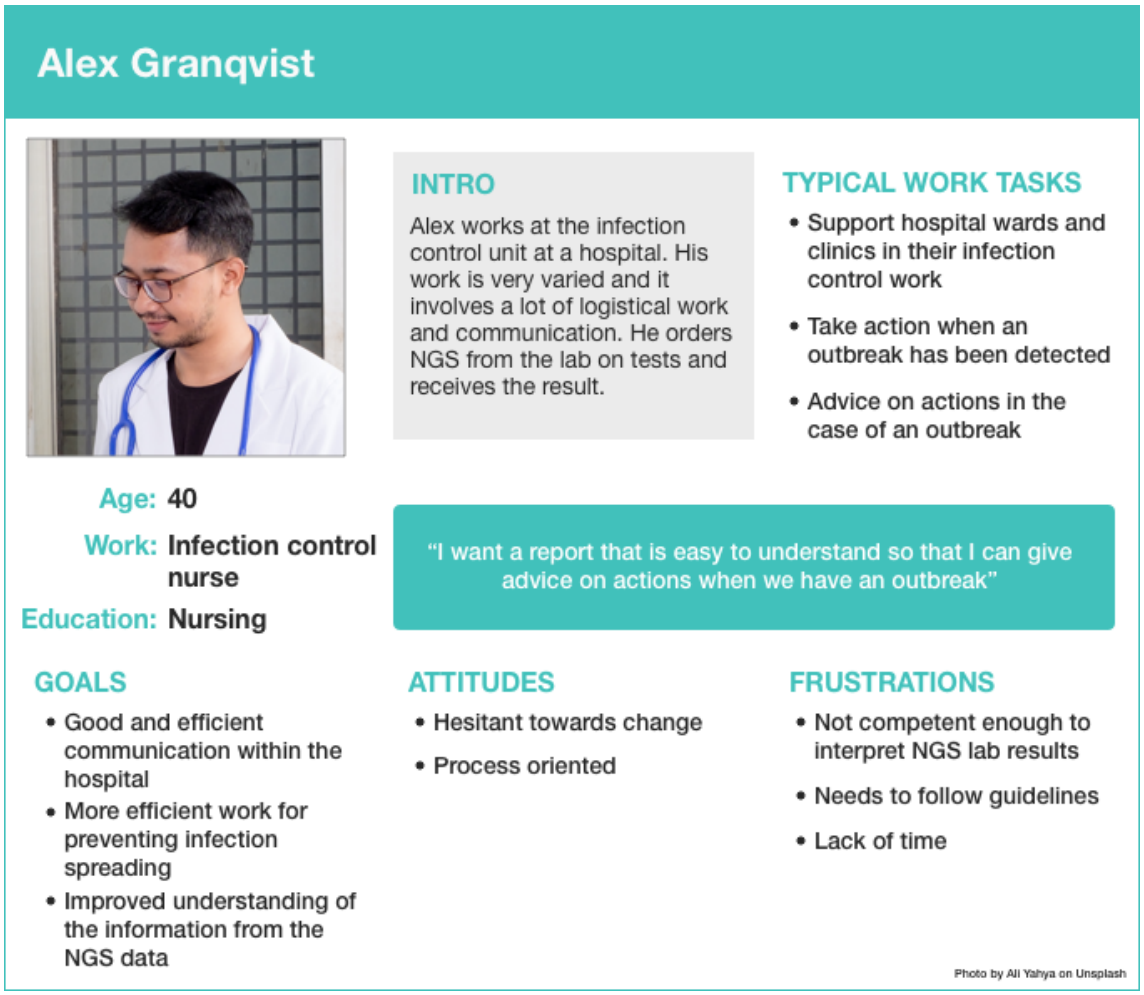


Figure 6.2: The secondary persona Alex who works at the infection control unit.

can explain the result to the infection control unit in person. In the second case the infection control unit is detached from the hospital and is located in different facilities from the lab. This means that the personal contact and relationships are missing and often results in misunderstandings between the two entities. In this case the lab must construct a report with all the information explained so that the infection control unit can understand it from reading. Since the results often are nuanced and complicated the lab simplifies it in order to not make the infection control unit frustrated from not understanding it. This can also be interpreted as patronising by the infection control unit.

6.2.3 Task analysis

To understand the work flow of the primary persona, Alexandra, better, the tasks that she wants to perform in the visualisation was broken down in a hierarchical task analysis [34]. The purpose of doing this analysis in this stage was to define a work flow of the user that could serve as a starting point for ideation. The analysis was performed in a workshop together with two representatives from 1928 Diagnostics that have good contact with the customers and a good idea of

what the customers wishes to do with the product. The tasks were discussed and written down on post its which then were placed in the right order, shown in figure 6.3.

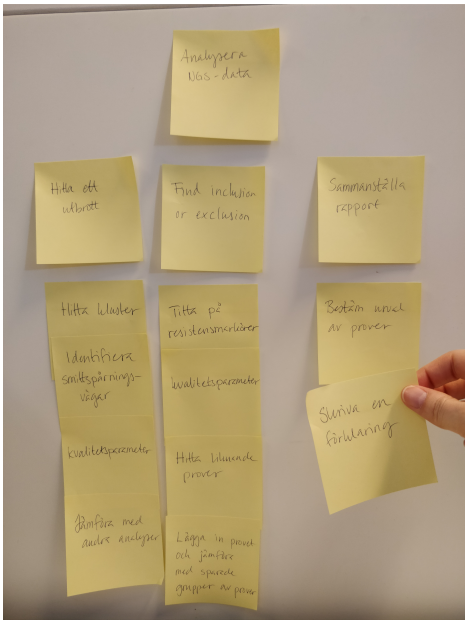


Figure 6.3: Diagram of the tasks defined through the hierarchical task analysis.

The main task that Alexandra wants to perform is to analyse the NGS data, this task is placed at the top in figure 6.3. In the workshop three sub tasks of the NGS analysis were identified; *find an outbreak*, *find inclusion or exclusion* and *compile report*. After further analysis the task *compile report* was moved up to the top level since it is was found to be a separate task. The final structure of the task analysis can be seen in figure 6.4.

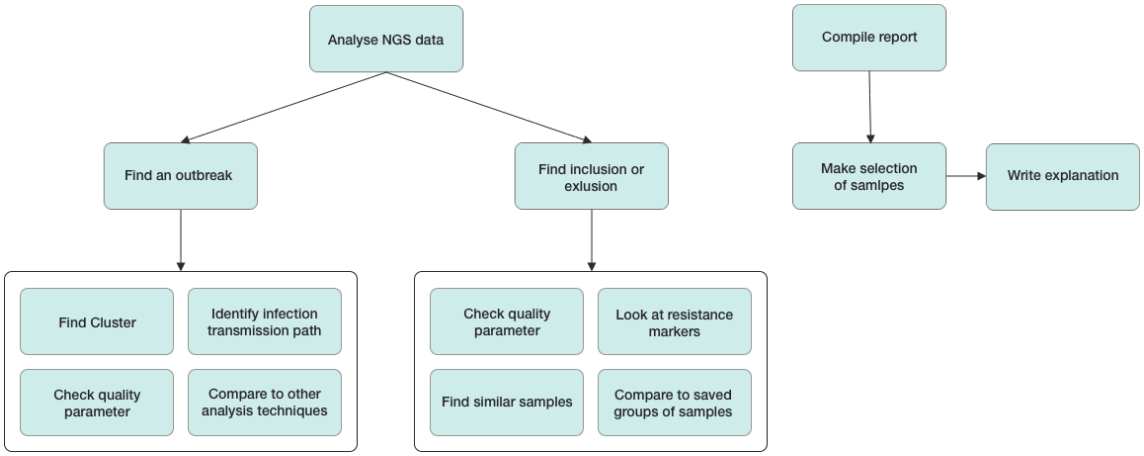


Figure 6.4: Diagram of the tasks defined through the hierarchical task analysis.

The task *find an outbreak* was broken down into four sub tasks that can be performed in any order; *find cluster*, *identify infection transmission path*, *check quality*

parameter and *compare to other analysis techniques*. This task flow represents the activity of identifying outbreaks in a large group of samples. When identifying outbreaks Alexandra looks for clusters that can represent an outbreak and then supports this theory by looking at other parameters, such as the quality parameter of the samples and compare to other analysis techniques. When an outbreak has been identified the user wants to identify transmission paths in order to better understand the nature of the outbreak.

The task flow *find inclusion or exclusion* represents the activity of determining if one or a small quantity of samples belong to an already identified outbreak or not. Four sub tasks were identified that can be performed in any order; *find similar samples*, *compare to saved groups of samples*, *look at resistance markers* and *check quality parameter*.

The second top level task is *compile report*, which was broken down into two sub tasks; *make selection of samples* and *write explanation*. This task flow represents the creation process of the report that is sent to the infection control unit. Since the infection control unit is not interested in all existing samples Alexandra must select the relevant samples first and then provide an explanation of her findings.

6.2.4 First Version of Guidelines

Based on the literature study and the insights from the empathising phase the first version of the design guidelines were developed.

1. Focus on tasks Design around the tasks performed when analysing the results for infection control, including outbreak identification and finding transmission routes. Since the product is a tool for microbiologists in their daily work it is important that they can perform all necessary tasks without distraction [29]. This approach was also successfully tried in the clinical report design project explained in section **??**.

2. Adapt to the Reader Have the recipient of the report in mind when designing the appearance of the diagrams in the visualisation. From the secondary persona in section 6.2.2 it is apparent that the visual language needs to help the infection control unit to interpret the results. This could be achieved by using different gestalt laws and colour to highlight and emphasise different characteristics [18].

3. Flexibility Since there are different infection control organisation types, explained in section 6.2.2, the tool should be flexible enough to suit different ones. From the report review in section 6.1.2 it can be seen that the reports are made differently in order to to suit the hospital's different workflows.

4. Affordance and Pliancy Work with affordance and pliancy in the interface in order to help the user understand how to interact with the visualisation. This is a common interaction design principle that should be followed [25][24].

5. Good Defaults Let the user enter the visualisation with good default values entered. Since, according to the primary persona in section 6.2.2, time is a limited

resource having good defaults value might enable the user to make the analysis faster.

6.3 First Iteration

The first iteration of the concept development phase focused on finding possible solutions for the tasks in the hierarchical task analysis. The design space of each task was explored and put together in a first iteration of the whole concept in a paper prototype. The prototype was tested with the focus on the navigation between features and the use of metadata in the visualization.

6.3.1 Ideation and Concept Development

The first stage of the ideation process was a brainstorming session that I carried out alone. The brainstorming was focused around one task at a time from the hierarchical task analysis in order to try to explore the design space around each task thoroughly. Only the bigger tasks were included, namely *find cluster*, *compare to other analysis techniques*, *identify infection transmission path*, *find similar samples*, *compare to saved group of samples* and *compile report*. The brainstorming resulted in one mind-map for each task, shown in figure 6.5.

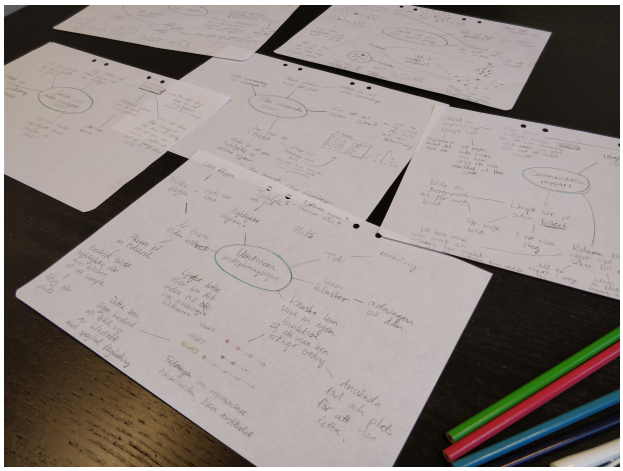
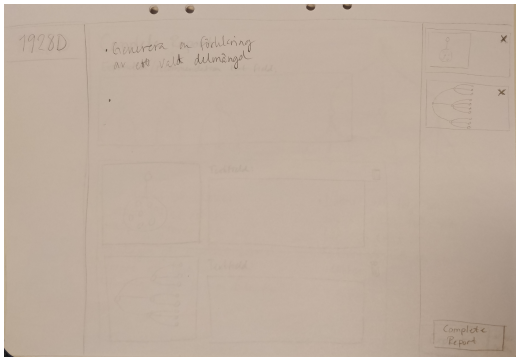


Figure 6.5: Mind maps

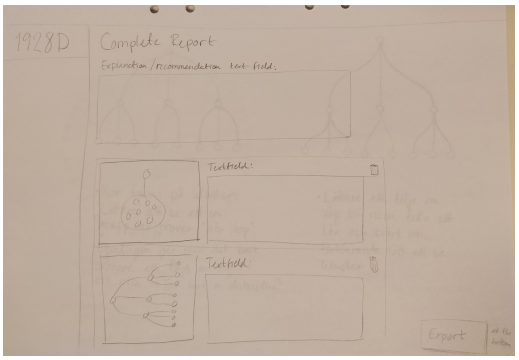
After the brainstorming the ideas were analysed in terms of which would best satisfy the persona's need and wants. Due to the nature of the mind maps it was possible to combine certain ideas at this stage to better suit the tasks that the persona aims to complete. After the analysis the best ideas were selected to be further developed. In order to develop the ideas they were sketched out with pen and paper since it helps to stimulate creative thinking. When sketching the different ideas a first draft of a cohesive interface developed as well as different suggestions for some parts, especially for how to visualise the relationships between the samples.

Although the personas provide great insights in what the users would look for in the product it was sometimes difficult to determine if an idea will add value for the users or not, especially different kinds of diagrams. When sketching, many different kinds of diagrams were produced and it was uncertain what kind of value some diagrams would add to the reporting process and also if they would add any value at all. Therefore, these issues were discussed with the CTO from 1928 diagnostics who has had much contact with users in order to ensure not spending time on useless ideas. During this discussion some ideas were dismissed while others were built upon further.

The idea developed for the reporting work flow is that, while analysing the data, the user can add snapshots of the view they are currently in to a staging area for the report. When the user is finished with the analysis and is satisfied with the graph selection they will go to a page for finalising the report, which means they can add explanatory texts to the graphs but also the report as a whole. From this page the report is exported so that it can be sent to the infection control unit. The staging area and the finalising page are shown in figure 6.6.



(a) Staging area for report.



(b) Screen for completing report,

Figure 6.6: The sketches for the reporting work flow.

A few ideas about redesigning the edges in the UPGMA tree were generated. The idea was to use the gestalt laws to make it easier for people, especially the infection control unit, to see relationships between samples. Two suggestions are shown in figure 6.7. There were also many ideas for improving the main tree view to fit the reporting work flow. Since the graphs that are added to the report will be read by people with less area knowledge than the people constructing the report the idea is that all the information displayed in the tree view should be flexible. This means that the user has relatively much control over what information to include in the tree and which information to highlight. Ideas to accommodate this is to colour code different attributes that can be perceived as groups, put highlights on samples with specific properties and provide a way to select what information to be written out next to every sample.

For finding clusters, and thus outbreaks, easily in the interface the ideas focused on how to provide a good overview of all the samples. One idea to solve this is

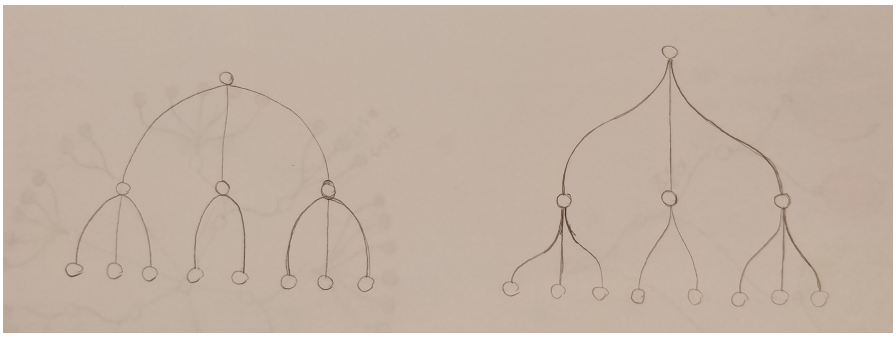
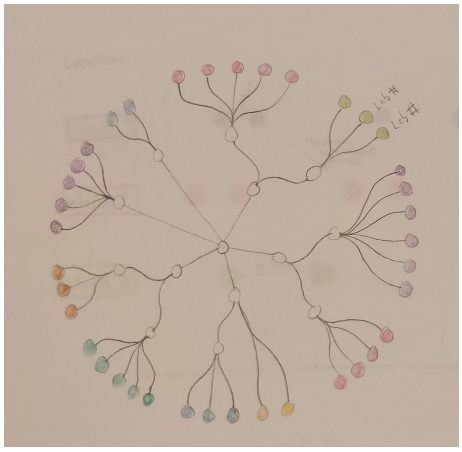
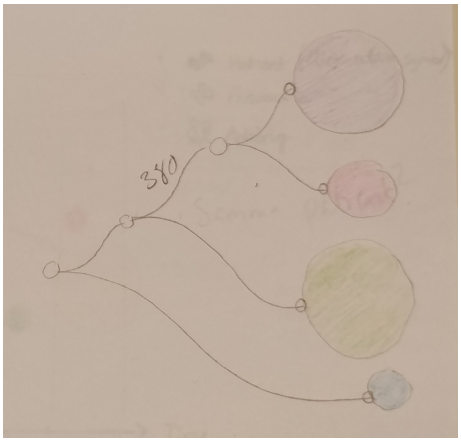


Figure 6.7: Sketches on different edge designs in a UPGMA tree.

to display the phylogenetic tree in a circle as shown in figure 6.8a. Even though this is very space efficient it is not possible to add much information to it while keeping it easy to read, therefore the tree will change to a regular tree layout when zooming into a smaller sub tree. An idea is also to show clusters as visual aggregates instead of individual samples in order to further emphasise where the clusters can be found, shown in figure 6.8b. The clusters will also be colour coded in every state of the tree.



(a) A circular UPGMA tree.



(b) A UPGMA tree with visual aggregates.

Figure 6.8: The sketches of diagrams for providing a good overview.

Another idea is to show the minimum spanning tree in order to provide an overview of all the samples. The visual aggregates of several samples can be implemented in this type of tree as well and when zooming in on a cluster the distances between the samples within it is shown as a smaller minimum spanning tree. A sketch of how a minimum spanning tree could look is shown in figure 6.9.

An idea for facilitate the identification of infection routes, i.e. try to understand where an outbreak started and in which order the patients were infected, was to incorporate time into the visualization. Since there can be many attributes visually encoded in the tree the idea is to have a time axis below the tree which shows

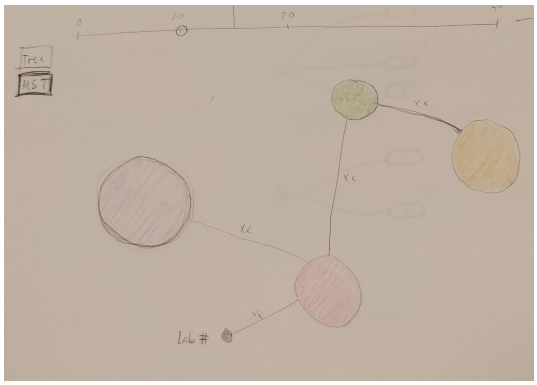


Figure 6.9: A sketch of a MST.

when the samples were taken. From the timeline the user can choose to show samples from a specific time period by selecting a time span with sliders. An idea to further make the identification of infection routes easier was to show the samples within a cluster in a diagram with time on the x-axis and place on the y-axis, shown in figure 6.10. The idea is that this will help the infection control unit to get a visual picture of the outbreak and improve their advice.

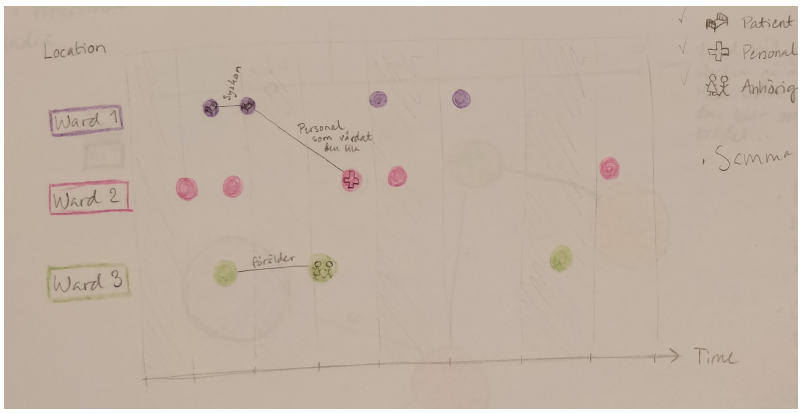


Figure 6.10: A sketch of a diagram that shows the samples in a diagram with time and location on the axes.

