UNIVERSITY OF BERGEN

Department of Information Science and Media Studies

Master Thesis

Human Computer Interaction Design for Data Mining in Cancer Registries

By: Grace B. Kanza

Supervisor: Ankica Babic

01st Jun. 2014

UNIVERSITY OF BERGEN
Department of Information Science and Media Studies

Master of Information Science

Abstract

Human Computer Interaction Design for Data Mining in Cancer Registries

by Grace B. Kanza

Cancer registries are created, managed and data mined to gain knowledge about long-term outcomes, effects of the medication and clinical factors influencing the patient well-being. Such clinical registries have been a good source of information and are used for research purposes in areas such as epidemiological research and healthcare planning and monitoring for evaluating and planning cancer control programs across the different regions. They contain a lot of data that require good visualisation techniques in order to extract meaning. Interest among different user groups (clinicians, medical students, and the public) creates expectations regarding the results and active role in the development and interactive use of the information.

Following the need for flexible and easy to use user interfaces, this thesis looks at reports based on the data mined from the Swedish Cancer Registry as a starting point in suggesting how this data could be presented to the audience in smart and user-friendly interfaces that enhance the utilization of information kept in such data resources.

Personal Kanban was used to manage the design process, which combined Parallel and Iterative design process models for designing a prototype solution. The resulting prototype provided users with a choice of several data visualisation possibilities following the users' tasks and goals. Heuristic evaluation with experts was used to evaluate elements of the user interfaces. The prototype had high scores on all the design dimensions: visibility of system status; match between the system and the real world; user control and freedom; consistency and standards; error prevention; recognition rather than recall; flexibility and efficiency of use; aesthetic and minimalist design; help users recognize, diagnose, and recover from errors; and help and documentation.

Table of Contents

List of Figures

List of Tables

Acknowledgement

I would first and foremost like to thank my supervisor Dr Ankica Babic. The work in this thesis was made possible by her invaluable support, motivation and guidance that she provided in the course of the thesis journey. I will always be grateful for her guidance role.
I wish to thank Dr Maria Albertsson from the Department of Oncology, Institute of Clinical and Experimental Medicine, University of Linköping, for her support, for pointing out the importance of registry work in the real life and for sharing her rich knowledge about the cancer registries.
I acknowledge the support of other members of The Department of Information Science and Media Studies of the University of Bergen without which my experience with the university would have been less fulfilling.
I wish to thank all who have, in one way or another, helped me during the project, especially the users who participated in the usability testing and all the experts that evaluated my prototype.
Finally, to my family, a special *thank you* to my beloved husband Bjørn, for his patience, understanding, motivation and continuous support during the long race; my parents-in-law, Tor and Luise, for their encouragement; and to my mother for her encouragement.

Chapter 1: Introduction

Long-term studies with a high number of subjects are trusted to give statistically meaningful results and a reliable insight into the patient outcomes. Medical registries have been developed using fewer data items, limiting the number of primary and secondary end points, but they keep the ambition of gathering data from several centres, both nationally and internationally. They allow several periodic analyses and some flexibility in generating reports.

However, for many years researchers and physicians have had to deal with more than one tool to obtain, analyze and visualize data from different medical registries. Reports have been published offline and once the gathered data was verified, data sets were sent away for analysis to the statisticians, which has often been time consuming. Reports resulting from those analyses typically come back in a form of a file, which could be tedious to follow in a data-mining manner. Having to look at never ending rows and columns of numbers is cognitively a demanding task that could be eased by human computer interaction methodology.

A hypothesis could be made to explore how a smart design for visualizing data could benefit mining the data from what is usually called big data (Eaton, Deroos, Deutsch, Lapis, & Zikopoulos, 2012). In this context, the notion of big data stands for the combination of various clinical data, images, and textual information found in these registries; all of which is the reality of the patient management, clinical research and education.

This thesis looks at the Swedish Cancer Registries and the report based on it. It provides several alternative user interfaces in which data mining could be implemented in order to enable efficient and user-friendly data visualisation.

The Cancer registries and web-based medical data sources containing information about cancer related studies, are subjects of change as new data brings new information and holds insights into the patient population, disease treatment, and changing patterns. Despite the challenges of data mining, there are possibilities to employ existing technologies. They could enable integration of distributed systems and assist users to contribute their data and track changes happening with time and new incoming data. This also helps compare already

existing information and analyse selected samples based on diagnosis, age, treatment and quality of life, to mention some of the possibilities.
This thesis has used the Waikato Environment for Knowledge Analysis – WEKA (Hall et al., 2009) project as a starting point for designing user interfaces and as an example of data mining software that provides graphical user that can be used to mine and present data in different computing environments.

1.1 Research Problem and Objectives

This thesis will be part of a new project that will be run by the Swedish Oncology Research Group at the institute of Oncology at the University of Linkoping. The project aims at enhancing data mined from the cancer registries.
The place of my thesis in this project will be to provide smart visualisation designs for data mining from the registries. A smart visualisation solution is not in use in this particular area. For the pilot study, the material of Swedish cancer register regarding the incidences of cancers in different regions of the country has been used.
The Swedish Cancer Registry is a collection of data about the incidences of cancer in the different regions of the country (Ericsson et al., 2011).
Currently, extracting data from the registry is done through sending a request to statisticians who will retrieve the data requested and then forward it to the researchers and physicians for their analysis. This is done with the help of statistical tools such as the SPSS (Coakes & Steed, 2009) and SAS software (Khattree & Naik, 2000). This process is rather primitive and time consuming as it involves several stages from data request to data availability. We are expecting that by giving the researchers and physicians the possibility of direct interaction with the system will speed up the process and make learning and knowledge extraction more efficient and more fruitful as it would be possible to spontaneously explore more data.
Another gain of good interaction design is users' freedom to test hypothesis, ideas and minor tasks that are seldom requested in the current forms of work.
In planning data mining research, the phenomenon of big data has to be mentioned since it brings new challenges for data mining. What seems to be easy for a physician might be more demanding in reality. For example, combining all sorts of data that are available under the umbrella term of 'big data' might prove technologically demanding as it would mean

combining and analysing mixture of quantitative, qualitative and textual image data. In comparison, cancer registries seem to be still a manageable challenge as they collect a high quality data using clinical criteria and follow well defined data collection and data maintenance standards. Research hypothesis of this thesis are the following:

- Data visualisation can support data mining from the cancer registries.
- There are different case studies following the demands of the current cancer practises.
- Patients are also user group to consider for exploration of clinical data.
- Human Computer Interaction design is a good practical framework for designing data visualisation.

1.2 User groups

There was three different user groups identified as potential users that this design targets: clinicians, medical student, and the public.

1.2.1 Clinicians

This is the user group with the widest spectra since they continuously use data mined from the cancer registries for research purposes and improving clinical practices. They do not have direct interaction with the registries, but use statisticians to obtain the data. For instance, for reporting certain cancer incidences at regional level, they would obtain the data after going through a series of steps and only then, can they analyze this data and make the report of particular interests to them. The data is often delivered with some delay and comes in many rows and columns, which makes it difficult to easily extract information. Therefore, by having smart and user-friendly interfaces changes could be made to improve interactions with the registry system. Direct interactions between the clinicians and the data will reduce the steps they are now performing. This will eventually prevent loosing the site over the information and it will provide users with alternative interfaces to view the data.

A good example is when researchers explore the mined data in order to detect trends or make hypothesis, it is important for them to have an overview of all the variables, which they think could be of impact.

1.2.2 Medical students

For students, the resulting prototype opens up learning possibilities as it can be used as an e-learning tool by defining tasks in which the mined data could be used, for instance, connect types of cancers, their treatments and the study outcomes to do research even at the student level. User interface with a good visualization technique will allow a student to explore relationships in the data to combine different representation forms and to understand relationships which otherwise could be hard to find.

1.2.3 The Public

The current system allows the public to view the data from the cancer registries in form of the annual reports. The public gets a limited access due to the nature and sensitivity of the data stored in these clinical data resources. Even though a lot of this data is publicly available, it is still not easy to extract information due to the way the data is presented. The many rows and columns of the data are tedious and not at all encouraging to go through in search for specific information. Therefore, the need to ease the ways of presenting data is of crucial importance. The designed prototype provides possibilities to look at the published reports resulting in several easy to follow interfaces.

The inclusion of other related data from several data resources will assist users in obtaining more information about the particular type of cancer selected. For example, the prototype provides the possibility for users to search for ongoing clinical trials that are related to the selected type of cancer where they can see the purposes of the trials, outcomes, enrolment criteria and the distance to the nearest centre performing clinical trials.