The infection control unit is often interested in a small set of samples and an idea to make it easier for the microbiologist to show the relationships between that smaller set was to let them select a number of samples directly in the tree to perform certain actions with. There would be a selection mode of the tree where checkboxes appear next to every node and a dropdown with different actions the user can perform on them, for example show in a minimum spanning tree or the infection route finder, this is showm in figure 6.11.

An idea to make it easy for the user to compare a sample to a group of other samples was to make a split view with a list of all the samples on one side and a small visualization on the other side with limited interaction. The sample that the user has chosen to compare with others is locked at the top of the list and is highlighted in the tree visualization.

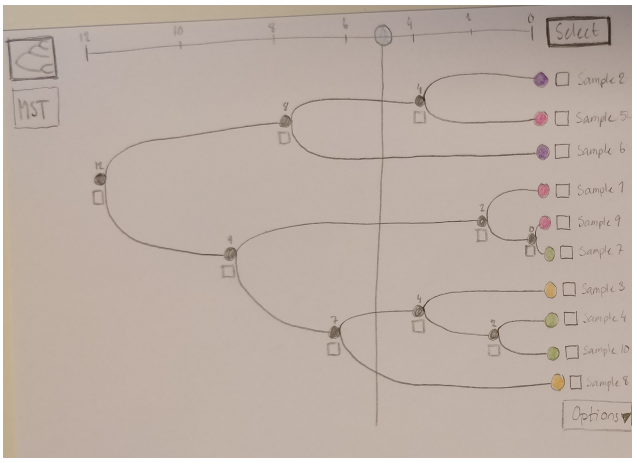


Figure 6.11: A sketch of the select feature.

6.3.2 Paper Prototype

The majority of the sketched and developed ideas were compiled into a paper prototype, shown in figure 6.12. The purpose of creating a paper prototype was mainly to explore how the ideas could be put together with intuitive navigation and interaction, but also to test the concept with user tests. Since the designs of the of the phylogenetic tree and the minimum spanning tree are difficult to draw at a good enough level for proper testing, only one design of each type for smaller trees were included in the paper prototype. Those trees were not based on real data but were included in order to illustrate the features that manipulates the trees in different ways, for example colour coding, highlighting and zooming. Different designs of the trees were prototyped and tested in the next iteration with a higher fidelity prototype in order to get visualizations that are as close to reality as possible.

The idea with the split screen for comparison of samples was not included in the paper prototype either since it was ruled out as redundant. It was found during the construction of the paper prototype that there are other features that provide a similar result, for example selecting a subset of samples in the tree and work with highlights and annotations.

6.3.3 Updated task analysis

After the ideation and concept development a few things changed from the original task analysis and many tasks could be broken down into another level of sub tasks. As a consequence of the idea of the report creation workflow explained in section 6.3.1 the task *Compile report* has become the only top level task, since everything that is done within the visualisation can be seen as a step towards completing the final report. The first four layers of the task analysis, including all sub tasks of *Compile report*, are shown in figure 6.13.

In the first version of the task analysis there were two top level tasks, namely *Analyse NGS data* and *Compile report*. In the updated version the task *Analyse*

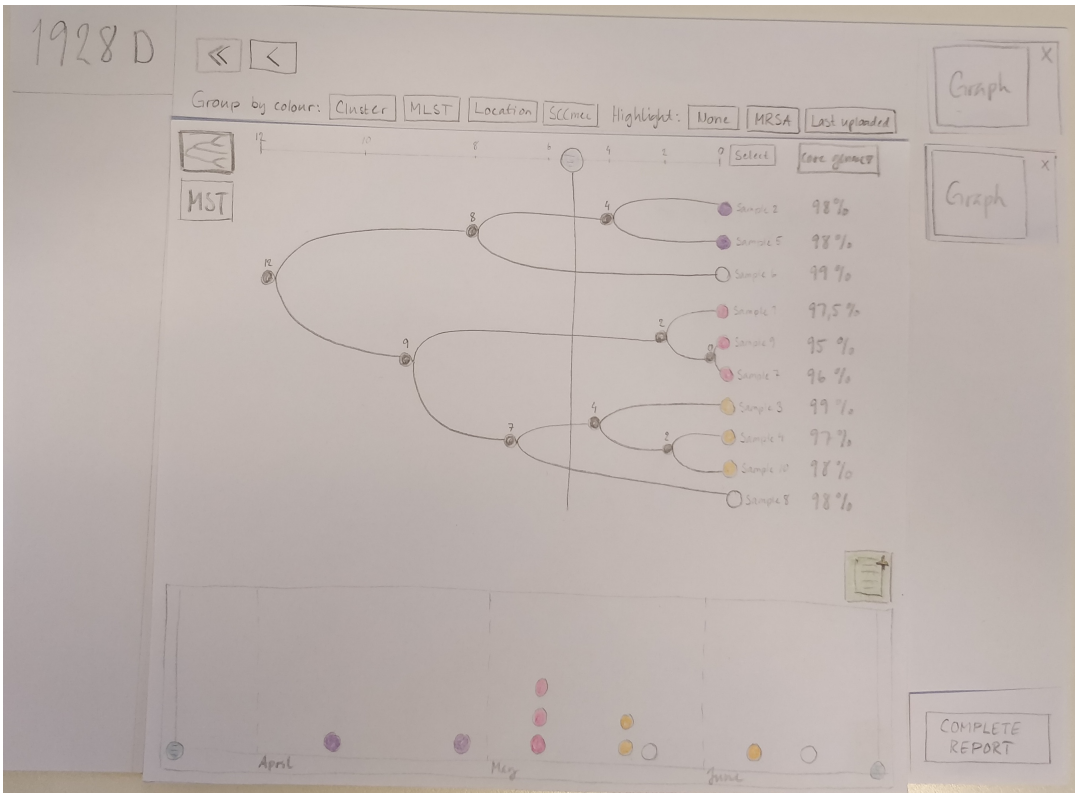


Figure 6.12: Paper prototype

NGS data has become a sub task of *Create different diagrams* since the diagrams are a result of the analysis. The tasks on the level below *Analyse NGS data* are the same tasks as before.

The task *Check quality parameter* was a sub task of both *Find outbreak* and *Find inclusion or exclusion* in the original analysis. In the updated version this task has been moved down as a sub task of some of the other tasks since it felt more intuitive in the interface to incorporate such a small task in the other workflows. The same thing was found to be suitable for the task *Check resistance markers*. The updated subtasks of *Find outbreak* and *Find inclusion or exclusion* are shown in figure 6.14, and the last level of tasks can be found in appendix ??

6.3.4 User test

Three user tests were conducted with employees at 1928 Diagnostics as participants. All participants had the required knowledge to understand the data and the associated information, but they had varied understanding for the final users of the product. One person has had moderate contact with users and had an understanding of their basic needs and wants, one person has been part of discussions about the users and has a vague image of them and the third person has very little understanding of the users. This was not seen as problematic at this stage since the focus of the test was on the navigation and if it was easy to understand.

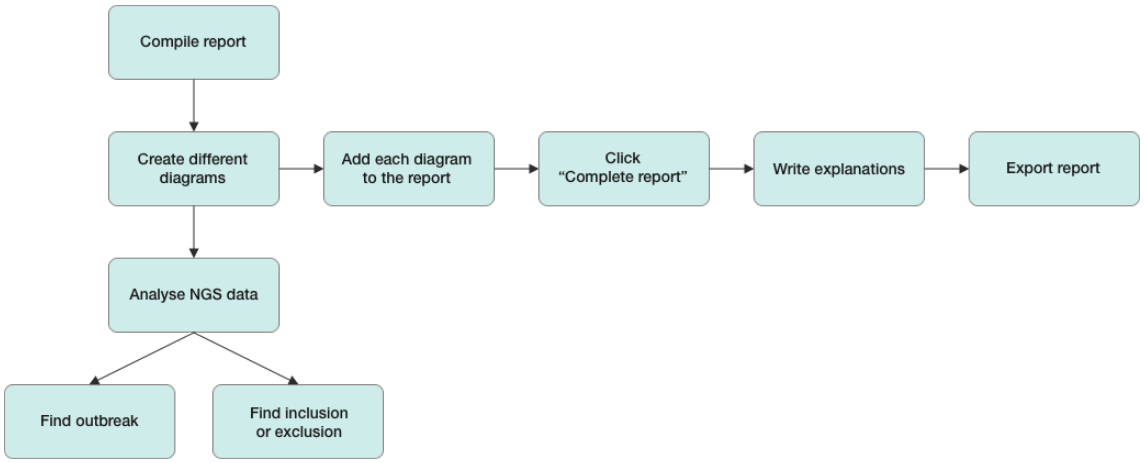


Figure 6.13: Task analysis

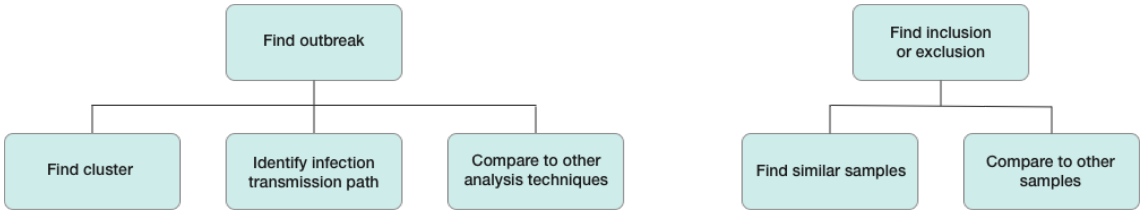


Figure 6.14: Task analysis

Every test lasted for approximately 30 minutes and the participants were asked to do simple tasks in the interface and think aloud while trying to complete them, shown in figure 6.15. For some tasks questions were asked before showing the result of the participants action in order to gain an understanding of what the participants expected the interface to do. The tasks and questions for the user test is listed in appendix A. Interesting actions and ways of completing the tasks were written down during the tests so that the results could be analysed easier afterwards. Since I conducted the tests alone I placed new elements in front of the test participants and took notes simultaneously, the set up of the test is shown in figure 6.16.

6.3.5 Focus Group

A focus group was held in order to get feedback on the design in the paper prototype from people that have a good insight into the users work. The focus group consisted of the CTO and the COO at 1928 Diagnostics, one person from sales was also supposed to participate but was not able to attend that day. The meeting lasted for an hour and started with a demo of the paper prototype followed by a discussion around the features in the interface.

6.3.6 Results

The results from the first iteration of the design process are presented in this section, which include results from the user tests and the focus group.

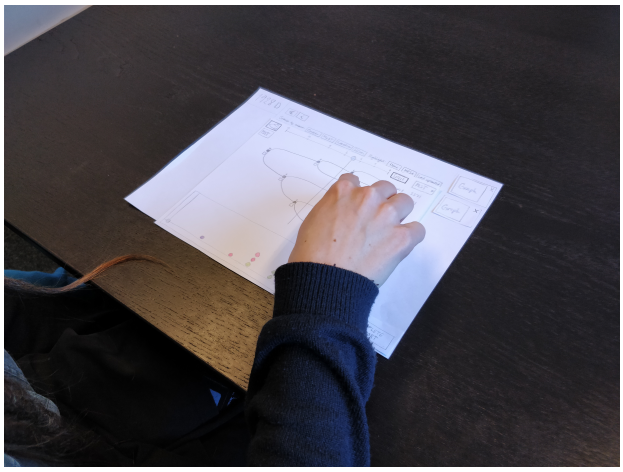


Figure 6.15: User test participant

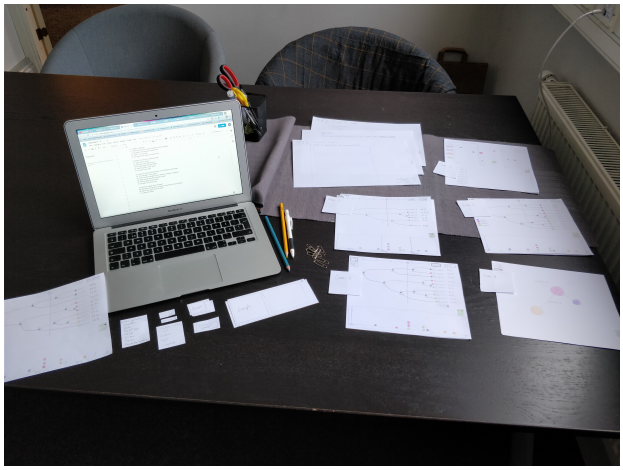


Figure 6.16: Set up for user test one.

6.3.6.1 User Tests

The user tests highlighted a couple of problems in the interface. For some of the problems the participants had it was clear that it was due to the quality of the prototype from what they said when trying to solve that specific task. These things included unclear drop down buttons and sliders. Apart from problems induced by the quality of the prototype there were a few features that the participants had difficulties of understanding.

The main issue identified in the user tests was that there was a great confusion around the select feature and the highlight feature. When asked to select a small group of samples and show them in another diagram two of the participants tried to highlight the samples first and expected to be able to do something with the highlighted samples. When they did not find how to create a new diagram with them they became slightly frustrated and one participant said he felt stupid that he did not understand what to do. One participant tried to highlight the samples and then click the select button, which did not produce the anticipated result. This issue is the most crucial to solve since it is a core feature for using many other

features in the interface, such as the infection route finder and showing a small selection in a new diagram.

It was clear from the user tests that it will be important to work with pliancy and affordances in the final interface. Since there are many things on the screen that the user can interact with it need to be clear which those are, and also that they communicate what will happen when the user interacts with them. All participants had a hard time understanding the difference between the toggle buttons and the "go back" buttons. When asked to go back to the original tree they pressed the toggle button for the tree view instead of the arrow buttons at the top. Another example of missing pliancy was that two of the participants did not try to interact with the internal nodes in the tree when asked to zoom into a sub tree. Instead they used the select feature and selected all samples one by one in a cluster and then entered the drop down to choose to show them in a new tree. This flow is not wrong but the interaction with the internal nodes in the tree was included in order to spare the user all these clicks for a common task.

None of the participants had any trouble with using the reporting work flow. When the report started building up at the side of the visualisation all participants understood how it worked. Two of them also started trying to delete and add diagrams to the report throughout the test which suggested that the interaction with the reporting feature was intuitive.

6.3.6.2 Focus Group

The discussions during the focus group meeting mainly revolved around how to incorporate the time dimension visually in the horizontal tree view and what information that should be visually encoded in the different tree views. In addition to that the compare functionality was revisited as well as the navigation between the different views and zooming.

The group thought that the time dimension should be visually encoded in the tree in addition to the time line below it. The idea was that this would provide the user with a context for the samples that would help them to trace transmission routes. An idea that would take the visualisation in that direction was to be able to highlight samples within a time interval. Either with a toggle button among the other highlight attributes or from the time line. This started a discussion around what the time line should be used for. The conclusion was that there are two possible use cases for the time line. Either the user wishes to create a new tree with the samples in the chosen interval or highlight the chosen samples in the tree.

The most important information that was missing in the visual encoding in the tree was the uncertainty of which samples actually belong to an outbreak or not. Within a certain distance interval it is very difficult to say if two samples are related or not. Incorporating that in the visualisation would greatly help the microbiologist to explain the result to the infection control unit. Other attributes that could be considered to encode are to show if two samples are from the same patient

and choose to group and colour code based on several sequencing types in the minimum spanning tree.

Although the compare function was discarded at the beginning of this iteration the focus group now missed this feature. However, this time it was discussed whether the samples could be compared to a pre-selected reference sample instead of a saved group of samples. When the original sketch of the compare screen was presented the group thought it was a good starting point for this feature. The fact that there has been very different opinions on whether this feature should be included or not suggests that they do not know if the user would use the feature. Therefore it should be discussed with a user if it would add value to their work.

When presenting how the zooming and navigation in the trees worked in the prototype some concerns were raised. The main worry was that the user would lose context of where in the tree they were currently working, and that the clicking of the nodes would make the experience choppy. An idea of using a toolbox was presented as something that could be explored in order to solve this issue.

6.3.7 Revised Guidelines

Based on the results from the first concept iteration the guidelines were revised and one guideline was added.

1. Focus on tasks Spend time to understand what tasks, within infection control, the users need to perform in the interface and design around them. Since the product is a tool for microbiologists in their daily work it is important that they can perform all necessary tasks without distraction [29]. This approach was also successfully tried in the clinical report design project explained in section **??**.

2. Adapt to the Reader Have the recipient of the report in mind when designing the appearance of the diagrams in the visualisation. From the secondary persona in section 6.2.2 it is apparent that the visual language needs to help the infection control unit to interpret the results. This could be achieved by using different gestalt laws and colour to highlight and emphasise different characteristics [18]. An efficient visual language could also help the microbiologist write explanations for the diagrams.

3. Flexibility Since there are different infection control organisation types, explained in section 6.2.2, the tool should be flexible enough to suit different ones. From the report review in section 6.1.2 it can be seen that the reports are made differently in order to to suit the hospital's different workflows.

4. Emphasise interaction possibilities Work with affordance and pliancy in the interface in order to help the user understand how to interact with the visualisation. By using different types of hinting in the interface the user can find the most efficient way of performing certain tasks [25]. This is a common interaction design principle that should be followed [24].

5. Good Defaults Let the user enter the visualisation with good default values entered. Since, according to the primary persona in section 6.2.2, time is a limited resource having good defaults value might enable the user to make the analysis faster.

6. Merge Workflows Incorporate the report creation workflow in the analysis workflow in order to reduce the amount of extra work the microbiologist has to do. The user test results, in section 6.3.6.1 suggested that this concept could be a good example of this.

6.4 Second iteration

In the second iteration of the concept development phase the focus was on the appearance of the circular and horizontal trees as well as the minimum spanning tree diagrams. Different designs for a few central properties were prototyped and evaluated in a user questionnaire.

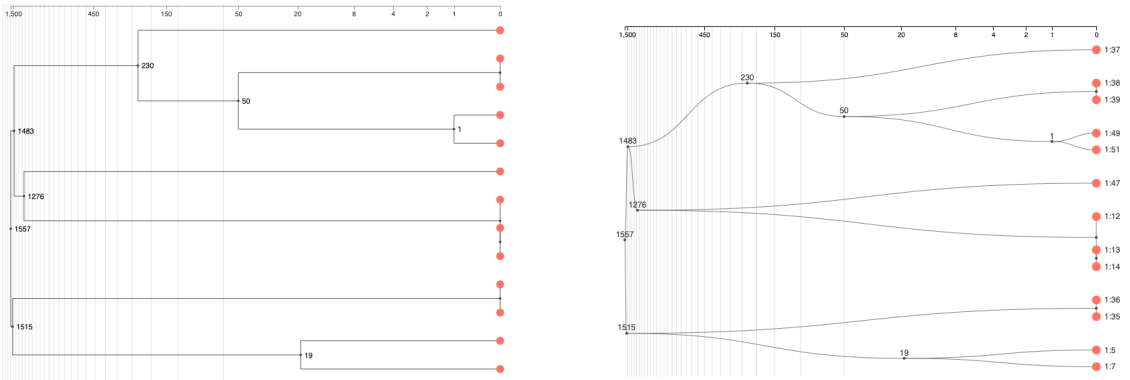
6.4.1 Ideation and Prototyping

The ideation and prototyping were divided into two parts, one for the phylogenetic trees and one for the MST. The process of each part is explained in this section.

6.4.1.1 Phylogenetic Trees

The ideation session in the first iteration resulted in a number of ideas for the general design of the diagrams. For the phylogenetic tree the ideas were concentrated around whether the edges should be curved or be in a right angle. In this iteration the focus was on how to visualise other properties of the phylogenetic tree namely, how to visualise clusters, colour code certain properties while still see clusters and highlight samples.

The phylogenetic trees were prototyped using the JavaScript library D3 and real data from the 1928 platform in order to create realistic diagrams that could be tested with users. While exploring the D3 library different ideas on how to visualise the different properties were produced. Thus, in this iteration, the prototyping served as an ideation tool. The ability to visualise the data realistically also allowed for exploring how many samples were suited for the horizontal phylogenetic tree, in order to minimise scrolling, and the circular phylogenetic tree and therefore also at what amount the minimum spanning tree would be the best option. It was found that the horizontal tree would be suitable for approximately 25 samples, the circular up to approximately 100 and above that the MST would be the most suitable.