1.3 Research Question

This research will address the following questions;

- How can Human Computer Interaction make interfaces for data visualisation in data mining more efficient and user friendly?
- How can the research of this thesis help transform the current state of presenting data from the cancer registries to a better and preferred state?
- How can smart user interfaces enable good data visualisation for different users groups?
- Will this kind of research and interfaces help to change the way users look at the registry data?

1.4 Structure of the Thesis

This structure of this thesis is as follows:

- *Chapter 1* provides an introduction to the thesis, and presents the research problem and objectives; the target user groups and the research questions.
- *Chapter 2* provides a description of the fundamental concepts underlying the theme of this thesis and that subsequently appeared throughout the thesis chapters when reviewing related works.
- *Chapter 3* describes the design science methodology and the research methods that were followed.
- *Chapter 4* explores the design process by presenting the tasks and workflow; tools and technologies used and the establishment of the requirements.
- *Chapter 5* provides a description of the prototype development process with Personal Kanban and the three iterations as the main components of this chapter.
- *Chapter 6* presents the evaluation results and the discussion.
- *Chapter 7* contains the suggested long and short-term future works, and the conclusion of the thesis.

Chapter 2: Basic Concepts and Related Work

This chapter presents the definitions of the basic concepts that will dominate this thesis and on which the literature survey is based.
Since this research project presents a solution that has not been previously used to visualize data from bigger data sources, the literature search will touch Information System technology and clinical domains such as cancer, organised after registries that are themselves huge data storages organised after sound domain concept. From the Information Technology (IT) point of view, the literature will focus on data mining, data visualisation, user experience, usability, design principles and big data. It will deepen into the area of design and human computer interaction, which are expected to provide approach and solutions to data visualisation and efficient user interaction with cancer registry data. The thesis will also reason about the possibility of individualized, and personalised convenient ways of data mining using small devices such as mobile phones and tablet to display data mining.

2.1 Data Mining

Data mining is sometimes referred to as data or knowledge discover. Shaw, Subramaniam, Tan, and Welge (2001), defines data mining as, "the process of searching and analyzing data in order to find implicit, but potentially useful, information" (p.128).
In support of this, Wilson, Thabane, and Holbrook (2003) use databases as the primary source of data and define data mining "as the application of statistical techniques, e.g. Predictive modelling, clustering, link analysis, deviation detection and disproportionality measures, to databases" (p.128).

Since they complement each other well, the two definitions will be used when referring to data mining in this thesis.

Data mining tools such as the WEKA software (Hall et al., 2009) and the Massive Online Analysis (MOA) software (Bifet, Holmes, Kirkby, & Pfahringer, 2010) facilitate extraction of meaning from different data sets.

Waikato Environment for Knowledge Analysis (WEKA) is a collection of machine learning algorithms useful for extracting information from large databases. It is open source software implemented in Java and is executable on different platforms either by using the command line or Graphical User Interface (GUI). In addition, its algorithms can be called from a customized Java code or can directly be applied to a dataset. The software can provide data in two main ways. The first is by allowing the possibility of loading data from databases, files and Universal Resource Locators (URLs) with the help of supported formats. The second is the possibility to generate data from artificial data sources where the generated data can be edited manually with a dataset editor (Kirkby, 2007).

According to Hall et al. (2009) the WEKA software has currently several user interfaces designed for different data presentations on PCs and large screens , but not much has been done to incorporate data presentation on mobile devices.

The Massive Online Analysis (MOA) is an open source framework for data stream mining that is meant for algorithm implementations and the execution of online learning experiments. It has many similarities to the WEKA software mentioned earlier and supports bi-directional interaction with the WEKA software (Bifet, Holmes, Pfahringer, et al., 2010).

As presented by Bifet, Holmes, Pfahringer, et al. (2010), the software's aspects that are interesting and attractive include the evaluation tools and the machine learning algorithms of clustering, recommender systems, regression and classification.

Both the WEKA and MOA systems give users the possibility to play around with data as they analyze it from different dimensions. These tools are applicable in different areas as they simplify the process of identifying trends, patterns and correlations when working with large amounts of data. This serves as an inspiration for the thesis work, which will also present several design cases in which data mining could be implemented in order to enable efficient and user-friendly data mining from these registries.

With the help of data mining methods and algorithms such as the support vector machines, decision trees and artificial neural networks, the medical field can use data mining in different research and studies to predict survival chances of patients with different types of cancers. For instance the prediction models for breast cancer survivability developed by Delen, Walker,

and Kadam (2005) is among the successful research projects that used such data mining methods and algorithms for the prediction of this particular type of cancer.
In addition to health care, other areas that benefit from data mining include pharmaceuticals, banking, finance, advertising, telecommunication, transportation and aerospace, engineering, insurance, e-commerce, and retail, where data mining can for instance be used in fraud detection, trend analysis, market interactivity, and market segmentation (Persidis, 2000).
This thesis will approach the topic by focusing onto the selected registries and provide easy to learn and easy to use interfaces to the data resources.

2.1.1 Data Mining Tasks and Techniques

As shown in the *Figure 2.1*, Shaw et al. (2001) divide data mining tasks into five categories of which, data visualization takes the main focus in this thesis with the help of graph based techniques that are used to extract and reveal the unseen data relations and trends behind the collected data. When working with such huge data sets as those of the cancer registries, other data mining approaches such as those offered by multivariate statistics and artificial intelligence(Kanza & Babic, 2014) may also be applicable depending on what data is extracted from the data sources for exploration.

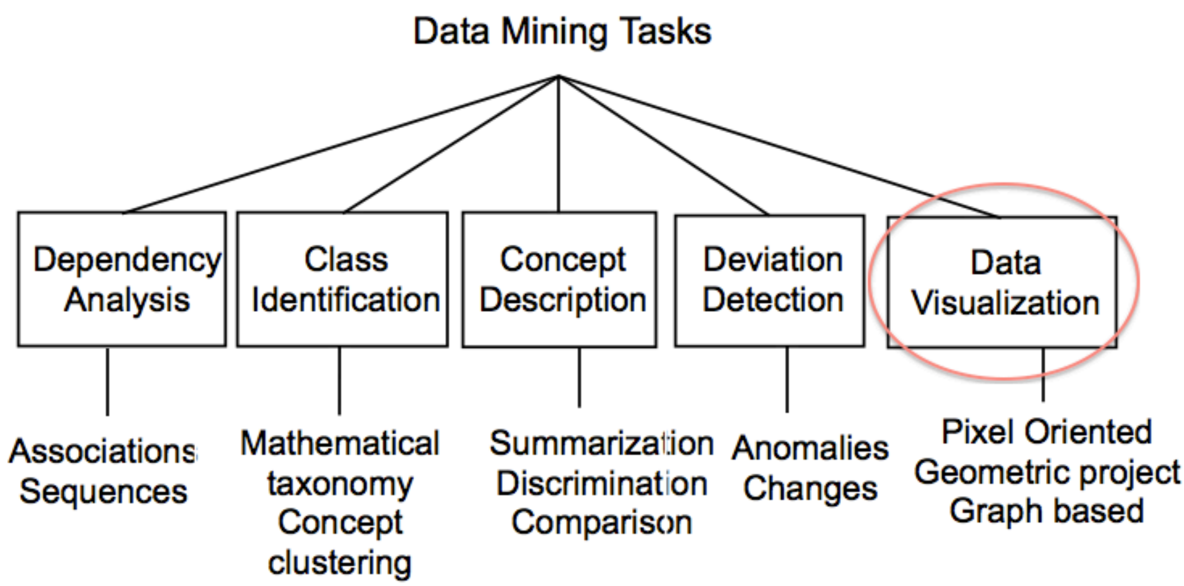


Figure 2. 1: A taxonomy of data mining tasks (Shaw et al., 2001)

2.2 Data Visualisation

Data visualisation is the graphical presentation of abstract information that assists users to reason about, explore, and communicate data. It essentially helps with the processes of thinking and communicating the displayed information (Few, 2013).
The graphical presentation includes the use of bar graphs, charts, diagrams, tables, histograms, and maps.
Under the domain of cognition, the term visualisation is referred to as the construction of visual image in the mind aiming at forming a mental model of the data that is being analysed. Therefore, mental models are humans' internal representations of the external visual world (Ware, 2000).
Graphical presentation relies on the visualisation tools used to accurately convey the information behind the data. Devices used to present such kinds of data are also important since they need to be capable of displaying all the required details in order for users to make sense of them. For example, depending on the type of data one is working on, too small screens might be difficult to interpret data from. Since different screen sizes could be used for presenting mined data from the cancer registries, a balance between the functionality and aesthetic form will be undertaken in order to allow user interpretation.
Marketing is another area that benefit from a well-organised combination of data mining and good data visualisation. With the help of good visualisation techniques, for instance, Shaw et al. (2001) discusses the use of data mining tools for marketing purposes where customer relationship management is substantiated by user-friendly data interactions. Such integrated frameworks as for knowledge discovery and management can help marketers to address customers' needs according to what markers know about their customers rather than following generalised customer characteristics. Even though this thesis does not fall under marketing domain, such a results are inspirational and encouraging to designing solutions that will support clinicians in their knowledge discovery and interactions.
In the pharmaceutical field, data mining, knowledge discovery and data visualisation have played a big role in, for instance, the detection of Adverse Drug Events (ADE) through working on the data mining techniques for sorting the data and finding how the different data in is related (Wilson et al., 2003).