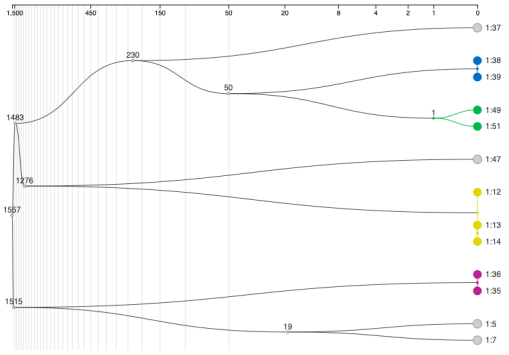


(a) Right angled edges **(b)** Curved edges

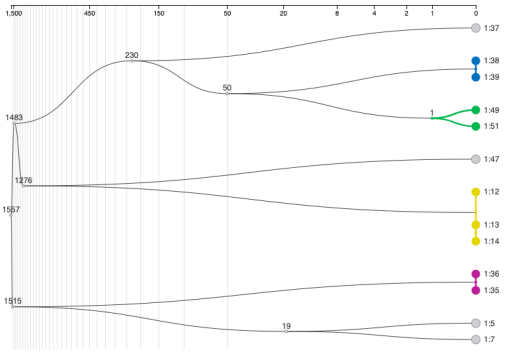
Figure 6.17: Two different designs on the edges in the phylogenetic tree.

Two edge designs were prototyped, curved and right angled. The right angled design is most common and is the design mostly used for this purpose. The idea behind the curved edges was to apply the gestalt laws of continuity and proximity, explained in section 3.3.3, in order to make it easier for the user to visually find clusters and see how closely related groups of samples are. The two designs are shown in figure 6.17.

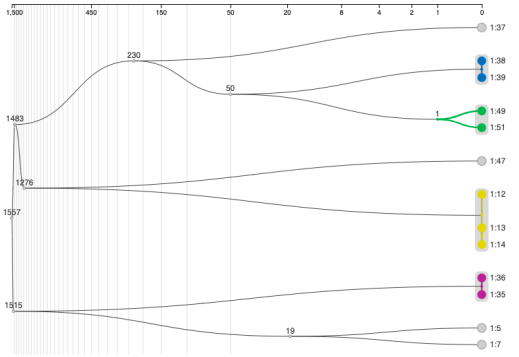
In order to make it easy for the user to identify outbreaks each cluster that represents an outbreak were assigned a colour, thus applying the gestalt law of similarity explained in section 3.3.3. Two prototypes were made where the nodes as well as the edges were coloured and one where only the nodes were coloured. In one of the prototypes where the edges were coloured the coloured lines were thicker than the rest of the lines, figure 6.18b, figure 6.18a shows the same design with thin coloured lines. In the last one a grey shadow was added behind the clustered nodes in order to emphasise it further, figure 6.18c.



(a) Coloured clusters with thin coloured lines.



(b) Coloured clusters with thick coloured lines.

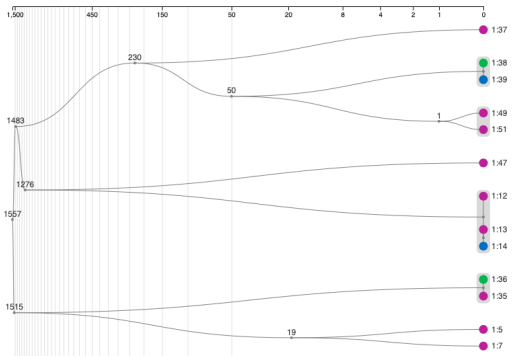


(c) Coloured clusters with grey shadow.

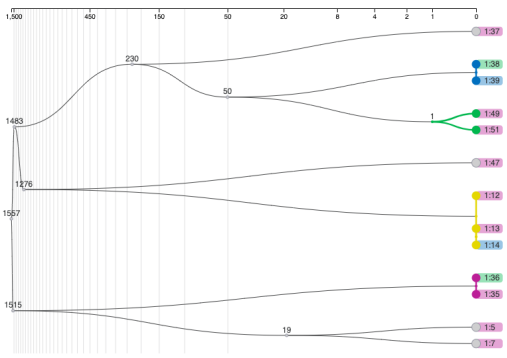
Figure 6.18: Three designs for colouring clusters representing outbreaks.

Since it is important for the user to be able to clearly see which samples are part of an outbreak at all times it was important to figure out how to colour code the samples depending on metadata, for example location, while not losing the outbreak

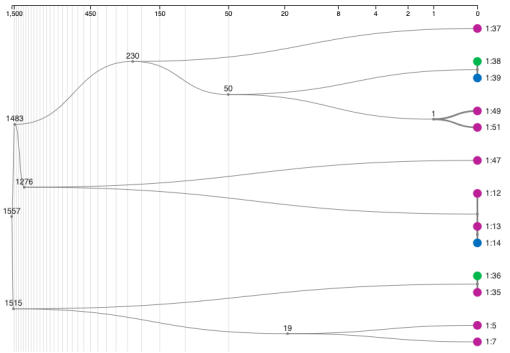
information. Three designs that could solve this problem were prototyped. In the first design, shown in figure 6.19a, the leaf nodes are colour coded depending on location and the cluster information is shown with a grey shadow behind the clustered samples. In the second design, shown in figure 6.19b, the samples and edges in a cluster are coloured and the location is shown by adding colour coded backgrounds to the sample IDs. In the last design, shown in figure 6.19c, the nodes are colour coded based on metadata and the outbreak clusters are shown through thicker edges between the nodes that are included in an outbreak.



(a) Colour coded sample nodes with grey shadow.



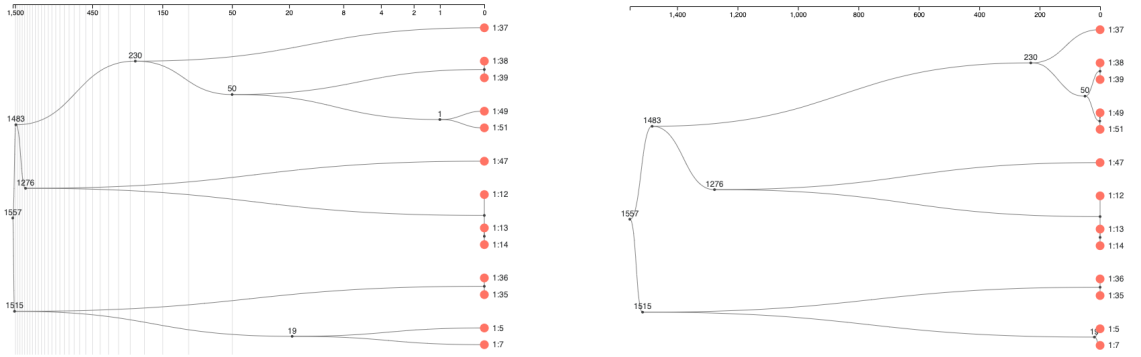
(b) Coloured outbreaks with colour coded background on lab ID.



(c) Colour coded sample nodes and thick lines representing outbreaks.

Figure 6.19: Three designs for colour coding metadata while preserving outbreak information.

Working with D3 also helped to identify that different types of scales can be used for directing the users attention to the important parts of the diagram, i.e. the outbreaks. The idea was to use a logarithmic scale when the difference between the largest and smallest distance is quite big. When the difference is smaller a linear scale can be used in order to show a more fair picture of the relationships. The two different scales are shown in figure 6.20.



(a) Logarithmic scale.

(b) Linear scale,

Figure 6.20: Two different scales used for placing the inner nodes in the phylogenetic tree.

6.4.1.2 Minimum Spanning Tree

The ideation and prototyping process was slightly different for the minimum spanning tree. When developing ideas for how to color code and highlight samples the main technique used was sketching. Since it was found in the first iteration that this type of diagram could be suitable for visualise a large amount of samples it was also considered how visual aggregates could be used in order to reduce the amount of visual entities on the screen. As a consequence it was also ideated upon how to combine the visual aggregates with the colour coding and highlighting.

For colour coding attributes when there are visual aggregates in the diagram, the idea chosen to take further was to show the distribution of the attribute as a pie chart on the visual aggregate, shown in figure 6.21. For highlighting, the visual aggregates that contains the highlighted samples will be highlighted so that the user can see where to zoom in in order to see their highlighted samples individually.

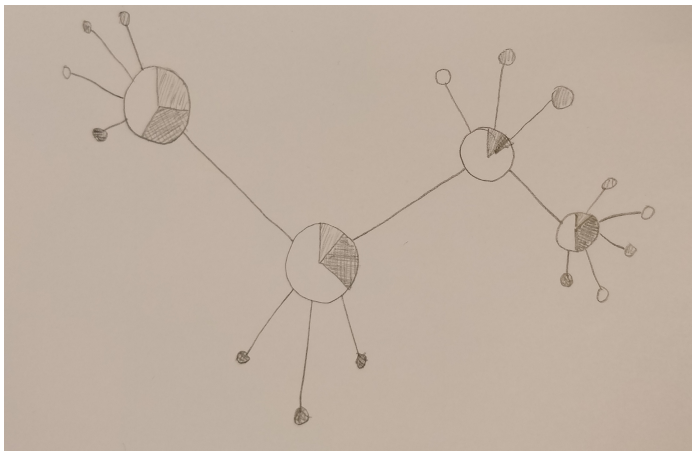


Figure 6.21: Pie charts on visual aggregates for colour coding attributes.

6.4.2 User Questionnaire

When different designs of the tree had been prototyped a questionnaire was created in order to get feedback from potential users. It was sent out to 9 current or potential users of the product and 18 employees at 1928 Diagnostics. This group was chosen so that there would be approximately the same amount of participants with and without experience in bioinformatics or microbiology. Since the participants come from different countries they could choose if they wanted to answer it in English or in Swedish depending on what they felt more comfortable in. The questionnaire was open for 7 days.

The questionnaire focused on different properties in the horizontal tree in order to keep it short. This way it was expected that the probability that people outside 1928 Diagnostics would take time to answer it would increase. The questionnaire was divided into 6 parts, starting with two questions about was profession and education the participant has. Each following part consisted of 1-3 images of the design suggestions created in the prototyping phase for a particular property of the tree. Following the image collection was a question of which design the participant preferred and the possibility to motivate the answer. The properties tested in the questionnaire were; design of the edges, type of scale, cluster identification, colour coding of metadata and level of information content in a circular tree, and the images used were the prototypes of phylogenetic trees shown in section 6.4.1.1.

6.4.3 Result

20 persons participated in the questionnaire and they could be divided into two groups; *Group 1* includes people that are knowledgeable in bioinformatics and used to phylogenetic trees, and *Group 2* includes people that have limited knowledge in bioinformatics and that are not used to phylogenetic trees. The groups were divided in this way in order to both represent the microbiologists that would use the platform and the infection control unit that would receive the report. The complete results are presented in Appendix B.

The total result of what type of edges the participants preferred when trying to understand the relationships between the samples is shown in figure 6.22. A majority preferred the right angled edges over the curved ones and the result was similar for the two different groups although slightly fewer of *Group 2* chose the right angled edges. The motivations for the right angled edges included that participants were used to that design and thus could interpret it faster and also that it was easier to compare to the scale above.

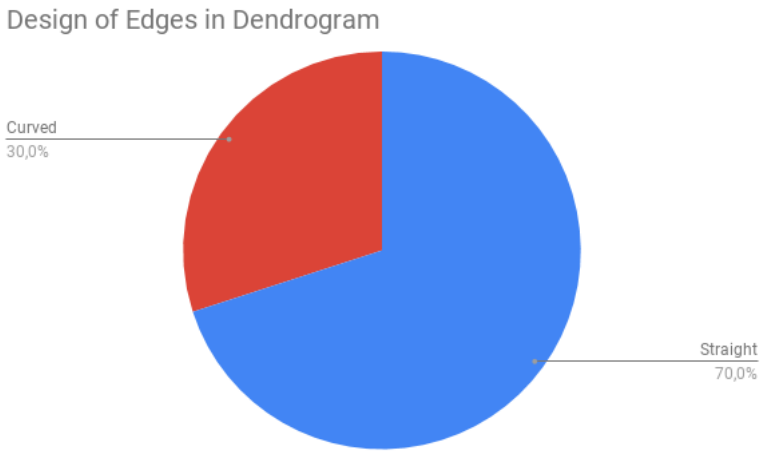


Figure 6.22: Total result for question 1.

Figure 6.23 shows the total result of which type of scale the participants preferred when the difference between the largest and smallest distances is large. A majority preferred the logarithmic scale because it provides a more detailed image of the outbreak information. The results were very similar for both *Group 1* and *Group 2*. The main argument against the logarithmic scale was that it could be confusing and easy to miss.

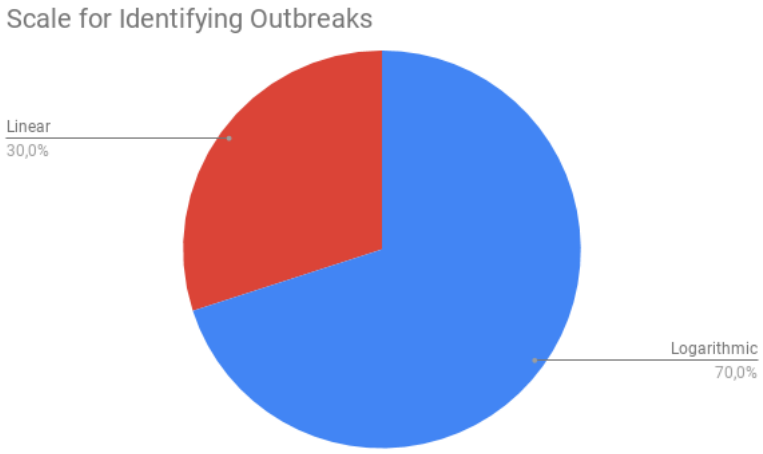


Figure 6.23: Total result for question 2.

When asked which design made it the easiest to identify clusters the two groups answered almost identical, the total result is shown in figure 6.24. Most participants preferred the design in image 3, where the sample nodes and the edges were coloured when part of an outbreak with an additional grey shadow behind the nodes in the cluster.

Figure 6.25 shows the total result of the question regarding colour coding of metadata. In image 1 the sample nodes were coloured according to metadata with

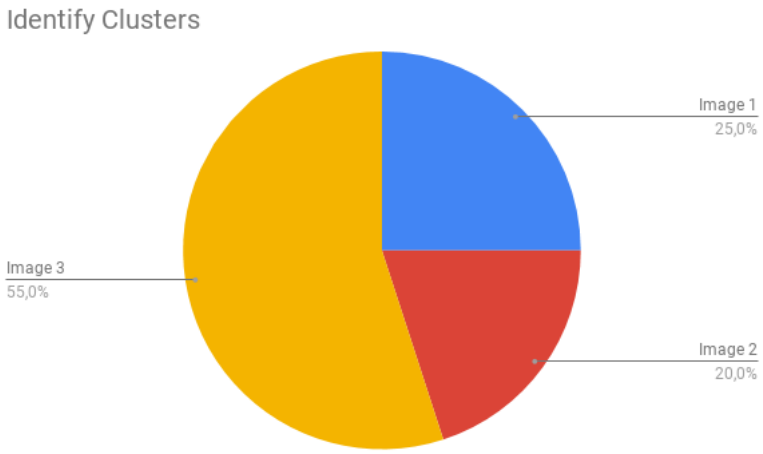


Figure 6.24: Total result for question 3.

a grey shadow showing the outbreak information and in image 2 the metadata was shown by colour coding the background of the lab ID while also keeping the colours for the outbreaks. The motivations for image 1 included that the grey shading made it easier to also take in the metadata information and that it was less visually busy than image 2.

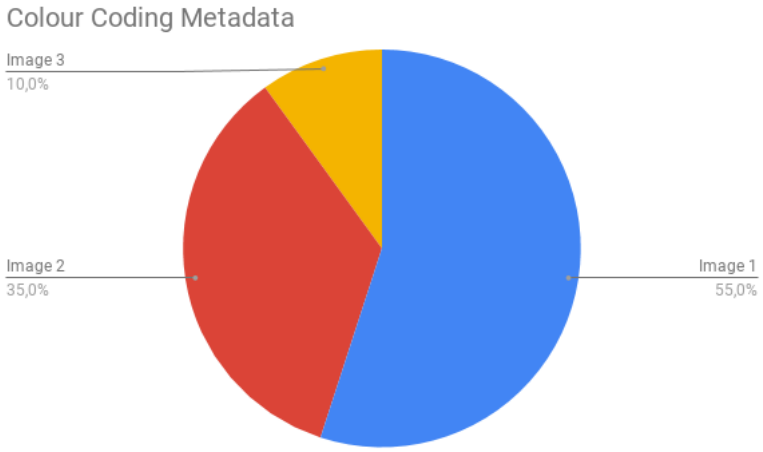


Figure 6.25: Total result for question 4.

The last question in the questionnaire focused on the amount of information shown in a circular phylogenetic tree. When showing the distance at every internal node the diagram was very cluttered so the image in the questionnaire only showed the distances at the root node as well as the nodes within an outbreak. For this questions *Group 1*'s answers are the most important since they have more knowledge of the domain. A large majority of *Group 1* thought that the amount of information was enough since it would be possible to zoom in to branches and get more detailed information.

6.4.4 Revised Guidelines

In this iteration two more guidelines were added based on the results from the user questionnaire.

1. Focus on tasks Spend time to understand what tasks, within infection control, the users need to perform in the interface and design around them. Since the product is a tool for microbiologists in their daily work it is important that they can perform all necessary tasks without distraction [29]. This approach was also successfully tried in the clinical report design project explained in section **??**.

2. Adapt to the Reader Have the recipient of the report in mind when designing the appearance of the diagrams in the visualisation. From the secondary persona in section 6.2.2 it is apparent that the visual language needs to help the infection control unit to interpret the results. This could be achieved by using different gestalt laws and colour to highlight and emphasise different characteristics [18]. An efficient visual language could also help the microbiologist write explanations for the diagrams.

3. Flexibility Since there are different infection control organisation types, explained in section 6.2.2, the tool should be flexible enough to suit different ones. From the report review in section 6.1.2 it can be seen that the reports are made differently in order to to suit the hospital's different workflows.

4. Emphasise interaction possibilities Work with affordance and pliancy in the interface in order to help the user understand how to interact with the visualisation. By using different types of hinting in the interface the user can find the most efficient way of performing certain tasks [25]. This is a common interaction design principle that should be followed [24].

5. Good Defaults Let the user enter the visualisation with good default values entered. Since, according to the primary persona in section 6.2.2, time is a limited resource having good defaults value might enable the user to make the analysis faster.

6. Merge Workflows Incorporate the report creation workflow in the analysis workflow in order to reduce the amount of extra work the microbiologist has to do. The paper prototype user test results, in section 6.3.6.1 suggested that this concept could be a good example of this.

7. Familiarity Diagrams that are commonly used in the field should keep familiar features. The results from the user questionnaire, explained in section 6.4.3, suggests that users appreciate information being visually encoded in traditional designs of diagrams. By keeping the most prominent features the user might be able to interpret the diagrams quicker.

8. Direct attention to the answer Use visual encoding in a way that shows the answer to the most common questions at firs glance. The user questionnaire

results, in section 6.4.3, shows that the user preferred redundant encoding for outbreak information since they immediately can see which samples are part of an outbreak.

6.5 Third iteration

The goal of the third iteration was to combine the results from the previous iterations into a high fidelity prototype of the entire concept. The prototype was then evaluated by a focus group, an expert interview, a questionnaire and finally a user test. Based on the results the final guidelines were formulated.

6.5.1 Ideation and Prototyping

The high fidelity prototype was made with the JavaScript library D3, Sketch and InVision. D3 was used for drawing the UPGMA trees, these images were then print screened and added to the views made in Sketch. All views as well as the prototypes of the MST were created in Sketch. Most of the interactivity was then implemented with the help of the InVision plug in and then synced to their website. A snapshot of the interface is shown in figure 6.26. Figure 6.26 shows a snapshot of the interface. The interface is divided into four parts, a toolbar at the top, the main visualisation area in the middle, a timeline with the samples at the bottom, and the column to the right which will show the progress of the report.

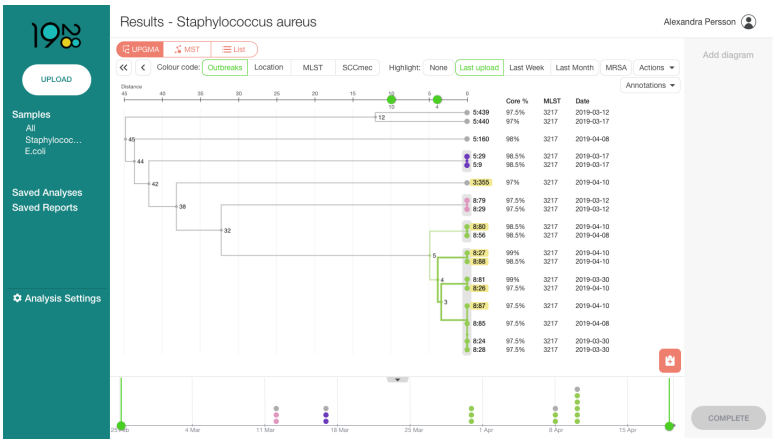


Figure 6.26: Prototype.

Three topics were ideated upon during this iteration. One of the most important as how to improve the time line functionality. Some ideas involved trying to incorporate the time aspect in the UPGMA tree and MST. These ideas were discarded in order to minimise the risk of encoding too much information in the same diagram. Instead, the idea that was chosen to develop further in to the prototype was that the user would be presented with options for what to do with the samples within the selected time frame on the timeline. When the sliders have been dragged to their desired positions a drop down button appears when hovering

over the timeline, as shown in figure 6.27. Here the user can choose to cluster the selected samples in a new tree, highlight them or dim the excluded samples, which is the default.

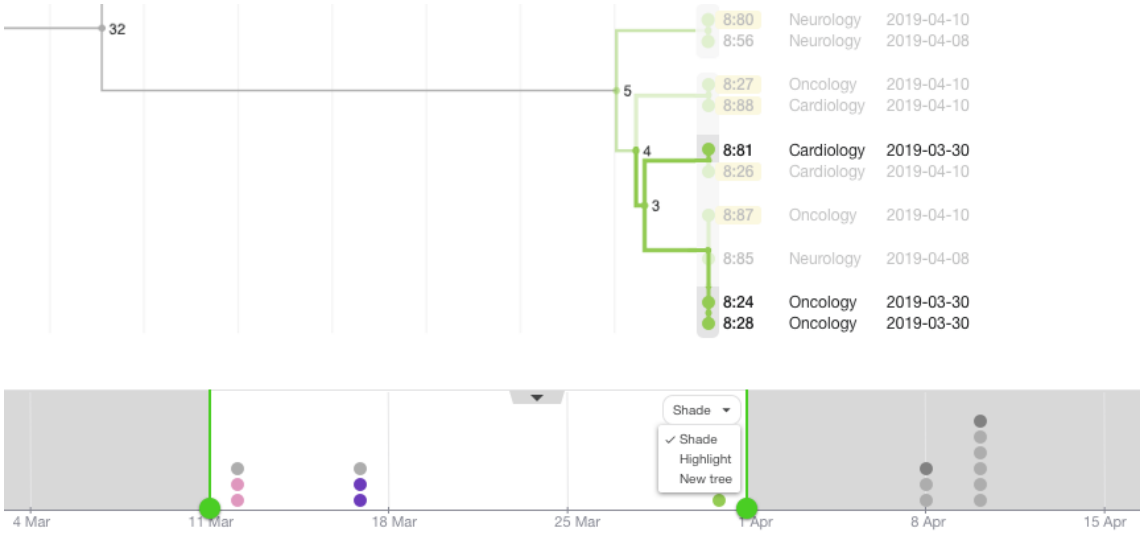
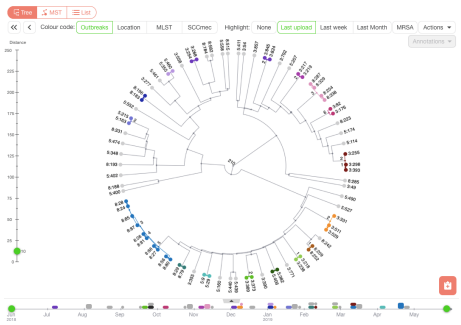
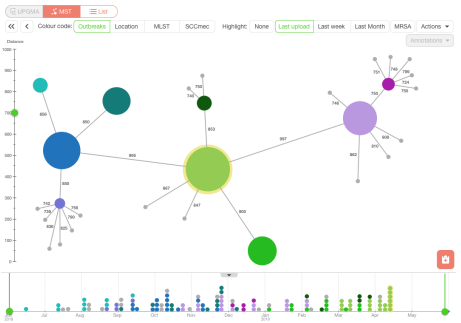


Figure 6.27: Time line.

Depending on how many samples the user has chosen to work with they enter the visualisation at different diagrams. Up to approximately 25 samples the user first sees the horizontal UPGMA tree shown in figure 6.26. This diagram is also used if all or a majority of the chosen samples belong to the same outbreak since the annotated information might be of interest in that case, even though this means that the user has to scroll to see everything. If the user has chosen between approximately 25 and 100 samples they see the circular UPGMA tree first, shown in figure 6.28a, and above that a MST, shown in figure 6.28b. At any time, if the amount of samples approve, the user can toggle between displaying the samples in a UPGMA tree, MST or a list with more detailed information, such as resistance markers. The list view is shown in figure 6.29.



(a) Circular UPGMA.



(b) MST.

Figure 6.28: The different diagrams for displaying the samples.

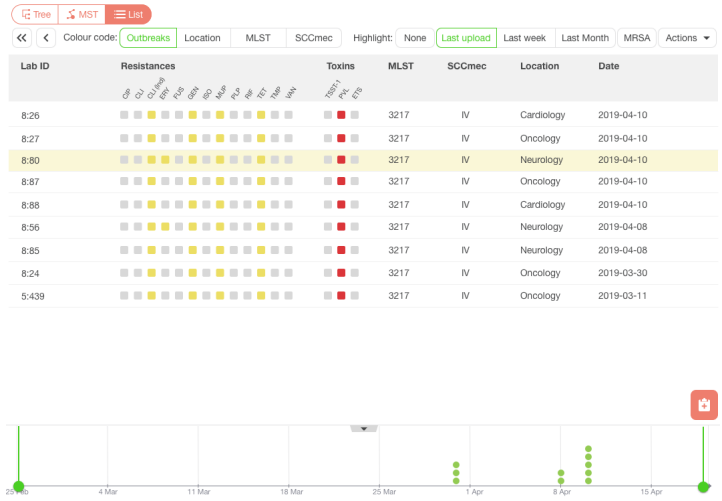
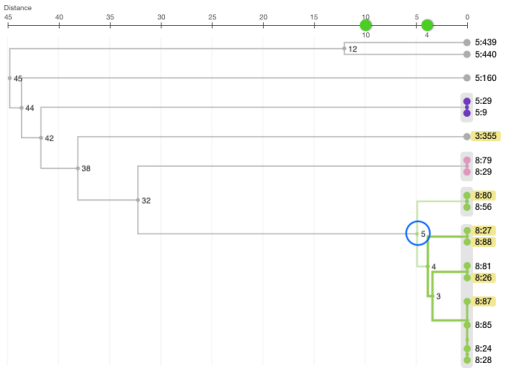
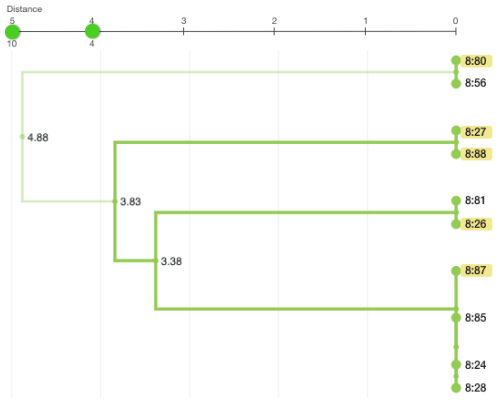


Figure 6.29: List view of samples.

An important feature is to be able to zoom into certain parts of the different diagrams. In the UPGMA the user can zoom into a branch by clicking an inner node of the tree, figure 6.30a. A UPGMA tree of only samples in the chosen branch is then displayed, figure 6.30b. In a MST the zooming is similar. Since every large circle represents a group of samples the user simply clicks a large circle in order to show only the samples inside, figure 6.31. In order to control how many samples you want to include in the zoom the user can use the distance threshold on the left hand side of the diagram. The threshold in this view determines which samples should be grouped together, so every sample that is within a shorter distance than the threshold are included in the group.



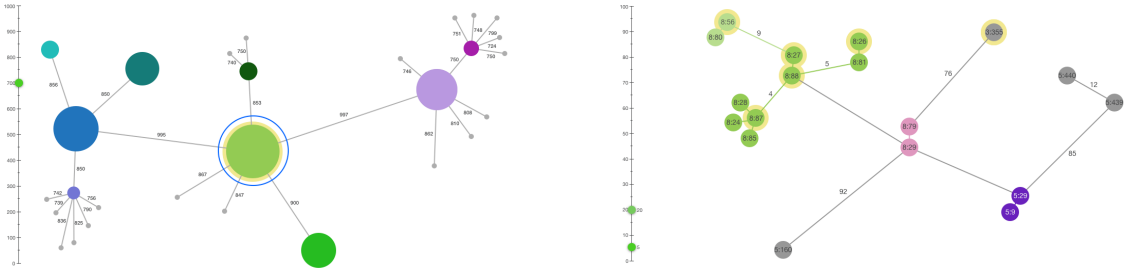
(a) The user can click an inner node to zoom.



(b) Zoomed into a branch of a UPGMA.

Figure 6.30: Zooming in the UPGMA tree.

The default value for the thresholds are adapted to the different views while the default values for colour coding and highlight are the same for all three views.



(a) The user can click on any big circle to zoom in.

(b) Zoomed in to the green circle.

Figure 6.31: Zooming in the MST

Based on the workflow of the microbiologist the colour code default is set to outbreak and the highlight default is set to last upload. Those values let the microbiologist immediately see if the latest batch of samples are part of an outbreak or not. The annotation defaults were set to core genome percentage, MLST type and date based on information obtained in previous iterations. In iteration one it was discussed that it is important to show the infection control unit the uncertainties in the analysis. There is often an interval of distances where it is difficult to say if the samples are part of the same outbreak or not. The best idea from the ideation to show this was to use two thresholds, the green dots on the distance axis shown in figure 6.30b. Below the lower threshold it is certain that the samples are part of the same outbreak and between the upper and the lower it is uncertain but possible. This is visually encoded in the trees by lower opacity on the edges and nodes of the uncertain samples.

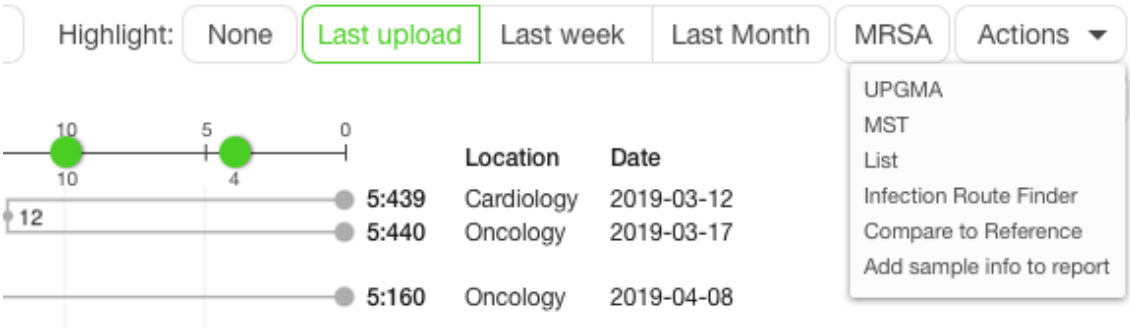


Figure 6.32: Actions menu for highlighted samples.

From the user tests in iteration one, section 6.3.4, it was apparent that the selection and highlight features were very confusing to the users. They did not understand the difference and the selection feature had too many steps in order for the user to reach their goal. In this ideation session, solutions for this problem was ideated upon. The result was to combine the two features into one. The idea is that the user can perform certain actions on the highlighted samples, these ac-

tions can be found in the actions drop down in the toolbox, shown in figure 6.32. Apart from choosing to show the highlighted samples in a new UPGMA tree, MST or list the user can choose to show them in the infection route finder or compare to reference.

The infection route finder, in figure 6.33, displays the samples in a diagram with time on the x-axis and location on the y-axis. In order to make it easier to see in what location the sample has been taken each location has a specific colour. The larger circles means that it includes several samples, just as in the MST.

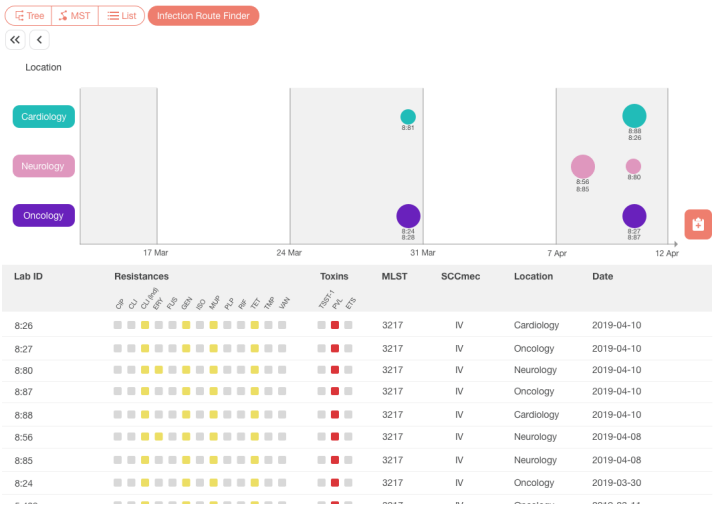


Figure 6.33: The infection route finder.

When the user chooses to compare the highlighted samples to a reference sample they enter a view that is split in half, shown in figure 6.34. The right side contains a list with all the samples information and the left side shows either a UPGMA tree or a MST of the samples. On both sides is the reference sample highlighted. This view was included based on the result of the focus group in the first iteration, section 6.3.5.

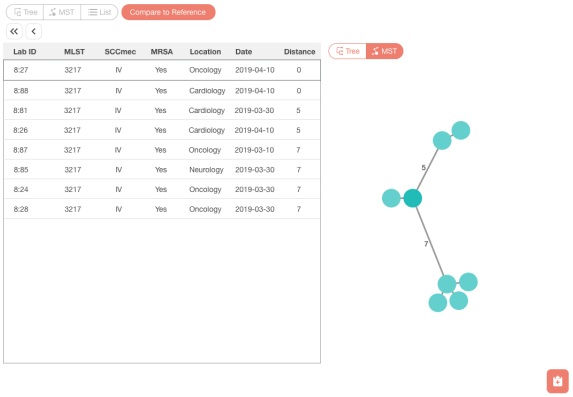


Figure 6.34: The compare to reference view.

At any time the user can add a snapshot of the current diagram to the report by clicking the coral button with a white "add to report" icon, shown in figure 6.35. A miniature of the diagram is then added to the column at the right hand side, so that the user always can see what is currently included in the report. When the user is happy with the selection of diagrams and has chosen the order of them by dragging and dropping they click the complete button. The user can now add explanations to the diagrams as a group by using the text box at the top in figure 6.36, or each individual diagram by using the text boxes next to the diagrams. Then the report is ready to export or to save in the platform.

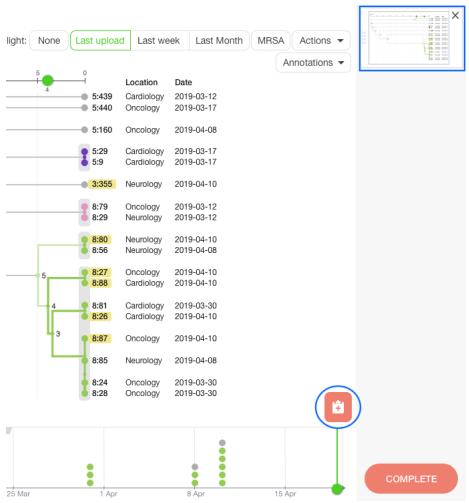


Figure 6.35: Add to report.

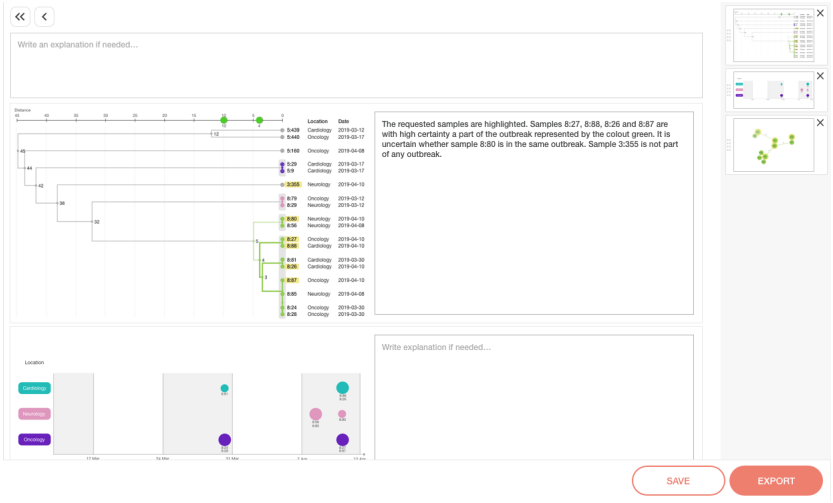


Figure 6.36: Complete report.

6.5.2 Focus Group

The purpose of the focus group was to discuss a couple of the guidelines and how well the prototype follows them. Before the discussion started a demo of the entire concept was given with the help of the high fidelity prototype. The guidelines discussed were Direct attention to the answer and Good defaults. The focus group consisted of the CTO, COO and Sales person at 1928 Diagnostics and the meeting lasted for 90 minutes.

6.5.3 Expert Interview

A 30 minutes long interview was held with a microbiologist who has NGS reporting as one of her work tasks. The purpose of the interview was to gather information that could be user for refining some of the guidelines even further. The interview started with a short demonstration of the concept and some of the features in order to provide a foundation for the interview. Following the demonstration was a number of questions that were focused around the guidelines *Merge workflows*, *Add their own explanations*, *Reporting flexibility*, *Focus on tasks*, and *Good defaults*.

6.5.4 User Tests

User tests were performed in order to test how easy the interface is to interpret and what result users would expect from certain interaction. The three participants were bioinformaticians that work at 1928 Diagnostics. These were chosen because they have similar education and background knowledge as a microbiologist at a lab has. None of the participants had seen the interface in advance and each test lasted for approximately 30 minutes.

In the first part of the user test the participants were presented with two images of the interface, one with the horizontal phylogenetic tree and one with a MST. Before trying to interact with the interface they were asked to tell what they believed they could interact with, how they would interact with those things and what they expected to happen. A few guiding questions were asked in order to ensure that the most important parts were covered. Following this the participants were asked to perform three tasks that required them to navigate through the interface and use several features.

6.5.5 Questionnaire

The majority of the design process had so far focused on the microbiologists and their work, therefore a questionnaire was made and sent to an infection control nurse. The purpose of the questionnaire was to find out if the new designs of the diagrams were more easily interpreted by the infection control unit. This was done by constructing the questionnaire in a similar way as the questionnaire in the second iteration. The questionnaire consisted of four sections, each section had one image of a diagram followed by 2-3 questions. The diagrams included were

a horizontal phylogenetic tree with coloured outbreaks, a horizontal phylogenetic tree with location colour coded, a MST of the relevant samples and the infection route finder. Each diagram had an explaining text that served as an example of what the microbiologist might write in a report.

6.5.6 Results

The results from the focus group, expert interview, user test and questionnaire are presented in this section.

6.5.6.1 Focus Group

The first guideline to be discussed in the focus group was *Direct attention to the answer*. Assuming that the most important thing is to find an outbreak in the phylogenetic tree and MST the conclusion was that the visual elements, such as colour, thicker lines and highlights, fulfils this guideline. The group thought that the infection route finder had problems with this since the colours had changed meaning from outbreak to location. That highlighted a problem with continuity within the interface and emphasised the importance of that issue. Another improvement point for the infection route finder was that, if admission and discharge dates were available, it could be shown how long a patient had been at one location. This requires the information to be available for the person who enters it into the platform, currently many hospitals in Sweden can not do this.

The discussion continued with the guideline *Good Defaults*. The conclusion was that it could be good with different defaults depending on how zoomed in you are. When there are only a few samples in the horizontal tree for example the defaults should be more adapted to the reporting and when there are more samples they should be adapted to the analysis. Since there are so many different bacteria that need different defaults it would probably be best if the user did not set them themselves, it is enough that the values can be adjusted in the interface.

One topic that has been discussed throughout the design process is how to show that two samples are from the same person. At the meeting the group was shown a couple of design suggestions with symbols representing the same patient. The common opinion was that it might be too much visual information in the diagrams. The group also felt unsure whether this was a desired feature among users or not.