2.2.1 Why Graph - based Visualization?

Of the three visualization techniques mentioned in the *Figure 2.1*, above, graph based techniques (Cook & Holder, 2006) visualizing data seems to have a broad use in the data intensive area in which combination of different kinds of data would be an impossible task. Köhler et al. (2006) have presented a challenge of interpreting genes-related knowledge and solutions offered by the graph theory. Many similar references could be found to confirm their findings that are of interest for this research. It was an expectation that the graph based techniques would enhance designing interfaces for visualizing data from the cancer registries.

2.2.2 Human Interpretation Issues

In order to visualise data effectively, design principles that are derived from an understanding of human perception should be followed when designing graphical displays that will cater for huge amounts of data. This will simplify the translation of the abstract data into physical attributes (Few, 2013).

This thesis will also consider issues that are related to human interpretation of visual data representation as a part of data visualisation tool design. The research has a mission to design for groups such as clinicians, medical students and the public.

The interfaces will be designed to fit the clinicians with the most relevant attributes paying attention to the specifics of the medical field.

This is a requirement that needs to be met in order to secure understanding and accurate interpretation of the clinical data. User groups may have different scientific and professional skills but they have interest to share the same data set from their point of view. In all cases it is important to make sure that the data is presented clearly and in ways that are meaningful and intuitive to the users, which will have a significant impact on the design.

2.3 Big Data

Big data is the term used to describe a collection of large quantities of heterogeneous and complex data sets. The terminology describes the exponential growth, variety and speed of information that makes up this kind of data. Some characteristics of this kind of data include the tendency to be difficult to be deleted and be highly complex (Eaton et al., 2012).
Data in the Swedish cancer registries is large; it is entered continuously and is of a well-defined structure. To fit the needs of the oncologists, pathologists health carers and sometimes patients when quality of life is of interest. The data comes from the different regions and their regional cancer centres where the quality of data is checked before it enters the national cancer registries (Ericsson et al., 2011).
These registries have been a good source of information and are used for research purposes in areas such as epidemiological research and healthcare planning and monitoring. For example they have been used to study trends and patterns of the disease over time in various populations. It is, for instance, possible to see the different types of cancers and their occurrences in the different regions of Sweden, something that can possibly help researchers to interpret cancer incidences. Highly populated areas where pollution is also high could for instance have a higher incidence of a certain type of cancer more popular than what could be seen otherwise.
The registries can be used as a guide for evaluating and planning cancer control programs across the different regions. They can for instance, be used to determine whether screening and treatment efforts should be organised for preventive reasons.
Data in these registries has as been used to improve the treatments for cancer patients and determine how to tackle new medical conditions, aging population, and different treatment strategies. All this is done by studying data collected over the years and has been reported in annual publicly available reports.

2.3.1 Challenges with Big Data

Challenges associated with big data generally vary from one field to another (Manyika et al., 2011). They are not only caused by the size of the data as big data can also be small, but also

caused by other factors such as; the type of data and its structure, the quality and reliability of the data and the technologies and tools used to store, manage, analyse and visualise the data. According to (Kalil et al., 2012), when not addressed properly, storage, harness, search, visualisation, analysis and sharing are some of the challenges associated with big data.
Some of the factors that could be considered to be a challenge regarding the use and handling of big data may include

- The data might be too large and too difficult to analyse.
- It can be difficult to find out how much of the data should be analysed.
- It can be tricky to find out which data points are really important.
- There is a possibility that collected data is biased or noisy.
- Authentication of the data can be a security problem that may require extra intelligence to tackle.
- Spending too much time on working on irrelevant data sets.

Annual reports, as they are useful, they still pose a problem since they can be approached best from the statistical point of view. They require concentration and memory and some knowledge of statistics to interpret the data correctly.

2.4 Human Computer Interaction (HCI)

According to (Dix, 2009), Human Computer Interaction (HCI) is a huge and significant field that deals with the ways in which humans interact with computers and how their activities are being influenced by computer technology.
In this thesis, it was attention to focus on the interactive information visualisation. It was intended to find best ways to visualise and interact with the cancer registry data that is available to the public.
The prototype's graphical interfaces should be used to enable direct interaction with the data and some degrees of personalisation to engage human cognition and help interpreting this kind of data as meaningful patterns.
A pragmatic approach to the visualisation and interaction would be to use graphical tools (Heer, Card, & Landay, 2005) with the capacity to present information through graphs, trees, body parts and/or regions.

The framework to be employed by this research thesis would be, as shown in the *Figure 2.2* where a user would send a search request (stage 1 on the diagram) to the specialised registries, then the data would get filtered (stage 2) before it gets transformed (stage 3) to be viewed. In this way, users can use the different view modes to personalise their search as they interact with the data directly.

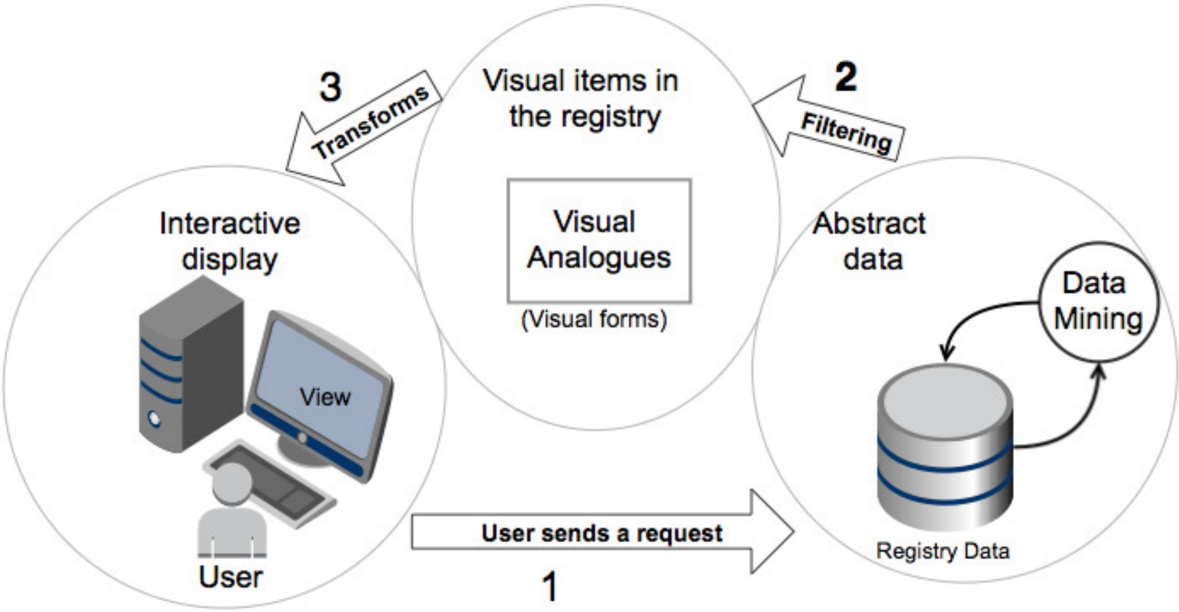


Figure 2. 2: Visualisation framework of the prototype concept

2.5 User Experience and Usability

According to Rogers, Sharp, and Preece (2011), user experience is a notion that engulfs the total understanding of user's impressions of a product, its behaviour and usage. Since it is not possible to design "the user experience" itself, this implies that user interface designers design for a user experience by creating design features that evoke a sensual experience as a user uses a product.

With this in mind, designing for a user experience is a continuous process that does not end when one single function is completed. Instead, it challenges the designers to evolve the product's design alongside its users' evolving behaviours, limitations and needs.

Usability is another key concept of user experience that has been considered when designing the interactive user interfaces for visualizing data from the cancer registries.

While user experience focuses on the impressions users get as they use a product, usability emphasises the criteria for improving the product's effectivity, ease-of-use and simplicity to learn. In addition, usability improves interaction between people and the interactive products and simplifies everyday life (Jacob Nielsen & Molich, 1990).

As applied and discussed later in the Chapter Five, the usability goals according to Rogers et al. (2011), include;

Effectiveness: Is the designed product successful in fulfilling the tasks it is supposed to do?