6.5.6.2 Expert interview

One of the design guidelines that was evaluated during this interview was *Focus on Tasks*. The microbiologist felt like she would be able to perform her most important tasks quite fast and easy, especially making the results visually simple for the recipient of the report. In the reporting workflow the communication with the infection control unit is the most important task, which is being simplified by both the flexibility in the creation of the report and the use of colours in the diagrams. This allows the microbiologist to adapt the report to the way the recipient has the

easiest to interpret the results. The flexibility, which also is expressed as a design guideline, also makes it possible to create reports for other purposes than only infection control. The microbiologist saw opportunities to use the reporting feature when writing about research results or presenting them for other scientists. This lies outside the scope of reporting results for infection control but could still add value to the user.

By merging the analysis and reporting workflows the idea was to let the microbiologists add content to the report while performing their regular analysis tasks. The microbiologist that was interviewed believed that it required some getting used to in order to benefit from the merged workflows. The reason was that it is different to the way many work today, which is analysing the results first and then constructing a report. She also mentioned that it would make it possible for the microbiologists to experiment with different workflows, which is beneficial since there are no standardised methods for how to work with NGS data within infection control yet. For this reason and that it is often very specific questions that need human interpretation being asked she would currently not be helped by an auto generated report or explanatory texts for the diagrams.

The last guideline being discussed in the interview was *Good Defaults*. The microbiologist said that having good defaults serves two purposes in this case. If the user is new to NGS and cgMLST analysis results it can greatly help if the defaults for outbreak thresholds, highlighting and colour coding are good. She would expect that the threshold defaults are based on the current literature and the other two set to latest upload and outbreaks respectively. The annotations on the other hand should be set to defaults that would help the reporting, for example only time and location.

The last part of the interview was not directly connected to the guidelines but touched upon some of them. When discussing the infection route finder the microbiologist said that it is important that the information about the relations between the samples is visible in the diagram, by having the same colour for example. The view for comparing a group of samples to a reference was disliked for a similar reason, she thought it would be confusing to have a different view for this purpose and would prefer to just view them in the standard views. She also said that showing if two samples are from the same patient is extremely important in some situations, for example in the infection route finder. That would make it possible to do research on how much a bacteria evolves within a person over time and other things not directly connected to the current infection control workflow.

6.5.6.3 User tests

The general result of the users test was that the participants find it quite easy to interpret the interface and point out what they could interact with. Two of the participants did not mention the green sliders on the distance axis and the timeline at first, but given that they were not surprised when asking specific questions about them it could have been that they just forgot to talk about them. One thing

that differed between the participants was the expectations about what would happen when shrinking the time interval. The idea is that the user can choose what to do with the samples within the chosen interval. The default would be to dim the excluded samples and after that the user can choose to redraw a new tree or highlight. One participant expected the dimming and the other two that the excluded samples would disappear from the tree, and thus form a new tree.

All participants expected a couple of interactions to have the same behaviour as in the current visualisation in the platform. They expected that when clicking on an inner node in the UPGMA tree it would zoom in to that branch, which is a behaviour I have chosen to keep. None of the participants thought of the possibility to right click and what would happen then. They also thought that by clicking a leaf node they would get to a page with all the details about the analysis of that particular sample. This would not happen but two of the participants said throughout the test that they probably would figure things out if they were given a couple of minutes to just click around.

When performing the tasks all three participants had some trouble with understanding the highlighting and the connection to the actions drop down button. They all managed to complete the tasks but could not understand how they achieved the result. This was explained after the test and two of the participants believed that if they could click around int he interface they would understand the connection quite fast. One participant also believed that if it was clearer that the date shown in the tree is not the analysis date but the date the sample was taken, it would be easier to understand that the highlighted samples were the last uploaded.

6.5.6.4 Questionnaire

The questionnaire was answered by one nurse at the infection control unit. The main conclusion that can be drawn from the answers is that the MST and the infection route finder are the most valuable for the infection control unit. Compared to the phylogenetic tree, the participant found the MST visually easier to interpret since the samples that are part of an outbreak not only have the same colour but they are also closer together. In order to be able to understand the phylogenetic tree the participant said that a more in depth and better explanation is needed than the one provided in the questionnaire. From the interviews conducted in the empathise phase it was clear that time and location data was important for infection control work. That goes in line with the appreciation for a time line in the infection route finder. But the current design of the diagram would just moderately useful in the infection control units work according to the answer in the questionnaire.

6.5.7 Updated Guidelines

After the evaluation of the high fidelity prototype the guidelines were refined further and two guidelines were added.

1. Focus on tasks Spend time to understand what tasks, within infection control, the users need to perform in the interface and design around them. Since the product is a tool for microbiologists in their daily work it is important that they can perform all necessary tasks without distraction [29]. This approach was also successfully tried in the clinical report design project explained in section **??**.

2. Adapt to the Reader Have the recipient of the report in mind when designing the appearance of the diagrams in the visualisation. From the secondary persona in section 6.2.2 it is apparent that the visual language needs to help the infection control unit to interpret the results. This could be achieved by using different gestalt laws and colour to highlight and emphasise different characteristics [18]. An efficient visual language could also help the microbiologist write explanations for the diagrams and thus improve the communication with the infection control unit which, according to the expert interview results in section 6.5.6.2, is the most important task in the reporting workflow.

3. Flexibility Since there are different infection control organisation types, explained in section 6.2.2, the tool should be flexible enough to suit different ones. From the report review in section 6.1.2 it can be seen that the reports are made differently in order to to suit the hospital's different workflows. Another advantage of the flexibility is that the microbiologist can make report for other purposes than only infection control, for example research papers which is a another common work task for microbiologists working with this technology.

4. Emphasise interaction possibilities Work with affordance and pliancy in the interface in order to help the user understand how to interact with the visualisation. By using different types of hinting in the interface the user can find the most efficient way of performing certain tasks [25]. This is a common interaction design principle that should be followed [24].

5. Defaults adapted to goal Let the user enter the visualisation with good default values entered. Since, according to the primary persona in section 6.2.2, time is a limited resource having good defaults value might enable the user to make the analysis faster.

6. Merge Workflows Incorporate the report creation workflow in the analysis workflow in order to reduce the amount of extra work the microbiologist has to do. The paper prototype user test results, in section 6.3.6.1 suggested that this concept could be a good example of this.

7. Familiarity Diagrams that are commonly used in the field should keep familiar features. The results from the user questionnaire, explained in section 6.4.3, suggests that users appreciate information being visually encoded in traditional designs of diagrams. By keeping the most prominent features the user might be able to interpret the diagrams quicker.

8. Direct attention to the answer Use visual encoding in a way that shows the answer to the most common questions at first glance. The user questionnaire

results, in section 6.4.3, shows that the user preferred redundant encoding for outbreak information since they immediately can see which samples are part of an outbreak. This also makes it possible to use some of the encoding for other attributes, for example colour coding location instead of outbreaks.

9. Continuity Similar attributed should mean the same thing when navigating between views in order to maintain a sense of continuity. In the results from both the focus group, in section 6.5.6.1, and the expert interview, in section 6.5.6.2, it was suggested that especially colours should mean the same things when switching view. Otherwise the user might perceive the interface as confusing which can have negative consequences on the reporting.

7
Results

The project produces two results, a set of design guidelines and a user interface prototype that demonstrates how they can be used. Both parts will be described in this chapter.

7.1 Final Guidelines

The design process resulted in seven design guidelines that should be considered as a starting point for further development and research. Therefore, they include suggestions for what to think about when designing an interface for analysing NGS data and create reports used for infection control at hospitals. The guidelines have been updated after every concept development iteration based on knowledge acquired from the user tests and focus groups, and are summarized in this section as the final set of guidelines.

7.1.1 Focus on tasks

> ***Focus on Tasks:*** *Design around the tasks that microbiologists need to perform when analysing the NGS data for infection control.*

When having an interactive visualisation that is used for analysing data it falls under the category of an exploratory visualisation [20]. In order for an exploratory visualisation to be a good tool the designer needs to make sure that the user can perform all necessary tasks to reach their goals. In this case the goal for the microbiologist is to find out if one or a group of samples are part of an infection outbreak or not. All functionality that does not provide value for this purpose should be considered to be removed in order to reduce distractions from the important ones.

Figure 7.1: A hierarchical task analysis can be a good tool when designing around tasks [34].

7.1.2 Adapt for Recipient

> ***Adapt for Recipient:*** *Acknowledge the report recipients knowledge level when designing the diagrams.*

In the case of reporting results from an advanced analysis to people who needs to act on the results without having sufficient domain knowledge, the designer needs to think about how to make the information easier to understand. For this purpose it might be valuable to take advantage of how the human brain processes visual images[18]. By working with different gestalt laws and colour, the designer can provide a visual language that both helps the recipient of the image to understand the information and helps the microbiologist to explain the information[18].

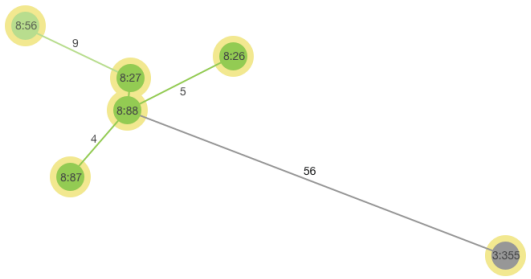


Figure 7.2: A visual language with information encoded in both colour and line length can be easier for the infection control unit to understand.

7.1.3 Immediate Answers

> ***Immediate Answers:*** *Provide the answer visually to the users most common questions immediately.*

Since time is a valuable resource for the microbiologist they do not want to spend an unnecessarily long time looking for the information they want to see. The designer can help the user with this by visually emphasise important information in the diagrams and provide good defaults for variable things in the interface. Redundant encoding, such as colour plus thick lines, is one way to draw the user's attention to the relevant parts of the diagram. During the user research it was found that the microbiologists wanted access to all available information about the samples. Although this might be good for their work as a whole only certain pieces of information is important when creating reports for the infection control unit. Therefore it should be considered to adapt the default values to the infection control reporting workflow if it is the main use of the interface.

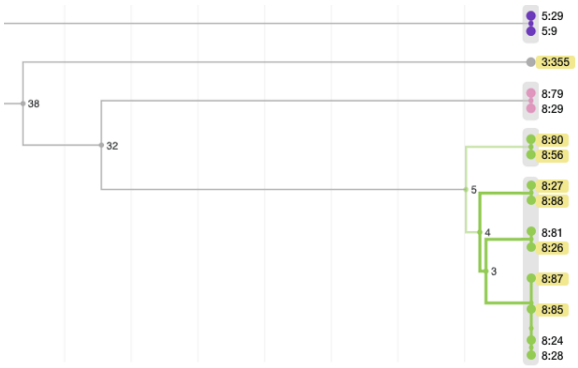


Figure 7.3: An example of when colour and line thickness emphasise the outbreaks in the diagram.

7.1.4 Flexibility

> ***Flexibility:*** *Consider introducing flexibility so that the user can shape their own workflow.*

The amount of flexibility the interface allows will determine the range of hospital organisation structures and purposes the interface can be used for. Since the infection control can be organised in many different ways there are many different requirements on what should be included in the reports and how that information is presented. If the microbiologist can shape the report as they want they can fulfil the requirement for their own hospital. This application of the NGS technology is also quite new so there are no standards for how to work with it. A flexible reporting workflow allows for experimentation with how the reports should be constructed for best infection control results. Another advantage that is outside the scope but should be considered is that with more flexibility in the report designs the reports can be used for other purposes as well, for example research papers and presentations.

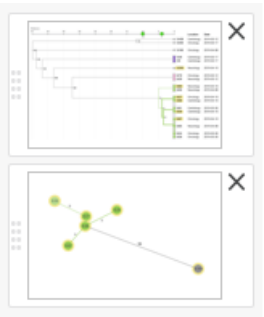


Figure 7.4: An example of flexibility is to let the user change order of diagrams in the report of delete them.

7.1.5 Domain Familiarity

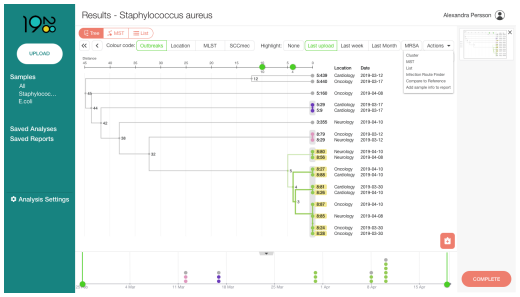
> ***Domain Familiarity:*** *Keep characteristic features of diagrams commonly used in the domain.*

By changing the design of fundamental characteristics of commonly used diagrams it might make it harder for the microbiologist to interpret the results because they are used to a specific design. Even though a new design is more appealing to and more easy interpreted for another group of people it should be considered to keep the traditional design. When being presented with something familiar the user might be more likely to reach their goals faster and easier, which is an important thing for the microbiologists.

7.1.6 Continuity

> ***Continuity:*** *Keep the meaning of visual attributes the same between screens.*

Continuity in the interface is important for decreasing the mental effort for the user to interpret the information in different views. When including many different diagrams this becomes even more important. The designer can achieve this by maintaining the meaning of colours when navigating through and between diagrams and keep the user in control of when they switch meaning.



(a) Highlighted cluster to be shown in infection route finder.

(b) Infection route finder of the green cluster.

Figure 7.5: Maintaining the cluster colour between views.

7.1.7 Emphasise interaction possibilities

> ***Emphasise interaction possibilities:*** *Improve the ease of use by clearly showing the user what objects are interactive in the interface.*

The designer should work wisely with affordance and pliancy of objects in the interface in order to help the user interact with it. If the user can perceive affordances easily it has the potential to increase their productivity.

7.2 Final Prototype

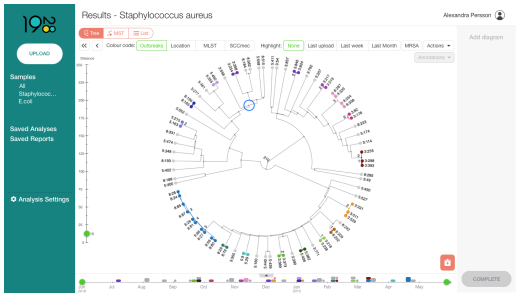
The final prototype in this project can be seen as a demonstration of how the design guidelines can be applied. The idea behind the prototype was to create an interface that allows the microbiologist to create an infection control report that is customized for the needs of their hospital and the current case. While the microbiologist explores and analyses the visualised results they can at any time add the current diagram to the report. The interface is divided into four parts, as shown in figure 7.6. A toolbar for certain action at the top, the main visualisation area in the middle, the time line at the bottom and the staging area for the diagrams to the report in the column on the right hand side. The different features of the interface and how they relate to the guidelines will be described in this section.

Figure 7.6: An overview of the interface with the four different part highlighted.

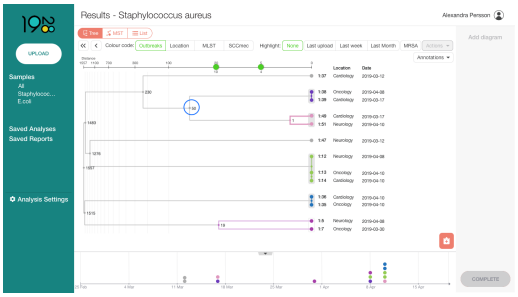
7.2.1 Levels and Zooming in Diagrams

Before entering the visualisation part of the platform the user chooses what samples to work with. Which diagram that is presented first to the user depends on the number of selected samples. If approximately 100 samples have been chosen the user first sees a circular phylogenetic tree. Because of the circular layout of the leaf nodes no annotated information is shown beside the samples since this would be hard to read and quite unnecessary when looking at this amount of samples, since the purpose often is to get an overview of the relationships between the samples. The user can zoom in phylogenetic trees by clicking the inner nodes, for example when clicking the circled node in figure 7.7a the user gets zoomed into the tree in figure 7.7b. Since the number of samples now have

decreased to under approximately 25 the phylogenetic tree is drawn horizontally. This view allows for adding extra information annotated beside each node. By clicking an inner node in this tree the user can zoom in further to for example the tree in figure 7.7c.



(a) Circular phylogenetic tree.



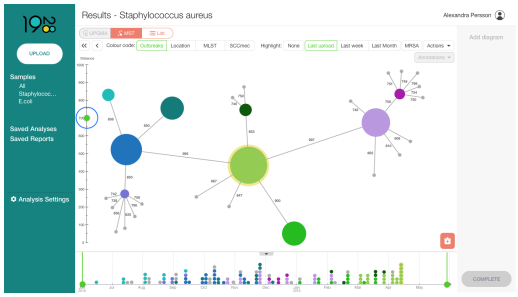
(b) Horizontal phylogenetic tree.

(c) Small branch of a phylogenetic tree.

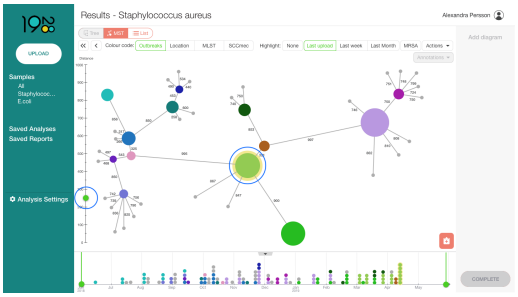
Figure 7.7: Zooming in a phylogenetic tree by clicking on the inner nodes.

If the user has chosen to work with more than approximately 100 samples the first diagram presented is a MST. In order to minimize the number of visual object on the screen and make the diagram less cluttered samples that are close together are grouped into visual aggregates. Every large circle is a visual aggregate and the size of it indicates how many samples it includes. The light green circle in the middle of figure 7.8a contains more samples than the dark green circle above it. By clicking the visual aggregates the user zooms into the samples within it. In order to control the number of nodes contained in a visual aggregate the user can change the threshold value by moving the slider that is circled in figure 7.8a. The threshold determines that all samples within that distance will be grouped together, so when lowering it some of the visual aggregated may split up into smaller ones, as shown in figure 7.8b. When clicking the green circle the user is zoomed into the MST in figure 7.8c.

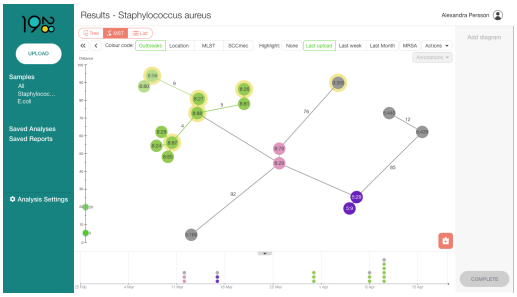
The user can at any time switch between the UPGMA view, the MST view and a list view by clicking the coral coloured buttons at the top of the screen. When there are too many samples for a phylogenetic tree that button is disabled. When switching between the different views the colour coding remains the same which



(a) MST with visual aggregates.



(b) MST with lower threshold for visual aggregate.



(c) Small MST with individual samples.

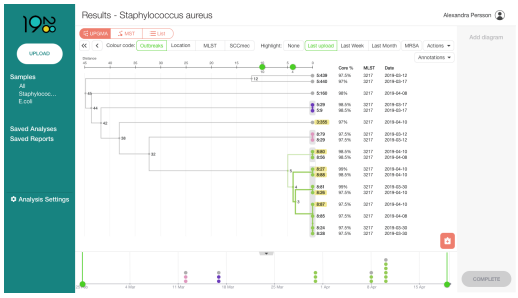
Figure 7.8: Zooming in a MST by changing distance threshold and clicking on visual aggregates.

is in line with the guideline *Continuity*. The fact that he user is able to choose in which diagram type they want to see their samples is a good example of *Flexibility* in the interface.

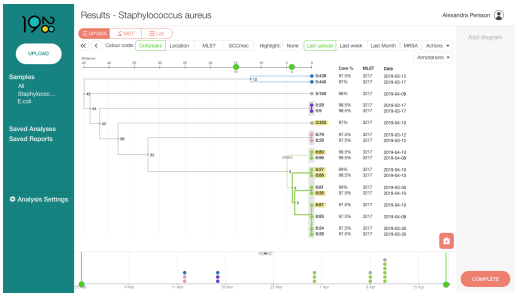
7.2.2 Cluster and Outbreak Detection

One of the most important tasks fort he microbiologist is to find outbreaks. To help them do this the samples are coloured according to if they belong to an outbreak and which one. What samples that are included in an outbreak is determined by a threshold value of the distance, which means that all samples that are closer in distance than a specific value are part of an outbreak. Between a certain interval of distances it is difficult to determine if the samples are included in an outbreak or not. Therefore, there are two sliders on the distance axis above the phylogenetic tree, which is shown in the images in figure 7.9. The user can at any time change the threshold values by moving the sliders. This will result in the cluster colour coding changing. If the user moves the upper threshold from 10, as shown in figure 7.9a, to 15, as shown in figure 7.9b two samples become blue because they might constitute an outbreak. To visualise the uncertain outbreak members the lines to the samples have lower opacity than the lines for the certain samples. There is also different thickness on the lines, the certain outbreak having the thickest, the uncertain slightly thinner and the samples outside the outbreak

having the thinnest. This is an example of how the guideline *Immediate answer* can be applied.



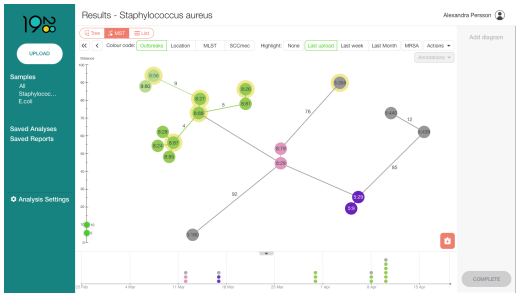
(a) Thresholds at 4 and 10.



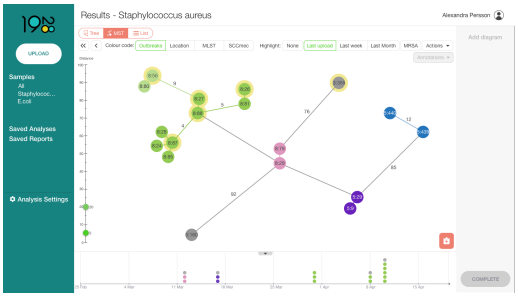
(b) Thresholds at 4 and 15.

Figure 7.9: Changing the scope of the outbreaks by changing the threshold values in a phylogenetic tree.

In an MST with only individual samples the cluster colours and thresholds work the same as in the phylogenetic tree. In the figures 7.10a and 7.10b the upper threshold is moved from 10 to 20 which results in the new possible outbreak coloured blue. A difference in this diagram is that the nodes also vary in opacity, which the uncertain samples having lower opacity.



(a) Thresholds at 5 and 10.

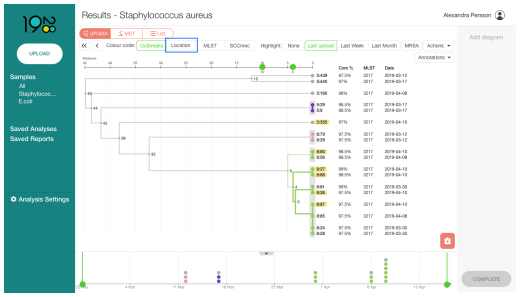


(b) Thresholds at 5 and 20.

Figure 7.10: Changing the scope of the outbreaks by changing the threshold values in a MST.

7.2.3 Colour Coding

In both the phylogenetic tree and the MST the user can choose to colour code certain attributes of the samples by using the buttons in the toolbar. The default colour coding value is outbreaks since this is what the microbiologist most often is interested in seeing first. When clicking the location button, as shown in figure 7.12a, the colour coding of the samples changes to representing location instead, as shown in figure 7.12b. Even though the colour coding is changed the other encodings for outbreaks remain since it is often relevant to see the location distribution in relation to the outbreaks.

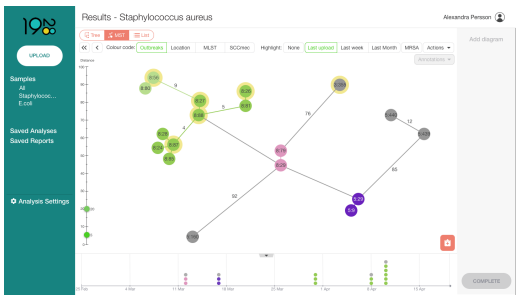


(a) Outbreaks colour coded.

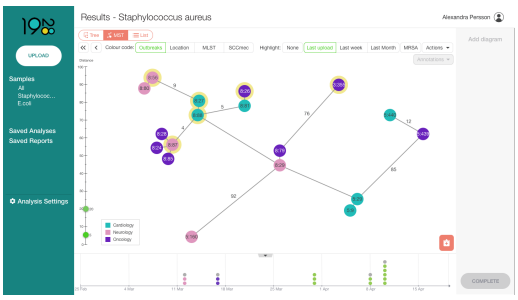
(b) Location colour coded.

Figure 7.11: Changing colour coding from outbreak to location in a phylogenetic tree by clicking the toggle button in the toolbar at the top.

The colour coding is identical in the MST diagrams with individual samples, shown in figure 7.12. When visual aggregates are present in the MST the distribution of the attribute is shown as a pie chart with the colours, as shown in figure **??**. In all diagrams a legend explaining what the colours mean appears in the bottom left corner of the visualisation. The colour coding is an example of how the guidelines *Flexibility*, *Adapt for Recipient* and *Continuity* can be applied since the user can choose what information to emphasise in the report and also that the colour coding stays the same between views.



(a) Outbreaks colour coded.



(b) Location colour coded.

Figure 7.12: Changing colour coding from outbreak to location in a MST by clicking the toggle button in the toolbar at the top.

7.2.4 Highlighting

The highlighting feature was originally added so that the microbiologist would be able to see where their latest uploaded samples are located in the different diagrams. Thus, the default value of the highlight is the latest uploaded samples, both in the phylogenetic tree and the MST. In both types of diagram the user can select to highlight samples within another time frame or with a special attribute in the toolbar above the visualisation. By clicking the leaf nodes or the sample ID in the phylogenetic tree the user can add or remove the highlight on individual samples. Figure 7.13 shows what the highlight looks like in a phylogenetic tree.

In a MST with individual samples, as shown in figure 7.14a, the user can add and remove highlight by clicking the nodes similar to the phylogenetic tree. However, the highlight in a MST where visual aggregates are present, as in figure 7.14b works slightly different. If a visual aggregate is highlighted it can in some cases mean that the user has highlighted all samples in the big circle by right clicking on it. In the other cases it indicates that the visual aggregate contains some of the samples that the user has chosen to highlight using the buttons in the toolbar **??**. In both diagrams the highlight is synced with the time line, meaning that the samples that are highlighted in the diagram is also highlighted on the timeline.

The highlighting feature is an example of how the guidelines *Adapt for Recipient* and *Immediate Answer* can be applied. The highlight can serve as a communication tool for the microbiologist to communicate certain information in a simple

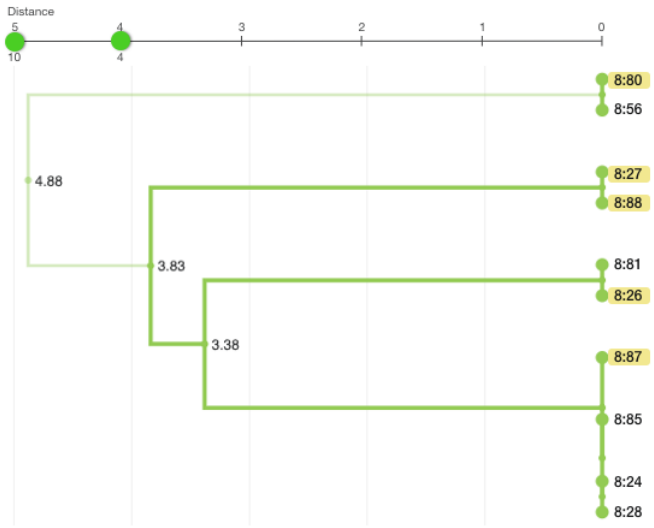
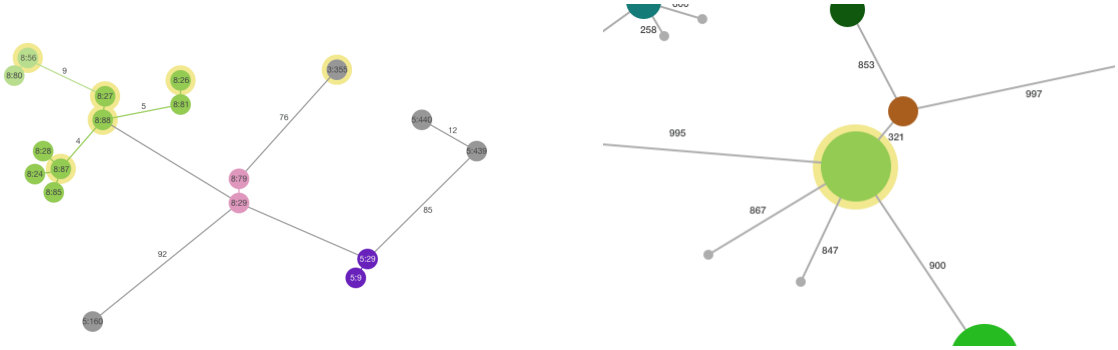


Figure 7.13: Highlighting in a phylogenetic tree.



(a) Highlight in small MST. **(b)** Highlight in large MST.

Figure 7.14: Highlighting at different levels of a MST.

way to the infection control unit. But at the same time it helps the microbiologist to see if the relevant samples are part of an outbreak or not.

7.2.5 Actions

The highlighting is not only used for making certain samples more visually prominent, the user can also choose to perform certain actions with the highlighted samples. When opening the drop down menu that is adjacent to the highlighting buttons the user is presented with a number of actions, shown in figure 7.15. When clicking an option in the actions menu only the samples that are currently highlighted will be included in the actions. Four of the actions will display the samples in a new diagram, namely UPGMA, MST list and infection route finder, while two of the actions will either add sample info to the report or create a saved group with the samples.

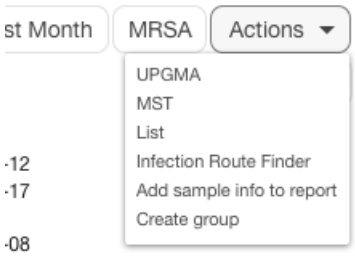
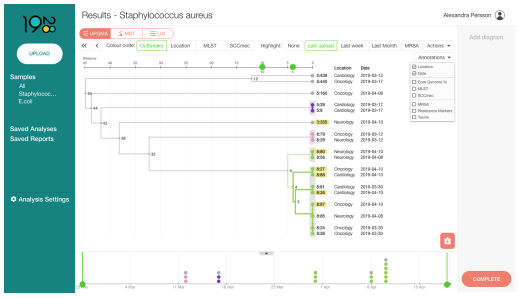


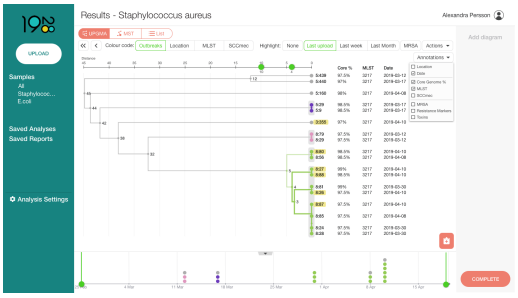
Figure 7.15: he infection route finder with a group of samples that constitutes an outbreak.

7.2.6 Annotations

In the horizontal phylogenetic tree there is room for a number of annotations on the right hand side. The default annotations are adapted to what kind of information the infection control unit is interested in, namely date and location of where the sample was taken.



(a) Annotated with date and location.



(b) Annotated with MLST and core genome percentage.

Figure 7.16: Choose what annotations to include in the diagram with the annotations drop down

The user can choose what annotations they want to show by opening the drop down menu for annotations, as shown in figure 7.16a. In figure 7.16b the user has selected show core genome percentage, MLST type and date. This is also an example of how the microbiologist can adapt the diagrams to the recipient and the default values are an example of the guideline *Immediate answer*.

7.2.7 Time Line

The time line displays all the samples in the UPGMA or MST according to the date they were taken. Thus it provides a way for the user to see the distribution over time. Since the colour coding selected for the trees applies to the time line as well they can also see the distribution over time for different locations. In figure 7.17 the sliders on the time line has been moved so that they create a window.

Everything outside this window is greyed out on the timeline and dimmed in the tree diagrams.

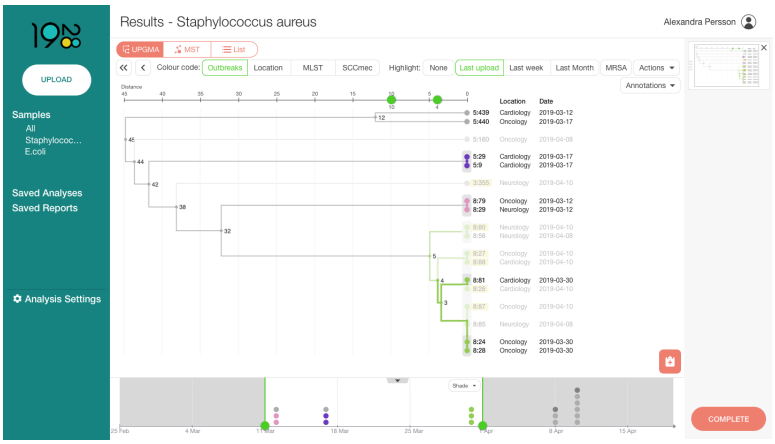


Figure 7.17: Using the time line sliders to filter in the tree.

Since the user might want to use this feature for different purposes depending on their task at hand they can choose what they want to do with the samples within the selected time frame. When hovering over the time line a drop down button appears, shown in fig 7.18, with which the user can choose to either highlight the selected samples or create a new tree instead of dimming out the not selected samples.

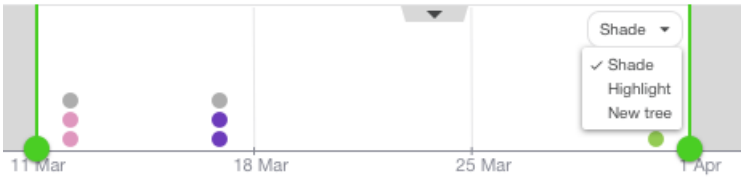


Figure 7.18: The different actions for the selected samples on the time line.

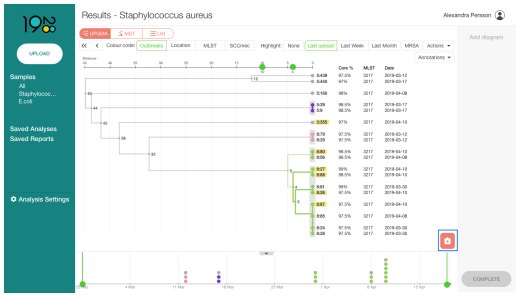
7.2.8 Infection Route Finder

The infection route finder is reached through the actions menu since the most likely scenario for using it is to investigate an outbreak further. The purpose of this diagram is to display how an infection outbreak has moved between places over time. With this information it is easier for the user to see what needs to be done in order to stop the outbreak. The samples in the infection route finder maintain the same colour coding as in the previous view, as shown in figure 7.19, and thus applying the *Continuity* guideline. This means that it would be easy to differentiate between different outbreaks if more than one is present.

Figure 7.19: The infection route finder with a group of samples that constitutes an outbreak.

7.2.9 Reporting

A central part of the interface is the way the user builds the report that they want to send to the infection control unit. At any time when working in the visualisation the user can add the current diagram to the report. This is done by clicking the coral coloured add button at the bottom right corner of the diagram, shown in figure 7.20a. When the button is clicked a miniature image of the diagram is added to the column on the right hand side of the screen. This column can be seen as a staging area for the report. The user can change order of the diagrams in this column as well as remove diagrams.



(a) Add diagram to report by clicking add button.

(b) Complete report by clicking complete button.

Figure 7.20: Adding a diagram to the staging area for the report and completing it by using the add and complete buttons.

When the user is happy with the selection of diagrams they click the *Complete* button, as shown in figure 7.20b. The next page contains larger images of the diagrams and text fields next to them where the microbiologist can add explanations to each diagram, shown in figure 7.21. There is also a text field at the top of the page that can be used for general explanation of the results in total.

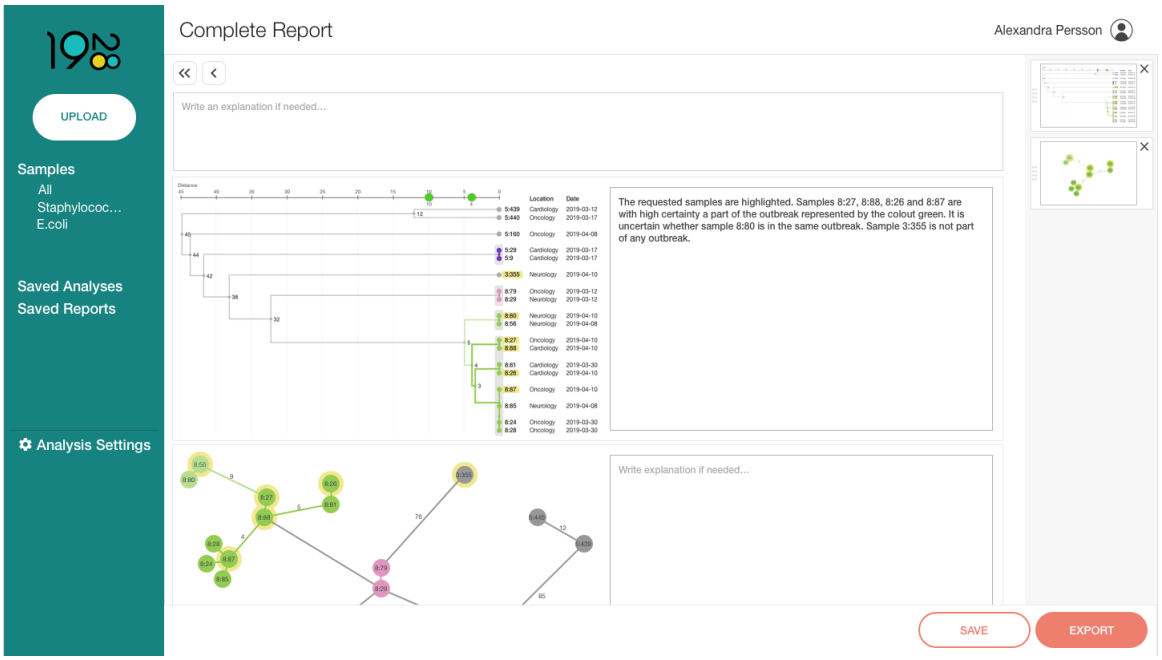


Figure 7.21: Add optional explanations to the report.

8
Discussion

The discussion is divided into four parts; results discussion, method discussion including the guidelines and the final prototype, ethical considerations, and lastly suggestions for future work.

8.1 Results

The results obtained from this project, both design guidelines and prototype, can be considered a good starting point for further research within the area. Although there are common practices of how to visualise this type of data it has not, to my knowledge, been investigated how they can be used and designed for the purpose of infection control at hospitals. The study described in section **??** is related to this area but is only focused on the final report, while this project has focused on how to create such a report. The guidelines defined in that study has been taken into consideration throughout the project and some of the final guidelines have been inspired by them.

Since hospitals only recently started to use NGS technology for infection control, it was learned from the interviews that, there are no common practices for how to work with it. Thus, every microbiologist has their own process that is shaped after how they work at their own hospital. This means that the design space for an NGS infection control tool has been quite open. However, it seems that several different concepts need to be designed and tested in order to reach a conclusion on what the best design practices are for this domain. Due to the lack of tools of this kind currently on the market it should be acknowledged that most designs that aims to serve this purpose probably would be considered an improvement from the microbiologists current workflow.

8.1.1 Design Guidelines

I believe that the design guidelines provide good suggestions for some of the most important aspects to consider when designing an interface for analysing NGS data and create a report that can be used when working with infection control. Based on the results in the study described in section 2.2.3 and the framework by Zhang et al. in section 4.2 the decision to perform a task analysis was made. Since it proved to be a good tool for understanding the most important tasks that

the microbiologist need to perform it helped shape the guidelines *Focus on tasks* and *Immediate Answers*.

In the beginning of the process the guidelines were inspired by common design principles from both interface design and information visualisation. The final guidelines are more focused on the domain specific aspects but they should still be used in addition to basic interaction design and information visualisation principles. The guideline *Emphasise interaction possibilities* however, is still very focused around the general importance of affordance and pliancy. The reason why it is included in the final guidelines is that I believe that these aspects are extra important when designing a tool that is meant to help the user to work efficiently.

There are a couple of aspects that are not covered in the guidelines and would need further research, the most important will be discussed in section 8.4 Future Work. In this project only one concept was designed, which the guidelines are based upon. This is one of the reasons why they can only be considered a starting point for further research. In order to make the guidelines more final they need to be evaluated with more design suggestions and further testing.