Efficiency: Does the designed product support high productivity for trained users? How quickly can they perform their tasks?

Safety: Does the designed product prevent the user from dangerous and unwanted situations?

Utility: Does the designed product provide the appropriate functionalities that are required for the user to be able to perform the set tasks?

Learnability: How easy is it for users to learn how to use or operate the designed product after the initial instructions?

Memorability: How easy is it to remember how to use the designed product once learned? How easy is it for users to establish proficiency after a period of not using the product?

2.6 Design Principles

Interaction designers use a set of design principles in order to design interactive user interfaces for user experience. When applied appropriately, these principles are meant to improve the quality of the user interface designed by providing some guidelines that helps the designer's thinking during the designing process (Rogers et al., 2011).

In this thesis the design principles according to Rogers et al. (2011) were followed as a guidance towards the design process. These included:

Visibility: Functions should be visible for users to easily find and use them. Hidden functions are difficult to locate and therefore users are likely to not know how to use them.

Feedback: The user should be informed by the system on whether the performed action has been executed successfully or not.

Constraints: The system should restrict users from performing invalid commands by proving command options suitable for the context at a given time.

Consistency: The interface should maintain the flow of similar elements and operations throughout the system.
Affordance: The attributes of the design elements used should be self explanatory with obvious clues that intuitively allow users to know how to interact with.

2.7 User Interfaces

This thesis utilizes the Information Technology that is close to the users' proximity and understanding in order to design user interfaces, which could improve the way data mined from the cancer registries could be presented to its audience. The resulting prototype from the design process presents several user interfaces that allow the different user groups to choose a way through which information could easily be obtained and extracted from the data they visualise.

This research is intended for the PC screens, which will allow quite complex presentations and combinations of data. In the future one should also consider small devices, as they are present in different working environment and in personal life. For example Medical doctors have the possibility to have a quick access to the internal medical journals from tablets in the middle of a consultation session.

It is possible to perform lots of mining tasks using even smaller devices as well as bigger like the PC and these two will influence the dynamic of designing screens which will probably become next step to consider. It will be demanding to understand what amount of data is still readable and presentable but the need to be quick and connected to the meaningful data resource such as registries is likely to become user requirement.

These design challenges can be tacked through applying what is usually called responsive design technology (Tidwell, 2010) (Gardner, 2011) that provides control of the computer generated layouts responsible for the detailed adjustments of the design parameters of the device layout attributes. Responsive design techniques (Gardner, 2011) enable user interfaces to respond to user actions without interruptions from page reloading.

2.8 Design Case Studies

According to (Walliman, 2006) and (Bryman, 2012), a study case is a comprehensive description and analysis of a single instance of a particular situation. Study cases are often used to narrow down broad fields in order to explore thoroughly all the possible details that can otherwise be missed.

In the context of this thesis, case studies were used to well define typical situations that are experienced by the current practises for visualising mined data from the cancer registries and that make the case for the design (Kanza & Babic, 2014). The cases used followed real life situations and covered most of the common tasks seen as they are conducted today.

In this research thesis, this paradigm was employed twice. Firstly, to understand the problems through the use of research based personas (Goodwin, 2011), which are described in the *Section 4.3.1*. Secondly, to demonstrate and compare the current practises to the newly suggested ways of solving similar tasks through the design case studies, which are explored in the *Section 6.2.1*.

2.9 Chapter Summary

This chapter has exhausted the fundamental concepts that support the basis of this thesis. It has provided some examples of inspiring projects that have been successful from area within and outside the clinical field. The most commonly used concepts and terms within this thesis have also been introduced.

Chapter 3: Methods and Methodologies

This chapter provides a description of the methodologies and scientific methods, which this thesis is based on. Its constituents are design science methodology; and genius design, Question, Option and Criteria (QOC) design method, prototyping, usability testing and expert evaluation methods. Data mining methodology will be touched in brief mainly to define their value and user experiences with it.

3.1 Design Science

Design science and behavioural science are the two paradigms that characterize Information Systems' research. They provide frameworks for analysis, investigation, evaluation and iteration guidelines for research projects. Design science is a problem-solving paradigm that deals with the creation of new and innovative artefacts in order to extend the boundaries of human and organisational capabilities. (Hevner, T.March, Park, & Ram, 2004). Alternative to consider, Behavioural science deals with the development and verification of theories explaining organisational and human phenomena within the discipline of Information systems (Hevner et al., 2004).
Of the two paradigms, this research follows Design science methodology and uses its seven design guidelines during the design and evaluation processes. The designed application addresses the problem regarding efficient data extraction and accurate visual representation of data from the cancer registries by providing different user interfaces that will assist users to visualise data efficiently. This is among the aspects of this thesis that differentiates it from the routine design. The evaluation of the end product will be based on its design and ability to perform the expected tasks.

3.1.1 Design Science Guidelines

Hevner et al. (2004), recommend researchers within the field of Information systems to use the design science guidelines to broaden their understanding about effective design science research and to use logical reasoning and be creative when applying every one of these guidelines. Below are the stated guidelines and a detailed presentation of how the design science guidelines have been applied to this thesis project. These design science guidelines are listed and described in the following text: Design as an artefact; Problem Relevance; Design Evaluation; Research Contributions; Research Rigor; Design as a Search Process; and Communication of Research.

Guideline 1: Design as an artefact

"Design-science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation" (Hevner et al., 2004, p. 83).

On this design guideline, Hevner et al. (2004) emphasise more on the artefact (March & Smith, 1995) itself than the organisations or users using it. In this way, the design process and the designed artefact are addressed in different design aspects. With this in mind, researchers are able to have a usable and identifiable product designed to evaluate in order to complete a research.

A prototype of a system that simplifies extraction of information and enhances accurate visual representations of data from cancer registries is the artefact produced in this research. The process of designing this prototype, the technologies involved and its adaptations to the challenging clinical environments will be discussed in the Chapters Four and Five. The user interfaces of the prototype addresses a demanding problem of handling bulk data to allow users to make sense of the big data more efficiently. Conceptual models showing data flows and data relations are also provided.

Guideline 2: Problem Relevance

"The objective of design-science research is to develop technology-based solutions to important and relevant business problems" (Hevner et al., 2004, p. 83).

In this context, Hevner et al. (2004) refer problems to the "differences between the goal state and the current state of a system" (p. 85).

The current system involves several stages and resources in order to perform simple tasks such as retrieving requested data sets from the cancer registries. The chain of commands, in this case, involves a series of stand-alone tasks such as sending a request to a statistician and waiting for a reply, which may take days or weeks before a reply is provided. This kind of problem affects the efficiency of those trying to carry out important tasks that require quick responses. This may delay overall tasks, which are dependent on such requested tasks and completing procedures.

The designed artefact in this thesis aims at trimming down the processes that researchers, physicians, medical students, the public and other users have to go through in order to obtain information from the registries. Direct interaction with the system will save time; and improve the quality of information extracted from these data sources, as information will be obtained directly, conveniently and more efficiently. By allowing users to be more involved, it will allow active online user choices.

This kind of solution in this domain is possible through the use of, for instance, appropriate techniques and methods such as data mining techniques and machine learning methods that can facilitate efficient data extraction and discovery of hidden knowledge within the medical databases.

Guideline 3: Design Evaluation

"The utility, quality and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods" (Hevner et al., 2004, p. 85). In this guideline, the authors emphasise evaluation as a crucial component of the evaluation process and therefore, encourage researchers to select evaluation methods that correspond to the designed artefact. The experimental and descriptive methods of evaluation (Hevner et al., 2004) are used in this thesis to evaluate the designed user interfaces. The detailed description of the evaluation process follows in the Chapter Six. These evaluation methodologies are chosen to fit the nature of the designed artefact and the geographical location of the target users. The evaluation criteria takes into account factors such as; utility, completeness of the tool, accuracy, design consistency, functionality, performance and usability.

Guideline 4: Research Contributions

"Effective design-science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and /or design methodologies" (Hevner et al., 2004,

p. 87). This guideline focuses on what the research can contribute to the already existing knowledge base.

With this thesis, the artefact itself is the contribution to the knowledge base as it provides a solution to the existing problems. As to the regional oncology research centre, this thesis will offer user-friendly solutions to the processes of presenting data from offline registries and provide several visualisation possibilities for the data mined from the cancer registries, something that is currently not offered.

This was achieved through applying a combination of already existing technologies and knowledge from other domains. As an example, this research will hopefully inspire other areas using medical registries, and that are still functioning in the typical way. Designed solutions offered in the thesis could make other users embrace this way of interacting with the data sources.

Guideline 5: Research Rigor

"Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact" (Hevner et al., 2004, p. 87). Here, the authors focus on the way the whole research is conducted, from the construction of the artefact to its evaluation, with emphasis on the methods applied.

The prototype presented in this thesis has been developed iteratively and incrementally with a total of three iterations as described in the Chapter Five. The iterations provided room for adjusting some elements of the prototype that did not fit well. The adjustments were based on the feedback provided from the usability testing, which was carried out in the second iteration.

Personal Kanban (Benson & Barry, 2011), a variation of Kanban, was used for the management of the design process. Even though this thesis was carried out solo, it was important to use this Agile development (Dingsøyr, Dybå, & Moe, 2010) approach in order to systematically keep track of all the design tasks and hold the whole thesis project within the set time frame.

As presented later in the chapter, Genius/ Innovative design (Saffer, 2010) and the Question, Option and Criteria (QOC) design method (MacLean, Young, & Moran, 1989) were also among methods used in designing the prototype.

Guideline 6: Design as a Search Process

"The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment" (Hevner et al., 2004, p. 88).
In this research, the iterations carried out during the design processes represented the *search.* The solution to the research questions, which was the goal of the thesis together with the constraints that came along represent the *ends*. The tools and other resources used to design the prototype represent the *means;* and the controllable environments represent the *law*.
This prototype was developed iteratively and incrementally aiding continuous improvements that evolved to reaching some sort of desired ends.
When designing an artefact, no satisfactory solution is perfect as the designed artefact keeps changing in accordance with the evolvement of the means, ends and the laws within the domain. Therefore, the artefact resulting from this thesis is a starting point that solves only the current problem.

Guideline 7: Communication of Research

"Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences (Hevner et al., 2004, p. 90).
In this thesis, the research documentation is written in such a way that it is presentable for both the technology-oriented and management-oriented audiences from both within and outside the domain. A thorough description of the processes involved in the construction and evaluation of the artefact are included in this thesis documentation in order to cater to the technology-oriented audiences.
On the other hand, in order to convince the management-oriented audiences that this project is worth developing further, emphasis is put on the importance of this project to the cancer research centre; how effective the designed artefact is and what can be achieved when it is implemented in relation to the current situation.
Technology-oriented audiences for a thesis project like this may include information systems developers, computer science engineers, researchers and system analysts whereas Management-oriented audiences may include project leaders, project coordinators, sponsors and the governmental institutions.

3.2 QOC Method

The Question, Option and Criteria (QOC) design method is an approach proposed by MacLean et al. (1989) for representing the design space around an artefact through identifying design problems and alternative solutions to these problems. The constituents of the abbreviation QOC are:

- *Questions* identifying the key design issues,
- *Options* providing possible answers to the Questions, and
- *Criteria* for assessing and comparing the Options.

In this context, design space is made up of decision space, which are the possible appropriate alternative options; and an evaluation space, which are the explicit reasons for selecting an option from the provided alternatives.

According to MacLean et al. (1989), when the QOC for a particular design problem are defined, the question is answered by matching the criteria set against the options to find out the most appropriate option to answer the question. The links between the options and the criteria define the assessments. The option that represents the answer to the question is considered to have a positive link and negative links are the remaining options with less logical answer to the question. Therefore, the designer should be able to solve these cognitive tasks by using the QOC method when designing the artefact.

In this thesis, the method has been especially useful in exploring the focused parts of the design space. This was when making decisions on how, where and what design elements should be presented within the different interfaces. Therefore, emphasis was not put on producing full descriptions of the design spaces, of the different prototypes produced during the iterations. This is an adaptation of the QOC method used to define the design space.

This method helped in providing rational explanations as to why the different design choices were made. There were a lot of questions, doubts and options around this, which was a challenge when it came to deciding the aesthetics of the interfaces. As this was a novel design and there were no existing design forms to follow, this design alteration opted. The formulation of questions for the QOC method is discussed in the Chapter Five.

3.3 Genius or Innovative Design

Interaction design has already made artefacts around us fun to use, useful and usable even though there are still thousands of interaction design problems we encounter. Due to the advancements in computer technology and an increase in the digitalisation of our workplaces, homes, gadgets and communication devices, interaction designers continue to innovate and refine their systems, designs and ideas in order to meet design challenges that come with these complex technologies.

To provide innovative solution to improve data visualisation from the cancer registries, this thesis project has opt for Genius design approach (Saffer, 2010) also known as innovative design for the very initial iteration. This is one of the four approaches to interaction design presented by Saffer (2010); where users will not be directly involved in the very initial design process, but instead participate as evaluators in the later iterations of the first design sketches. Rogers et al. (2011) suggests that even though users often can not tell designers what they want, they will know what they do not want after seeing and trying out a product. Therefore, presenting the users with the concept that will eventually solve the current problem has given users views on what they really want.

Research based personas (Goodwin, 2011) and scenarios (Rogers et al., 2011) were therefore constructed in order to establish the initial system requirements.

Innovative design was chosen specifically for the initial design stages because in this particular case, it was easier to find out what the users want through presenting them with a tentative solution to compare with what they currently have rather than directly involving them in the very early stages. The speed of technology changes is another reason for choosing innovative design since users often tend not to know what they want and it is difficult for them to envision and predict the possible solutions that can be achieved from using the right technologies.

3.4 Prototyping

Rogers et al. (2011) defines a prototype as, “a limited representation of a design that allows users to interact with it and explore its suitability” (p. 530). In software development, prototypes can vary from simple hand drawn sketches to fully functioning pieces of software.

According to (Rogers et al., 2011), prototyping phases can lead into the development of low-fidelity or high-fidelity prototypes depending on the medium used and how close the prototype looks and feels like the final product. While low-fidelity prototypes do not look very much like the final product they are simple and quick to develop. High-fidelity prototypes resemble the final product, are conveyed in the same medium as the final product, and take long time to develop.

When prototyping, compromises are always made in order to produce something quickly that can be used to test the aspects of the product. These tradeoffs involve how much functions should be provided for the developed prototype versus how detailed the functions provided should be presented on the prototype.

This thesis has portrayed the design ideas for visual presentation of the data mined from the cancer registries through the development of several prototypes in several iterations. Due to schedule and time constraints, limited and distributed resources, and the nature of the complexity of the cancer registries and medical data sources as a whole, the final prototype will not be fully implemented in the specs of this thesis. Nevertheless, emphasis was put on suggesting features that are technically feasible for the development team to implement.

The initial iteration consisted of rough sketches of the solution concept on paper, which kept on evolving into rich electronic interactive vertical prototypes by the end of the last iteration. The first design iteration resulted into both hand drawn sketches and electronic sketches as shown in the *Appendix 1*. The second iteration produced the first interactive prototype and the third iteration produced a fine tuned prototype with more consistent flow of design elements than in the first two iterations. The detailed description of the evolution of the prototypes is provided in the Section 5.4.3.

3.5 Evaluation

Evaluation of any designed product is important because it helps to check the quality of the product, if the product is usable, if the intended users like the product and if using the product give users a good experience. It is also useful for providing feedback before the product is put into use or for sale. There are several evaluation methods that fall under the three main approaches: usability testing, field studies and analytical evaluation (Rogers et al., 2011).

The aim of evaluating the user interfaces was to see whether the designed user interfaces will enable efficient and user-friendly knowledge extraction from data through the way this data is presented to users and therefore test the prototype's utility, quality and efficacy.
According to Pickard and Childs (2007), it is not unusual for a system to be evaluated using more than one evaluation method. This depends on the stage of the development process in which the evaluation takes place and it also depends on what design aspects should be evaluated.
In order to get external input about the design features and do some adjustments before finalising the prototype, usability testing was applied as the prototype was being developed. Expert evaluation was performed on the final prototype. The details of the usability testing are explored in the Section 5.3.3 and the evaluation of the final prototype in the Section 5.4.3.

3.5.1 Usability Testing

According to Rogers et al. (2011), usability testing is an approach used to evaluate how usable the designed product is. This is done through users performing a set of tasks in a controlled environment. After interacting with the designed product through performing the tasks, users are asked to complete a user satisfaction questionnaire in order to rate the designed product and point out the usability problems. Rogers et al. (2011) recommend five to twelve users as an acceptable number of testers on a usability study.
In this thesis, usability testing aimed at evaluating the ease of use, learnability, effectiveness, memorability and satisfaction aspects. Since users and designers do not think alike, it was important to get feedback based on others and not just on the designer's point of view. The feedback obtained from the users, on all the three design suggestions, in the second iteration was used to fine-tune the design aspects of the prototype during the third iteration. The details of the whole usability testing are presented in the Section 5.3.3.