8.1.2 Final Prototype

The final prototype is a good illustration of how the design guidelines developed alongside it can be applied. Even though the guidelines are based on this prototype there are probably many ways in which they can be applied. For demonstration and testing purposes the prototype has interaction for specific use cases implemented. That level of fidelity was excellent for communicating the ideas and receive feedback on them during the focus groups and the expert interview in the last design iteration. However, it turned out that the user testing suffered from it. In order to get really good results from the user tests the prototype would have had to be more interactive and mimic a full implementation even more.

The main drawback of the final prototype is that the minimum spanning trees are not based on real data, but only prototyped using Sketch. Therefore, it has not been tested how well they would scale with a large amount of samples. However, as seen in section 2.2.2, this kind of diagram is often used to visualise similar data and has proven to work quite well for providing an overview of a large sample set, which was the part of the purpose for them in this prototype. Many characteristics of the MST are inspired by already existing designs, e.g. the visual aggregates and the pie charts showing distribution. Even though the minimum spanning trees in the prototype are not real they still manage to illustrate how the different features would work with it, for example the highlighting and colour coding. The zooming feature, however, would have needed more testing in order to see if the combination with the distance threshold and clicking on visual aggregates is a good solution. The brief tests in this study suggest that it is a solution that is easy for the users to understand.

There are a few concrete things that still at the final stage of the development has significant room for improvement. The most important one is the infection route finder. In the beginning of the design process it was not clear how valuable this diagram would be, therefore less time was spent on developing that concept. During the last evaluations of the prototype it was clear that this was one of the features that would add significant value to both the infection control unit and the microbiologist. If this had been discovered earlier more effort could have been put into how this type of diagram could be scaled for larger sets of samples and also for what situations it would be used. The current design assumed that only samples from one outbreak would be interesting to view in the diagram. However, the interview with the microbiologist in the last design iteration indicated that it might be interesting to display a larger amount of samples in it from a variety of outbreaks of that species. The diagram could also be used for research purposes such as investigating how much a bacteria evolves within a patient over time. It can however be argued that these things are outside the scope since it does not directly contribute to the creation of the reports but rather a deeper understanding of how the bacteria strands within the hospital behaves.

Another feature that needs further improvement and user testing is the highlighting and select feature. It was not clear for the participants that the actions in the actions menu as performed on the highlighted samples. The purpose of this feature was mainly to provide a way for the user to quickly, from anywhere in the visualisation, be able to choose a selection of samples and show only them in their own diagram. The confusion might be due to the quality of the prototype since the user test participants had limited possibilities to navigate through the interface and experiment with how things work. Nevertheless, this feature would need be developed further and preferable tested with a prototype with more functionality implemented. I find that it is difficult to test the actual usability of a feature with a prototype of this fidelity, instead it most often focuses on how easy it is to learn how to use the interface. Perhaps this kind of interface will always need some instructions or a tutorial before using it so that the user knows about all the features beforehand.

8.2 Execution

The iterative process explained in section 4.1 was chosen due to its flexibility which would allow the process to be adjusted throughout the project. The initial plan was to iterate over the entire design three times, since this seemed reasonable within the given time frame. However, since iteration two only focused on the diagram designs the interface at a whole was only iterated upon two times. The consequence of this was that it felt like it was two design processes going on in parallel instead of one combined.

The user study in this project was based on qualitative data from interviews, both conducted by me and by 1928 Diagnostics. I believe that it was a good approach to get an understanding of the users and what they would want in a product fo-

cused on infection control. The first plan was to create a questionnaire as well in order to get some quantitative data to back up some theories and ideas with. However, after reading interviews and conducting my own I realised that it would be very difficult to construct a questionnaire that would provide any valuable input to project. Instead I analysed the interviews and constructed personas based on my insights from them. One thing that I believe would have improved my user study is to conduct another interview with someone from a hospital where the organisational structure regarding infection control is different from the hospital where the microbiologist I interviewed worked. This information was unfortunately received too late in the process and I prioritised constructing the personas and starting ideating on solutions as soon as possible. Instead I have made assumptions on what the requirements would be in such an organisation.

During the initial research phase I noticed that the design process for information visualisations is often based on a task analysis and then focused on the tasks defined in it. I adopted this into my process by doing an hierarchical task analysis with the help of the primary persona and then base the initial brainstorming session on the identified tasks. Although the personas were a good way to summarize the user research and create a clear image of the users in the beginning of the project the most valuable thing for the remainder was the task analysis that was based on them. Brainstorming around each task was a good way to make sure that the most important tasks that the users need to perform in the interface are included. Also when coming up with new features it was good to go back to the task analysis to see if the feature could be used for any of the tasks or not. Since it is an exploratory interactive visualisation that would be used as a tool it was important to make sure that no distracting features were included and I feel like defining the tasks before hand helped me to ensure that.

The first design iteration resulted in a paper prototype on which user tests were performed. I chose to work with paper prototypes because it helps me to find good solutions to problems when I can try out different placements of objects physically or by drawing. While I was creating the paper prototype I also discovered some problems with different features which could be solved by immediately sketching on solutions and trying to combine them with the prototype. Even though the paper prototype provided significant value to the creative process the quality of the user tests might have been higher with a higher fidelity prototype from the beginning. It was apparent in the tests that the quality of the paper prototype made it unnecessarily difficult for the participants to understand the interface. Despite that, the user tests at this stage highlighted some important problems that had to be solved in later iterations.

The high fidelity prototype was created in two steps, namely using D3 for the UPGMA tree diagrams and Sketch for the MST diagrams and the interface as a whole. I found D3 to be very good for experimenting with different designs for the UPGMA trees since the learning threshold for this was quite low. It was also very valuable to be able to use real data to construct diagrams in order to see how they would fit the overall interface. However, when starting experimenting with creating

a MST in D3 it showed that the functionality needed was more time consuming to learn. At this point I had already spent quite some time on the other diagrams and since MSTs are commonly used for this purpose in other programs, I decided to create mock ups in Sketch instead in order to illustrate the other functionality.

When the prototype in Sketch was done some interactivity was added to it using InVision. The level of interactivity that this type of prototyping tools offer is very good for demonstrating the prototype and communicating the concept that has been designed. As a result it was very valuable to have discussions around it during the focus groups and the expert interview in the last iteration. However, since it is very time consuming to create an interactive prototype that behaves as if it is implemented, and also easy to make mistakes, the user tests may not have given a fair image of how users would interact with the finished product. I tried to work around this issue by asking questions and encouraging the participants to explain their train of thought throughout the user test.

Furthermore, in order to improve the quality of both the prototype and the guidelines more user tests should be carried out. This project would have also benefited from a larger amount of participants. Due to poor time management and planning towards the end only people from 1928 Diagnostics, with similar education as the persona, participated in the test. With better preparation real users might have been able to test the prototype. But since this was thought of quite late and the users of the 1928 platform are located in many different cities, it was decided that it would be enough with the employees at the company. And, as mentioned above, the quality of the prototype might not have been suitable for testing with real users of the 1928 platform at this stage.

8.3 Ethical Considerations

When working with healthcare products there are always ethical issues to consider. In this project the main ethical concern was the risk of people misinterpreting the information in the visualisation itself but also the resulting report. Both cases can have severe consequences for the patients as well as the hospital. Due to the level of control given to the microbiologist in the report creating process much of the responsibility to minimise the risk of the infection control unit to misinterpret results lays with them. Even though the design in this project has given the microbiologist the tools to create a clear and informative report the designer has lost its power to ensure that this actually happens. This raises the question whether the reports should be more automated and standardised which will be further discussed in section 8.4 Future Works.

The consequences of misinterpreted information, at any point of the reporting process, can be severe for the patients at the hospital. When there is an outbreak the wards need to make a decision on what they should do in order to stop it from spreading further. A common action is to close a ward and decontaminate it to get rid of the bacteria causing the outbreak. A closed hospital ward means that

the hospital's capacity to treat patients is temporarily decreased, which means that people in need of case might not be able to get it in time if they need to travel to another hospital. When a serious outbreak has occurred this is a necessary action that will protect all patients in the hospital from the disease. However, if this decision is based on inadequate or misinterpreted information, and there is no outbreak, these hospital beds have been unnecessarily removed. On the other hand, if the outbreak is undetected due to the analysis results being misinterpreted the disease might keep spreading, which leads to more patients falling ill and the hospital might have to close down more wards.

8.4 Future Work

Since this project can only be considered a starting point for the research in how the interactive information visualisations can be used for improving the reporting of NGS data for infection control purposes, there are areas that need to be further explored. In addition to keep improving the guidelines suggested in this project two main ares of future work have been identified.

When the number of samples increases so does the importance for helping the user keeping track of where in the visualisation they are. A topic for future research could be to investigate if and how animations should be used to solve this issue. By animating the transitions between different views and when zooming in a diagram it might be easier for the user to keep track of where the samples they were interested in are located in the new view. It also has the potential of making the user experience smoother and more cohesive. The designer need to make sure that the animations add value to the visualisation instead of distracting the user from what is important.

Another question that has emerged during the project is whether or not the report that is being sent to the infection control unit should be automated and standardised. In the prototype in this project the level of automation is none, the microbiologist is still in control of the content. Perhaps only a certain part should be automated and the microbiologist make the final touches before sending it. It should be investigated what level of automation works best. This is a quite extensive and complicated question since the findings of this project suggests that every case is so different that a microbiologists analyses are needed in order to create a good report for the specific question being asked. The risk is that the auto generated report will be too general for many cases.

9
Conclusion

1928 Diagnostics has developed a platform that performs an analysis on NGS data that can be used for infection control in hospitals but their current visualisation of the results does not satisfy the needs of the microbiologists that use it. Therefore, the aim of this project was to design a tool, that can be incorporated in the 1928 platform, for creating infection control reports of NGS data through interactive information visualisation. The research question that this project aimed to answer was:

"What should be considered when designing an interactive visualization of NGS data used for facilitating the creation of reports for infection control in hospitals?"

To answer this research question an iterative design process was carried out in which a prototype of an interactive visualisation was designed. In order to make informed design decisions background research within the domain and a user study were carried out.

The core of the concept is that the microbiologist can construct the report by adding diagrams to it while exploring the data and analyse the results. This approach gives the microbiologist the tools and flexibility to create a report that is suitable for their hospital's organisation and infection control workflow. The design of the diagrams helps the user to discover outbreaks and to decide whether a new patient is part of an already existing outbreak by smart use of gestalt laws and colour.

With the help of the prototype development the research question was answered by providing a set of seven design guidelines that can be considered when designing a similar user interface:

> ***Focus on Tasks:*** *Design around the tasks that microbiologists need to perform when analysing the NGS data for infection control.*
>
> ***Adapt for Recipient:*** *Acknowledge the report recipients knowledge level when designing the diagrams.*
>
> ***Immediate Answers:*** *Provide the answer visually to the users most common questions immediately.*
>
> ***Flexibility:*** *Consider introducing flexibility so that the user can shape their own workflow.*

Domain Familiarity: *Keep characteristic features of diagrams commonly used in the domain.*

Continuity: *Keep the meaning of visual attributes the same between screens.*

Emphasise interaction possibilities: *Improve the ease of use by clearly showing the user what objects are interactive in the interface.*

Bibliography

[1] (2019, February) Antibiotic resistance. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance

[2] A. J. McAdam, D. C. Hooper, A. DeMaria, B. M. Libago, T. F. O'Brien, and B. McCaughey, "Antibiotic resistance: How serious is the problem, and what can be done?" *Clinical Chemistry*, vol. 58, no. 8, 2012.

[3] (2019, May) The burden of health care-associated infection worldwide. [Online]. Available: https://www.who.int/gpsc/country_work/burden_hcai/en/

[4] J. Zhang, R. Chiodini, A. Badr, and G. Zhang, "The impact of next-generation sequencing on genomics," *Journal of Genetics and Genomics*, vol. 38, no. 3, pp. 95–109, 2011.

[5] J. Besser, H. Carleton, P. Gerner-Smidt, and E. Trees, "Next-generation sequencing technologies and their application to the study and control of bacterial infections," *Clinical Microbiology and infection*, vol. 24, no. 84, p. 335, 2018.

[6] (2019, May) 1928 diagnostics. [Online]. Available: https://1928diagnostics.com/

[7] M. E. Pearce, N.-F. Alikhan, T. J. Dallman, Z. Zhou, K. Grant, and M. C. Maiden, "Comparative analysis of core genome mlst and snp typing within a european salmonella serovar enteritidis outbreak," *International Journal of Food Microbiology*, no. 274, Jun 2018.

[8] C. Allende, E. Sohn, and C. Little, "Treelink: data integration, clustering and visualization of phylogenetic trees," *BMC Bioinformatics*, vol. 16, no. 1, p. 414, Dec 2015. [Online]. Available: https://doi.org/10.1186/s12859-015-0860-1

[9] S. Argimón, K. Abudahab, R. J. E. Goater1, A. Fedosejev, J. Bhai, C. Glasner, E. J. Feil, M. T. G. Holden, C. A. Yeats, H. Grundmann, B. G. Spratt, and D. M. Aanensen, "Microreact: visualizing and sharing data for genomic epidemiology and phylogeography," *Microbial Genomics*, no. 2, Nov 2016. [Online]. Available: https://doi.org/10.1099/mgen.0.000093

[10] I. Letunic and P. Bork, "Interactive tree of life (itol): an online tool for phylogenetic tree display and annotation," *Bioinformatics*, vol. 23, no. 1, pp. 127–128, Jan 2007. [Online]. Available: https://academic.oup.com/bioinformatics/article/23/1/127/188940

[11] Z. Zhou, N.-F. Alikhan, M. J.Sergeant, N. Luhmann, C. Vaz, A. P.Francisco, J. A. Carrico, and M. Achtman, "Grapetree: visualization of core genomic relationships among 100,000 bacterial pathogens," *Genome*

Research, no. 28, pp. 1395–1404, Jul 2018. [Online]. Available: https://genome.cshlp.org/content/28/9/1395.full

[12] B. Ribeiro-Goncalves, A. P.Francisco, C. Vaz, M. Ramirez, and J. A. Carrico, "Phyloviz online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees," *Nucleic Acid Research*, Apr 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4987911/

[13] A. Crisan, G. McKee, T. Munzner, and J. L. Gardy, "Evidence-based design and evaluation of a whole genome sequencing clinical report for the reference microbiology laboratory," *PeerJ*, Jan 2018. [Online]. Available: https://peerj.com/articles/4218/

[14] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, Dec 2012. [Online]. Available: https://ieeexplore.ieee.org/document/6327248

[15] R. Buchanan, "Wicked problems in design thinking," *Design issues*, vol. 8, no. 2, pp. 5–21, 1992. [Online]. Available: https://www.jstor.org/stable/1511637

[16] J. Kolko, *Thoughts on Interaction Design (Second Edition)*. Elsevier, Inc., 2011.

[17] W. Gaver, "What sould we expect from research through design?" *HCI*, May 2012.

[18] C. Ware, *Information Visualization - Perception for design*, 3rd ed., ser. Interactive Technologies Ser. Elsevier Science Technology, 2012.

[19] G. J. Myatt and W. P. Johnson, *Making Sense of Data III : A Practical Guide to Designing Interactive Data Visualizations*. John Wiley Sons, Incorporated, 2011.

[20] N. Iliinsky and J. Steele, *Designing Data Visualizations*. O'Reilly Media, Inc., 2011.

[21] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," *Proceedings of IEEE Visual Languages*, 1996.

[22] N. Elmqvist and J.-D. Fekete, "Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 3, May 2010.

[23] A. D. Norman, *The design of everyday things*. Basic books, 2002.

[24] ——, *The design of everyday things - Revised and expanded edition*. Basic books, 2013.

[25] A. Cooper, R. Reimann, D. Cronin, and C. Noessel, *About Face, The essentials of interaction design*, 4th ed. John Wiley Sons, Inc., 2014.

[26] S. Dakin and A. Herbert, "The spatial region of integration for visual symmetry detection," *Proceedings of the Royal Society of London*, vol. 265, no. 1397, Apr 1998.

[27] R. Dam and T. Siang, "5 stages in the design thinking process," Feb 2019. [Online]. Available: https://www.interaction-design.org/literature/article/5-stages-in-the-design-thinking-process

[28] A. Pratt and J. Nunes, *Interactive Design : An Introduction to the Theory and Application of User-centered Design*. Quayside Publishing Group, Nov 2012.
[29] J. Zhang, K. A. Johnson, J. T. Malin, and J. W. Smith, "Human-centered information visualization," 2008.
[30] C. Chen, "Top 10 unsolved information visualization problems," *IEEE Computer Graphics and Applications*, vol. 25, no. 4, Jul 2005.
[31] J. Nielsen, "Interviewing users," Jul 2010. [Online]. Available: https://www.nngroup.com/articles/interviewing-users/
[32] J. Blomberg, J. Giacomi, A. Mosher, and P. Swenton-Wall, "Ethnographic field methods and their realtion to design," *Participatory design - Principles and Practices*, 1970.
[33] R. Dam and T. Siang, "Personas – a simple introduction," Jan 2019. [Online]. Available: https://www.interaction-design.org/literature/article/personas-why-and-how-you-should-use-them
[34] A. Komninos, "How to improve your ux designs with task analysis," Mar 2019. [Online]. Available: https://www.interaction-design.org/literature/article/task-analysis-a-ux-designer-s-best-friend
[35] H. Lamm and G. Trommsdorff, "Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review," *European Journal of Social Psychology*, Oct 1973.
[36] A. Stein Wellner, "A perfect brainstorm," *Boston inc.*, no. 10, Oct 2003.
[37] R. Hartson and P. Pardha S., *UX Book*. Elsevier Science technology, 2013.
[38] C. Snyder, *Paper Prototyping, The Fast and Easy Way to Design and Refine User Interfaces*. Interactive Technologies, 2004.
[39] (2019, May) D3 - data-driven documents. [Online]. Available: https://d3js.org/
[40] (2019, May) Sketch. [Online]. Available: https://www.sketch.com/
[41] (2019, May) Invision. [Online]. Available: https://www.invisionapp.com/
[42] Y. Wadsworth, *Do it yourself social research*, 3rd ed. Allen Unwin, 2011.

A

Tasks for the Paper Prototype User Test

Scenario 1:

1. Cluster all samples in the list.
2. Show only the samples from May
    - *Question: What do you expect to see? The same tree with samples faded or a new clustered tree with only the samples in may?*
3. Show all he samples again
4. Find all clusters with a threshold of distance 3 and highlight the newest uploaded samples.
5. Show the MLST type of all samples
6. Add this graph to the report.

Scenario 2:

1. Colour code the samples by location
2. Highlight only sample 3 and 7.
3. Add this graph to the report.
4. Remove the highlight

Scenario 3:

1. Select 3 samples
2. Show the minimum spanning tree of these samples.
3. Go back to the tree you started with
4. Zoom in to a cluster
5. Go back to the tree you started with

Scenario 4:

1. Show the minimum spanning tree
2. Zoom into a cluster
3. Go back to the original minimum spanning tree
4. Find a sample that has MRSA
5. Add this graph to the report
6. Go back to the tree

Scenario 5:

1. Select a cluster and show them in the infection route finder.
2. Find sample 1 in the graph
3. Find out which sample is from a family member of a patient

    - *Question: What do you think the big blobs mean?*
4. Add the graph to the report.
    - *Question: What do you expect to be added to the report?*
5. Go back to the tree view

Scenario 6:

1. Remove a graph from the report
2. Complete the report by writing an explanation of your findings
3. Remove a graph from the report
4. Export the report

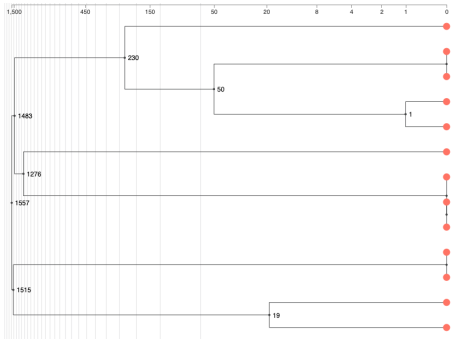
Discussion questions:

- What did you think was good?
- What did you find difficult?
- How would you want to improve it?

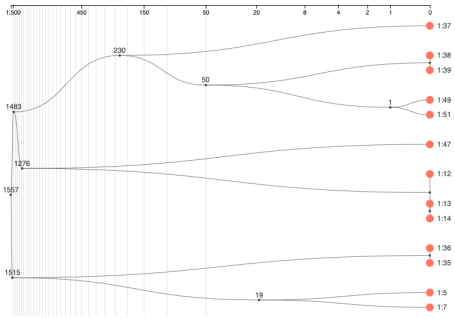
B

Questionnaire Results

Question 1: In which design do you find it the easiest to see the relationships between the samples?