3.5.2 Expert Evaluation

In this thesis project, analytical evaluation of the final prototype was carried out through expert evaluation where an inspection method of heuristic evaluation was used. A set of heuristics developed by Jacob Nielsen and Molich (1990) were followed as a guideline for analysing the interfaces and identifying the usability problems of the final prototype. These ten heuristics were used:

- H1: Visibility of system status.
- H2: Match between system and the real world.
- H3: User control and freedom.
- H4: Consistency and standards.
- H5: Error prevention.
- H6: Recognition rather than recall.
- H7: Flexibility and efficiency of use.
- H8: Aesthetic and minimalist design.
- H9: Help users recognize, diagnose, and recover from errors.
- H10: Help and documentation.

This thesis project followed the heuristic evaluation steps suggested by Rogers et al. (2011), which involved: briefing, evaluation and debriefing sessions. The evaluation results were then analysed in order to extract some useful findings that could answer the research questions.

3.6 Chapter Summary

This chapter has presented the design science methodology and described how the design science guidelines were applied on this thesis. It also provided the research methods used in this thesis project and reasons for choosing them.

Chapter 4: Design Process

The process of interaction design is a practical and creative activity, which results into a product that aims at assisting users in achieving their goals. The four basic activities for interaction design according to Rogers et al. (2011), are:

- Identifying needs and establishing requirements
- Developing alternative designs to meet those requirements
- Building interactive versions of the designs so that they can be communicated and assessed.
- Evaluating what has been built and the user experience it offers.

The design process in this thesis resulted in an innovative interactive artefact, which was later evaluated by experts. Prior to the design process, it was important to gain a thorough understanding of how the current system works, who the users are and the challenges encountered when handling data from the cancer registries. Therefore as part of the design process, this thesis has also consider workflow, tasks, procedural steps and organisations and people involved in managing and analysing cancer registries. Resulting patterns such as resource, data, and interaction patterns will be assessed for their potentials to influence the design process.

Since Innovative design approach (Saffer, 2010) was applied, research based personas (Goodwin, 2011) were used in order to obtain the initial user requirements.

Alternative conceptual models describing what the product does and its behaviour and physical models showing the appearances of the product were considered throughout the design process in order to secure the best possible solution.

Several versions of the prototype were developed during the iterations. Due to the nature of the domain problem, the very first versions of sketches were not interactive. The functionalities were improved in the later iterations where interactive features were embedded

from the second iteration. Usability testing was performed during the second iteration once the prototype was made interactive.
Experts carried out evaluation of the final prototype in order to assess the aspects of prototype's utility, quality and efficacy and the user experience it offers. These aspects were measured in accordance to the heuristic evaluation method.
In this thesis, activities of the design process were merged into several sections of the Chapters Four, Five and Six. This chapter explores how user needs were identified and requirements established, in addition to the workflow, and tools and development technologies.

4.1 Tasks and Workflow

Prior to the design process, the main goal was to understand the clinical data and indication for analysis, as well as to understand clinical and other related work practices. This was useful in obtaining a detailed understanding and orientation of the nature of the current practises, constraints, the ways users interact to obtain data, and the kind of data being handled.
Resulting patterns such as resource, data, and interaction patterns were assessed for their potentials to influence the design process. A set of design features (Shaw et al., 2001), were identified after exploring how data mining is conducted currently.
In addition to the analysis of the data and documents, this thesis research has been focused on the interaction design with emphasis on the objectives, elements, services and workflow; and interaction design with emphasis on functions, user behaviours and workflow as shown in the *Figure 4.1*.
All these series of events lead into the new automatic, user-friendly artefact for data visualisation from the cancer registry, which was then evaluated by experts within this specific domain.

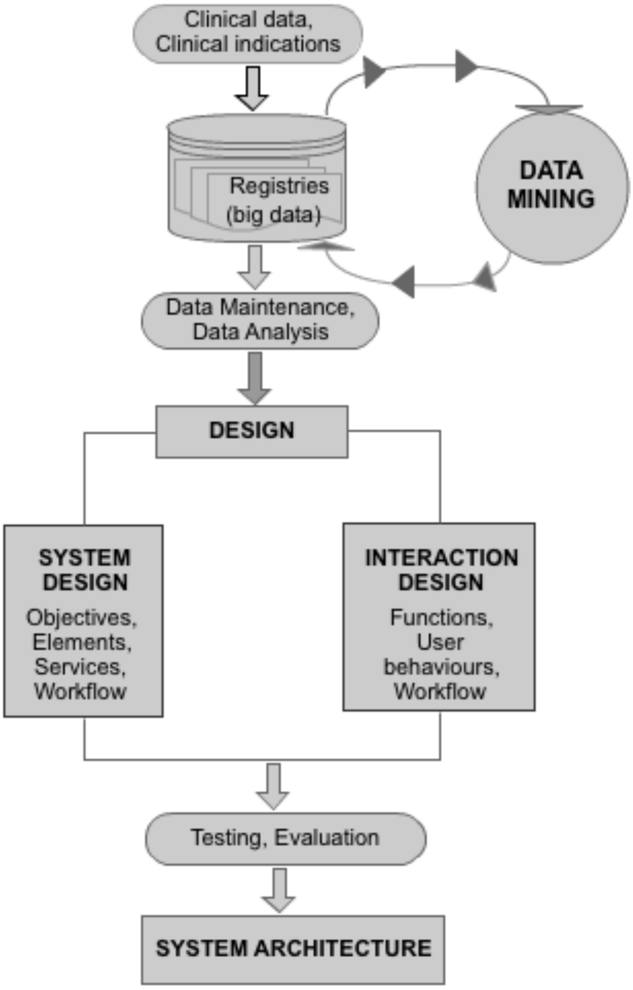


Figure 4. 1: Design steps in data mining (Kanza & Babic, 2014)

As illustrated in the *Figure 4.1* the main research activities of this thesis project included;

- Understanding the clinical nature of the data and indication for analysis.
- Understanding clinical and related work practices.
- Designing a new automatic, user-friendly system for data mining from the cancer registries.

4.2 Tools and Technology Used

In order to produce a viable artefact that could be evaluated, an interactive vertical prototype was developed. The actual tools and technologies used to develop the prototype included Komodo Edit (Apers & Paterson, 2010), Balsamiq Mockups (Guilizzoni, 2010), Justinmind

Prototyper (Justinmind, 2014), Photoshop (Album, 2004), and KanbanFlow (KanbanFlow, 2014). Below follows a brief explanation of how, when and why these tools and technologies were used for designing this prototype.

4.2.1 Balsamiq Mockups

Balsamiq mockups is a wireframing tool that was used to create the first electronic sketches of the prototype. It offers some level of interactivity, but it has limitation since it provides not enough support to fully implement the prototype. On the positive side, the software was easy to use. It allowed sketching the screens quickly and with that it allowed the sketching process to develop and result in what became the most suitable design solution.

4.2.2 Komodo Edit

This is open source web editor software that supports multi-programming languages and simplifies writing quality codes. In this thesis it was used for creating and testing the Hyper Text Markup Language (HTML) (Freeman & Freeman, 2006) and the Cascading Style Sheets (CSS) (Freeman & Freeman, 2006) codes, which were later used on the prototyping tool as snippets. Customised snippets were used to improve the features of the prototype since the main prototyping tool lacked useful functionalities necessary to create features on the prototype for demonstrating this design concept.

4.2.3 Justinmind Prototyper

Justinmind Prototyper was the main tool used for developing the interactive prototypes. It is an authoring tool for creating high fidelity prototypes for software projects. Among the attractive functions and possibilities that the software offers are: a drag and drop functionality and a server where the project can be stored and deployed for review. These possibilities were especially useful for communicating the prototype with the experts that were located in another country. It also allows the inclusion of separately written HTML codes that were incorporated into the prototypes to improve interactivity. HTML codes were used in most of the components of the prototype in order to specify the exact characteristics and interactions desired, for instance, the graphs were all made from incorporating separate HTML codes. This tool was used because of its simplicity; experience obtained from using it in previous projects; and the variation of functionalities it offers.

4.2.4 Photoshop

Photoshop was used for editing images that were implemented within the prototype. With the batch operation, it was fast to process many images at once without having to repeat the same action several times. Photoshop was chosen because of the previous experience in using it, which made it fast and easy to handle.

4.2.5 KanbanFlow

This is a web-based application for managing projects based on Kanban Agile methodology (Williams, 2012). Since Personal Kanban (Benson & Barry, 2011) was followed as a project management technique, this tool was used for keeping track of all the tasks that were to be completed during the project run. Among the interesting features, which this tool offers is the combination of the Pomodoro technique (Šmite, Moe, & Ågerfalk, 2010) popular in Kanban method for time tracking. This provided, for instance, an overview of time spent on each task and serve as a reminder of when to take breaks in order to work productively.

4.3 Requirements Establishment

In software development, requirements are specifications that help to explain what the purpose of the system is in terms of what it should accomplish and how it should work in order to meet the users’ needs. These specifications must be clearly described to avoid any ambiguity and prevent misunderstandings when designers and developers try to fulfil them (Sommerville & Sawyer, 1997).

Since this thesis followed genius interaction design approach, the initial system requirements were gathered through the use of research based personas (Goodwin, 2011). The requirements were grouped into functional and non-functional and later used as a foundation for the tasks on the Personal Kanban board.

"Functional requirements describe the behaviour of the system or what the system should do and non-functional requirements describe the performance characteristics of a system and what constraints there are on the system and its development" (Rogers et al., 2011).

4.3.1 Research - based Personas

The designed artefact is mainly to be used by clinicians, medical students and the public. Rich descriptions of these typical users of this artefact are provided in order to design a new solution to the domain problem. The research-based personas and scenarios described below present the initial requirements for each of the user groups based on the existing documentation study of the state of current practises.

Persona 1: Researcher
Name: Ola Nord
Age: 44
Residence: Linkoping
Occupation: Researcher
Hobbies: Swimming, Bicycling, Reading.

Ola has worked as a researcher at the Institute of Oncology for over fifteen years. In addition to this, he has also built into his job some features of an academic career like publishing academic articles, lecturing and supervising masters and doctoral researchers with the university partners. He has a great enthusiasm for his job, enjoys what he does and likes to use appropriate new technologies to achieve maximum productivity. Reading is one of his favourite hobbies, be it something to do with the research he is involved in or a sci-fi novel.

He is also a member of the Linkoping bicycling club and likes to go swimming, with his family in the weekends.

Scenario 1: A day in life of Ola Nord

Ola's current research involves understanding how and why one particular type of cancer is increasingly found in one specific age group of people living in the most populated regions of the country. He started with this research project one year ago and it is therefore still in its infancy.

Since the data of the national cancer registries are not publicly accessible and Ola does not have permission to access them instantly, he uses data from the publicly available annual reports whenever he needs to either analyse some data or produce small reports about certain cancer incidences. He has been looking at reported data on the national cancer incidences and would like to be able to compare it with the trends observed from other countries. Since he does not have access to the foreign cancer registries either, he therefore uses the publicly available annual reports from the selected countries. Instead of going through each one of these long reports year after year, he uses the newly designed interactive system that was recently implemented at Ola's office. With this system, Ola is now able to search for the reports per countries and years of publication. The retrieved reports will automatically load to the system and allow him to start working on the content of the report.

He can for instance, with simple commands; he can select two regions on a map and use the selected variables such as age, sex, domicile, type of cancer and reporting hospital to figure out the different occurrences of that particular type of cancer. He could also be able to expand the dashboard to get visual presentations of the selected data in graphs, charts and highlighted patterns.

Ola is very happy because he can finally interact with the system directly and obtain the information he is looking for instantly without having to go through the tedious tasks of looking through the never-ending rows and columns of numbers, something that was not possible a few months ago.

Persona 2: Physician / Clinician

Name: Bob Joback

Age: 60

Residence: Stockholm

Occupation: Physician

Hobbies: Fishing, Diving, Sailing and Cooking.

Bob has been working as a physician/clinician for over thirty years now and is hoping to continue for at least the next twelve years. He is a hard working physician, likes his job and has always accomplished everything he started. He is also willing to take leadership roles whenever needed. He has many friends and has saved many people's lives as a medical personnel. He owns a small sailing boat and likes spending his holidays and long weekends with his family at sea especially in the summer.

Scenario 2: A day in life of Bob Joback

Bob has recently received a relatively young patient with a severe running stomach but also suffering from multiple medical conditions and malignant melanoma of skin being one of them. These other conditions include asthma, diabetes and epilepsy. In addition, the patient is allergic to penicillin. This patient has also participated in one clinical trial in Norrbotten before he moved back to Skåne, but he has lost all his trial documentations in the moving process.

Bob runs some medical tests but he would very much like to find out more medical information about his patient before he prescribes any medications. In addition to the information found in the electronic patient journal, he is interested in knowing more about this particular trial his patient has participated in; the trial's intervention methods used; the outcomes of the trial; and what clinical guidelines are available for patients with similar conditions.

Knowing what he is looking for, Bob uses the available system to access some different medical databases in order to mine useful data that may guide him in finding the right prescription. Despite geographical and electronic barriers, Bob uses only a few minutes to obtain all the details of the trials his patient participated in while living in Norrbotten and is also able to obtain the relevant clinical guidelines addressing such a situation. All this was possible with the help of a few clicks of the mouse and Bob is now able to prescribe something to his patient that will ease his running stomach.

Persona 3: Relative (non-patient)

Name: Oliver Ansatt

Age: 30

Residence: Uppsala

Occupation: Farmer

Hobbies: Horse riding, antique collections, gardening and dancing

Oliver grew up in a typical farmer's family and most of his family and relatives live nearby in the same village where everyone knows everyone. Unlike his younger siblings who opted to move to big cities, Oliver stayed in the village after completing his education at the Uppsala Agricultural College to help with the family farm business where he obtained a lot of his current farming experiences. He is currently self-employed and lives in his small farm with his wife, a newborn baby boy and a dog. He has solid computer skills and tries to stay active in the social media where he communicates with his siblings and friends from college.

Scenario 3: A day in life of Oliver Ansatt

Last week, Oliver's seventy five year old grandfather was diagnosed with prostate cancer. Since there are many waiting to be treated, this old man has to wait in a queue for at least three months before his treatment can begin. The news has shocked the whole family and even though no one knows much about the disease, everyone is trying to be supportive to the poor old man. Oliver has been searching the Internet to see if there is a faster alternative or some treatment discovered somewhere. He comes across the interactive online cancer reports where he discovers a function for visualising data from several online data sources for clinical trials. He realises that the tool has other interesting functions, which can, for instance, allow the user to have an overview of the different types of cancers and their occurrences across the country and also link these cancer types to their respective clinical trials. Luckily he finds some useful information about clinical trials related to this particular cancer in the nearby town that are still recruiting. He shows this to his grandfather who decides to contact the responsible institute for the trials and hopes for the best.

Persona 4: Medical student

Name: Rosemarine Timian

Age: 25

Residence: Halland

Occupation: Medical student

Hobbies: Reading, Shopping, Fashion and Walking.

Rosemarine is in her last year as a medical student at the University of Halland. She has been working very hard throughout the years as a student and is looking forward to finishing her studies and start working. She spends ten hours weekly at the regional oncology research centre helping out as her part time job. She has relatively good computer skills and she has obtained a good amount of experience in data analysis from working at this research centre, where she got introduced to a good number of computer programs and systems that the researchers use especially when working with huge amounts of data.
Like most girls of her age, Rosemarine likes shopping and fashion and she enjoys taking a stroll with her mother's dog outdoors.

Scenario 4: A day in life of Rosemarine Timian
She is busy finalizing her findings for a student project that she has been involved in for a year now, where she needs to find out the intersection of aging and prostate cancer in the different regions of Sweden and then compare the results with the other Scandinavian countries. With the help of this well designed system available at the university hospital, she is able to obtain some useful information by analysing a combination of data accumulated over several years from several registries. She is very pleased with the system's simple design solution because it enables good data mining from the registries and simplifies the whole process of data analysis. It also provides alternative ways of looking at the data and which does not require special computer skills.

4.3.2 Functional Requirements

- The artefact should allow direct interaction with its users.
- The designed artefact should provide a simple way of viewing data presented on the annual reports.
- The artefact should be able to filter search requests following criteria as specified by the user.
- The artefact should be able to group and display mined data as requested by the user for example; region-wise, disease-wise or age-wise.
- The artefact should provide a possibility for comparing the data reported from the cancer registries from different regions around the country.

- The artefact should have a search function for users to perform free searches.
- The artefact should provide a possibility to search for the nearest clinical trials related to the selected type of cancer.
- The artefact should have different ways of visually presenting the data.
- The artefact should have clear and visible display of the important functions.
- The artefact should allow the possibility to compare data presented in various annual reports.

The artefact should use icons and terminologies that are easy to understand even for persons without medical background.

4.3.3 Non-functional Requirements

- The designed artefact should be able to access publicly available yearly reports.
- The designed artefact should be effective to allow users get instant responses to their requests.
- The designed artefact should provide up to date information based on the data obtained from the regional and/or national cancer registries.
- The designed artefact should be available for everyone since the data presented in the yearly reports is already public.