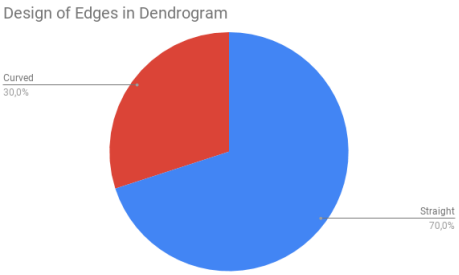


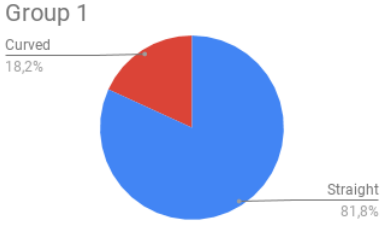
(a) Right angled edges



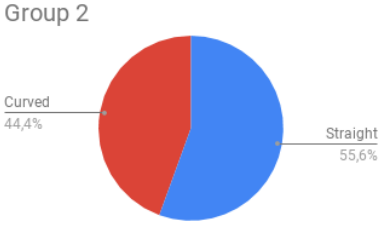
(b) Curved edges

Figure B.1: Two different designs on the edges in the phylogenetic tree.





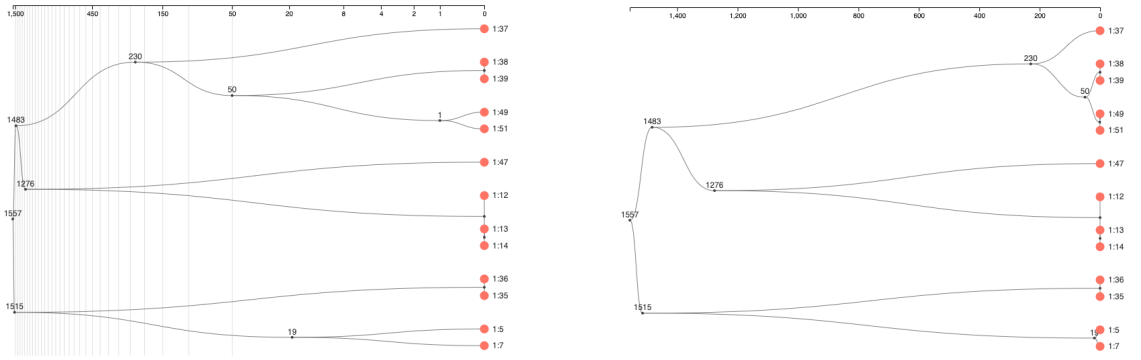
(a) Result for group 1.



(b) Result for group 2,

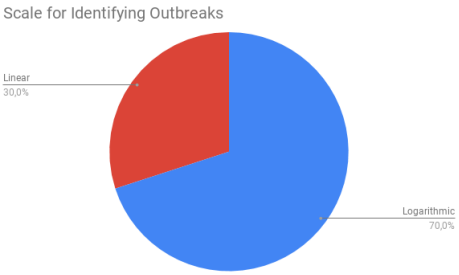
Figure B.2: Results for question 1.

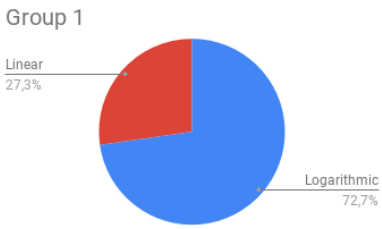
Question 2: Which scale do you prefer in order to quickly get an overview of the clustering?

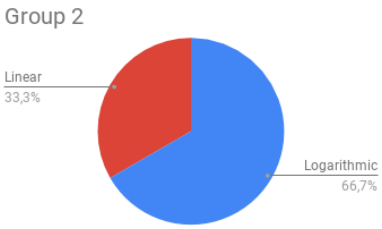


(a) Logarithmic scale. **(b)** Linear scale,

Figure B.3: Two different scales used for placing the inner nodes in the phylogenetic tree.



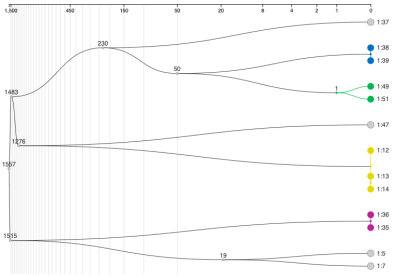




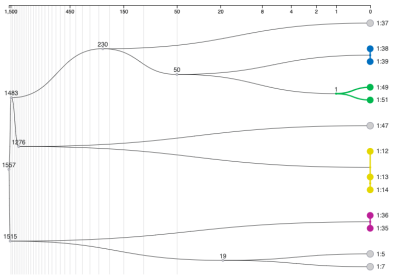
(a) Result for group 1. **(b)** Result for group 2,

Figure B.4: Results for question 2.

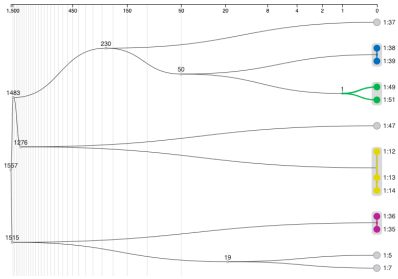
Question 3: In which image did you find it the easiest to identify a cluster?



(a) Coloured clusters with thin coloured lines.

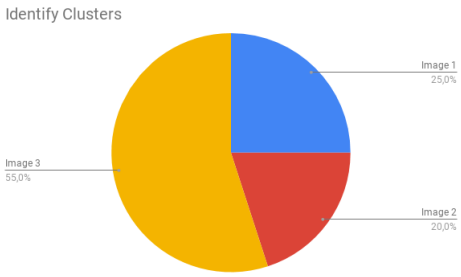


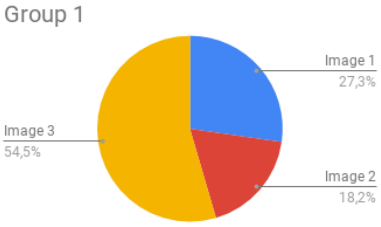
(b) Coloured clusters with thick coloured lines.



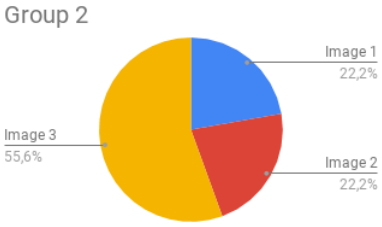
(c) Coloured clusters with grey shadow.

Figure B.5: Three designs for colouring clusters representing outbreaks.





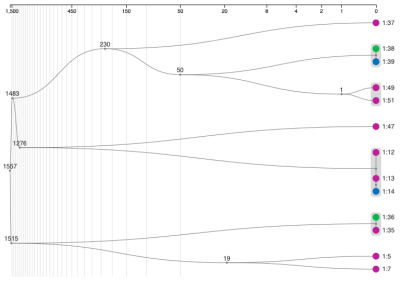
(a) Result for group 1.



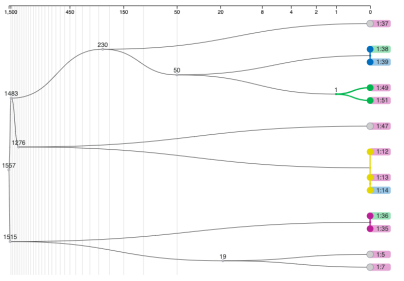
(b) Result for group 2,

Figure B.6: Results for question 3.

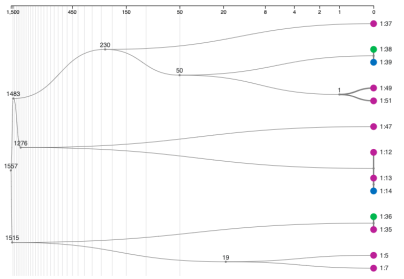
Question 4: In which image do find the colour coding of metadata the clearest, while clusters are easily identified?



(a) Colour coded sample nodes with grey shadow.

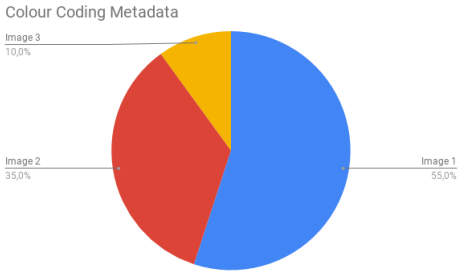


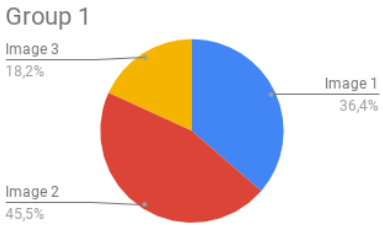
(b) Coloured outbreaks with colour coded background on lab ID.



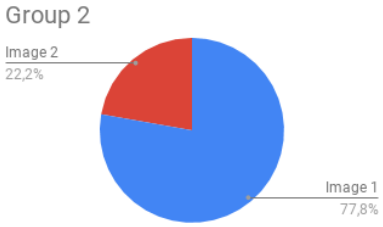
(c) Colour coded sample nodes and thick lines representing outbreaks.

Figure B.7: Three designs for colour coding metadata while preserving outbreak information.





(a) Result for group 1.



(b) Result for group 2,

Figure B.8: Results for question 4.

Question 5: Is the amount of information enough for an overview at the level described above?

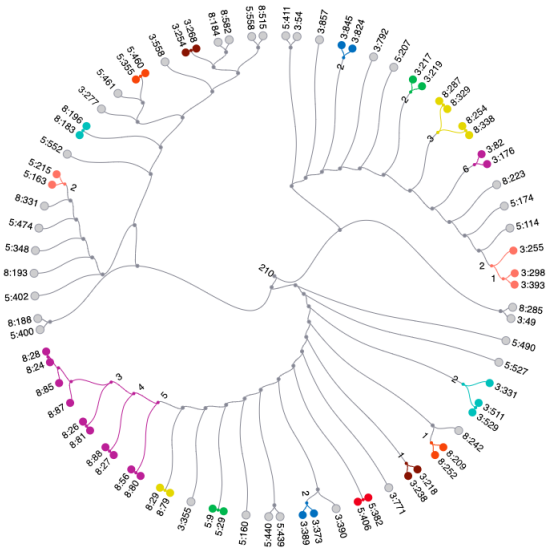
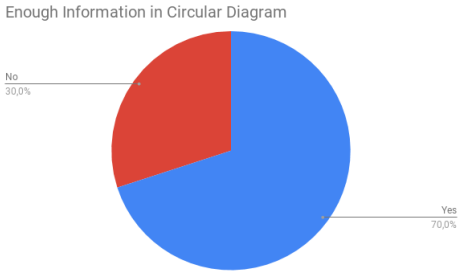
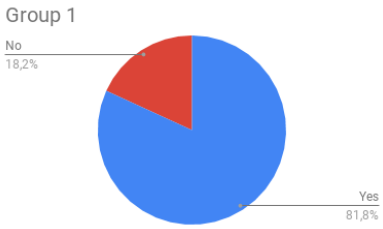
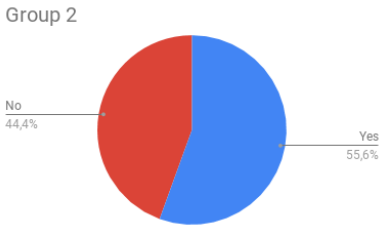


Figure B.9: Circular UPGMA tree with reduced information.





(a) Result for group 1.



(b) Result for group 2,

Figure B.10: Results for question 5.

C

Questionnaire for the Infection Control Unit

Introduction
My project has focused on investigating how an interactive data visualisation could be designed in order to facilitate the creation of reports that are used for infection control within a hospital. The concept that I have developed is an attempt to combine the reporting workflow with the analysis workflow that the microbiologist already performs. The purpose of the test today is to see how well users understand the interactions in the interface.

This is only a prototype, which means that some interactions are not exactly as they would be in a fully implemented product, but pretty close. I will start with showing two images of the interface and ask a few questions about them. Then I will ask you to perform 4 tasks in the interface and lastly ask three follow up questions and think aloud.

Questions for horizontal tree:
What do you think you can interact with?

How do you think you can interact with timeline sliders?
What do you expect to happen when you interact with timeline sliders?

How do you think you can interact with the threshold sliders?
What do you expect to happen when you interact with the threshold sliders?

How do you think you can interact with the inner nodes?
What do you expect to happen when you interact with the inner nodes?

How do you think you can interact with the leaf nodes?
What do you expect to happen when you interact with the leaf nodes?

Questions for MST:
What do you think you can interact with?

What do you expect to happen when you interact with timeline sliders?

What do you expect to happen when you interact with the threshold slider?

How do you think you can interact with the blobs?
What do you expect to happen when you interact with the blobs?

Tasks:
Add a diagram, which includes your latest uploaded samples, with location, date and annotations to the report.
Add a diagram of the minimum spanning tree of your last uploaded samples.
Open the infection route finder with the samples in the outbreak where most of your latest uploaded samples are included.
Complete and export the report.

Follow up questions:
What was your general feeling when interacting with the interface?
Did you find anything confusing?
How would you like to improve that?

D

Questionnaire for the Infection Control Unit

Each section consists of an image of a diagram with an accompanied explanatory text. 2-3 questions were asked about each of the diagrams. One infection control nurse answered the questionnaire and this appendix includes all the images and questions as well as the answers to them. The questions and the answers have been translated from Swedish.

D.1 Section 1

The samples in question are highlighted and are shown together with all staphylococcus aureus samples from the last month. Four of the samples are part of the green outbreak. It is uncertain whether the samples 8:80 and 8:56 are included in the outbreak or not.

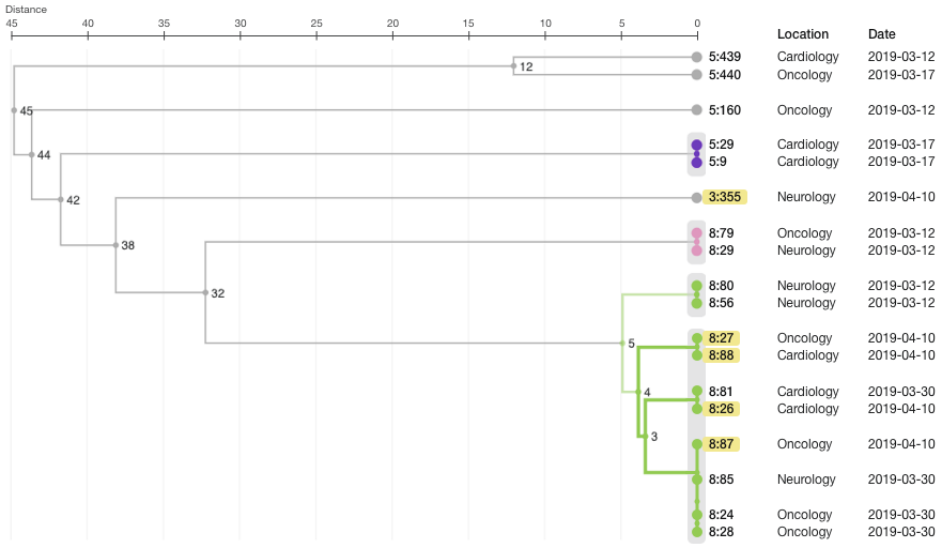


Figure D.1: The image for the questions in section 1.

Question 1: Is it clear to you how your samples are related to each other? Why?
No

Question 2: How useful is the information in your work?
Maybe it is clear if I have a "lab-person" with me that explains?

Question 3: How could the diagram be improved in order to make your work easier?
Don't know. I'm a nurse and not used to interpret this kind of diagrams. I can understand that the green lines mean that the samples are grouped together but not more than that.

D.2 Section 2

The samples in question are highlighted and are shown together with all staphylococcus aureus samples from the last month. It is indicated if they have MRSA.

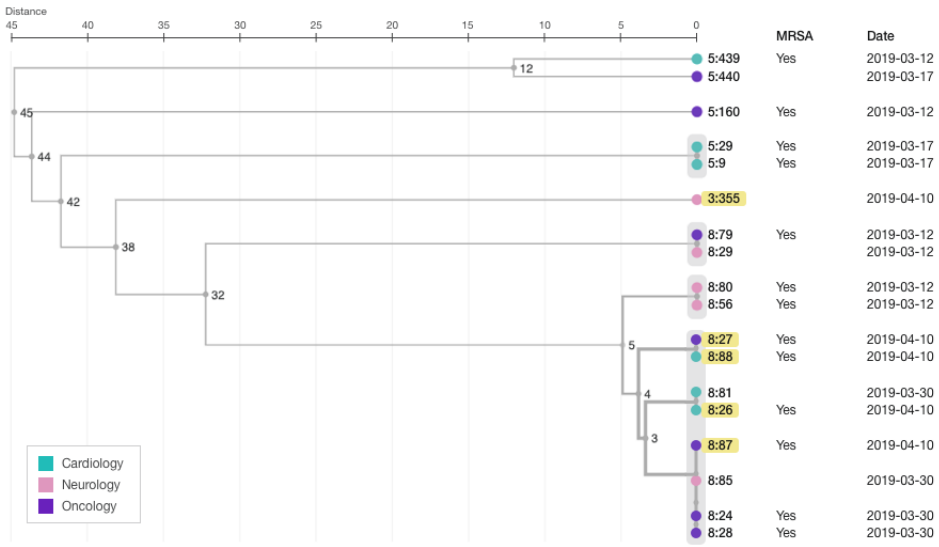


Figure D.2: The image for the questions in section 1.

Question 1: Is it clear what samples are part of an outbreak? Why?
No. I guess that the "staples" show that

Question 2: How useful is the information in your work?
Only if someone shows and explains.

Question 3: How could the diagram be improved in order to make your work easier?
Don't know

D.3 Section 3

The diagram shows the genetic distance between the samples and we can see that the blue samples with high certainty are part of an outbreak. Sample 3:355 are too different from the other to be part f the same outbreak.

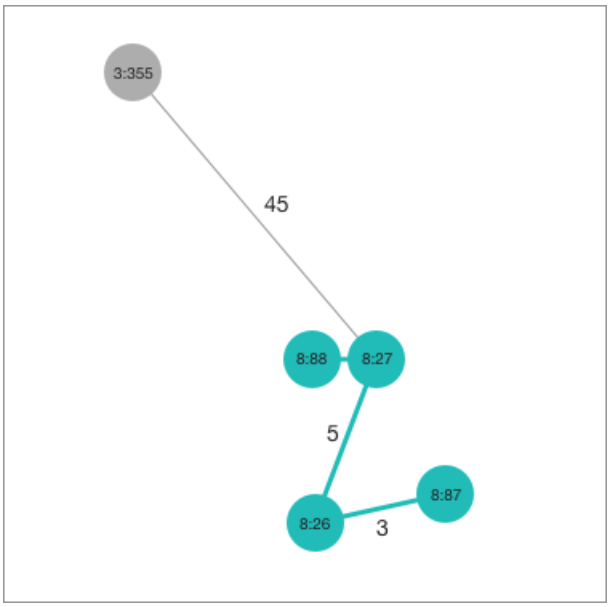


Figure D.3: The image for the questions in section 1.

Question 1: Is it clear to you how the samples are related to each other?
This is more clear. The length of the lines shows it.

Question 2: How useful is the information in your work? Why?
I still need an explanation but I think this is more visual to me.

Question 3: How could the diagram be improved in order to make your work easier?
Don't know

D.4 Section 4

All samples in the outbreak in which some of the samples in question are part of are displayed in this diagram in order to show how the disease has spread over time in the different wards.

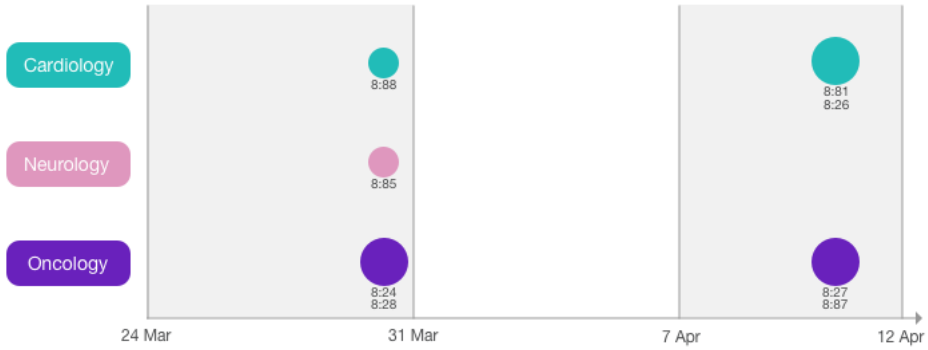


Figure D.4: The image for the questions in section 1.

Question 1: How useful is this representation of when and where the samples were taken? (scale 1-5)
3

Question 3: How could the diagram be improved in order to make your work easier?
A timeline is good.