4.4 Demand Specification

Pre-defined requirements needed to support the existence of the designed artefact include;

- The availability of Internet connection.
- A computer is the recommended device especially when displaying dashboards with a lot of content or for analysis purposes.
    - Other devices such as smart phones and tablets may also be used for simple informative tasks but would not be recommended for analytical tasks that require larger viewing surfaces.

4.5 Chapter Summary

This chapter has introduced the basic activities that were involved in the process of interaction design and accounted for how they will be distributed in the rest of the chapters in order to create a smooth flow. It has explored the tasks and workflow; tools and technologies used for the designing activities, and presented the requirements and how they were established.

Chapter 5: Prototype Development

This research relied on the multi-methodological approach to Information Science research as presented by Nunamaker Jr, Chen, and Purdin (1990); Systems Development and Experimentation, where a prototype was developed, tested and evaluated. Eventually some theoretical hypothesis could be generated on how the results of this research will influence the way of interacting with clinical registries.

The prototype developed in this thesis aimed at presenting several user-friendly interfaces that can be capable of providing new ways of viewing data mined from the cancer registries in order to enhance information extraction and knowledge discovery. Due to time constraints and the nature of the data stored in these clinical registries, the designed prototype was not fully implemented in this thesis, instead, it has been used to demonstrate the feasibility of a potential coming project. Therefore, product development and technology transfer were not a part of this thesis work but may be completed separately later on.

The prototyping phase described in this chapter involved three iterations through which the design elements of the artefact evolved from hand drawn sketches to highly interactive features. Those iterations are preceded by a description of how the prototyping tasks were visualised and managed on a Personal Kanban wall.

Usability testing in a confined environment, as described in the Section 5.3.3, was performed where users were able to interact with the artefact and provided feedback. The evaluation of the final prototype comes separately in the Section 5.4.3.

5.1 Personal Kanban

Personal Kanban was used to manage the development of the actual user interfaces. According to Benson and Barry (2011), it is suitable for projects carried out by individuals and small teams and it emphasises on work visualisation and limiting work-in-progress (WIP).

By following this agile approach, it was easy to visualise the amount of tasks involved in prototyping; limit the work in progress; do the right prioritization when executing the tasks and focus on the work at hand. The user interface design tasks were based on the established requirements. Since this thesis was a one person's project, Personal Kanban fitted better than Kanban or other Agile approaches for managing tasks. KanbanFlow (KanbanFlow, 2014) was the tool used as the electronic Personal Kanban board.

As shown in the *Figure 5.1*, the Personal Kanban board was divided into three columns; *backlog*, *in progress,* and *done*.

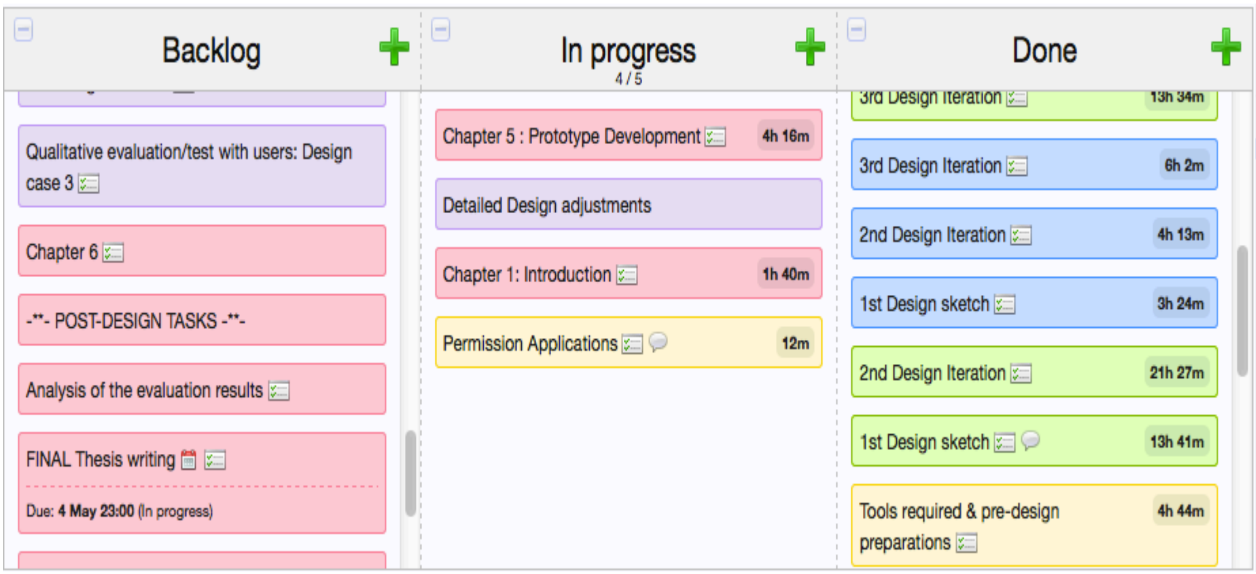


Figure 5. 1: Personal Kanban board

The backlog column consisted of all the tasks to be completed during prototyping. They were colour-coded and added to the backlog column then moved from left to right across the board following their order of priority.

A WIP limit of five tasks was set to the 'in progress' column in order to control the number of tasks to be handled at the same time and to save time by eliminating too much task switching. In this way, the tasks were executed systematically and it was easy to prevent having too many unfinished tasks, which could easily lead to losing control of the workflow. For instance, a task was first to be moved to the 'in progress' column when it was ready to be

processed, but only if there was room for it. Upon its completion, it was moved to the 'done' column.
The Pomodoro Technique was useful not only for tracking time used on each task, but also was a reminder of taking in-between breaks in order to enhance productivity while working on the tasks. A series of short 5-minute breaks were taken after devoting full concentration on a task for 25 minutes.
In this context, these short breaks are known as the Pomodoros. A 20-minute break followed after working 5 pomodoros. This approach was an encouragement to breaking down the work into simple and focused tasks.
Even though the pomodoros were a bit distracting at the beginning, it was an interesting and positive experience to try out Personal Kanban when prototyping.

5.2 First Iteration

This iteration began by taking a close look at one of the annual reports for cancer incidences in Sweden, which consisted of aggregated data presented in many rows and columns. A thorough look through the data was tedious and time consuming, but it was a crucial stage for defining a new way of visualising this kind of intensive rich data. Due to the sensitivity of the data stored in these medical records, the annual report was used in order to identify the key *variables* that matter when it comes for presenting this kind of data.
It was also important to understand the way this data is currently medically grouped in order to follow the registry patterns. This is because this thesis does not restructure the whole cancer database that communicates with other systems such as the Electronic Patient Record (EPR) systems, but it introduces user-friendly interfaces for data visualisation mined from these databases. Therefore, it was important not to alter the already established patterns of data from different systems involved even though these systems influence the design results. For example, not all the data collected from the EPR systems is presented in various reports, which are resulting from mining the registries, but rather a selection of the whole data set.
The results of this iteration are, the first hand-drawn as well as electronic sketches of the three design alternatives. Tools used during the first iteration are pen, paper and Balsamiq mockups.

5.2.1 Interface Designing

After the system's key variables and their relations were identified, it was time to convert the rows and columns of numbers into user-friendly interfaces that are intuitive and less tiresome to extract information from. The QOC method was used in order to define the design space for the different design alternatives.

Interface designing started with sketching different design solutions on paper. After selecting the design alternatives, which best met the requirements described earlier, Balsamiq mockups was used to convert the hand-drawn sketches into electronic ones.

5.2.2 Questions, Options and Criteria (QOC) Formulation.

Before sketching could begin, decisions were made regarding the features that could be included in the interface's design space. This was through the use of the QOC method where questions, options and criteria were formulated in order to rationally guide the decisions regarding the *where, why*, *what* and *how* of the design features in relation to purpose of the interface being designed.

The formulation of the QOC started with a thorough understanding of the established requirements and how the system components were related. This was a starting point to the ideation of design features. From the requirements, important interaction problems were outlined and questions formulated. The questions were used to structure the design space by addressing the key design issues.

The Options created for each question represented the design options to answer the particular design question. Each question had at least two options created, but in general there were more than three options for each question.

The Criteria presented the necessary characteristics of the artefact. It also described the requirements, which needed to be met. Therefore, Criteria were used to assess and compare the provided options. When designing the criteria for each question, consideration was put on making sure that the criteria did not overlap and that they were objective, and specific.

In order to find out which design alternative fits best for a particular design problem, the QOC for that particular problem was first defined then the options were matched against the set criteria and the answer was provided by the positive links between the options and the criteria. *Figure 5.2* shows an example of the QOC presentation of the design space for solving a problem regarding how a button should be displayed. The boxed option with the most positive links (solid lines) indicates the decision made. The remaining options with the negative links (dotted lines) were the ones considered to have had less logical support to the options provided for this particular design issue.

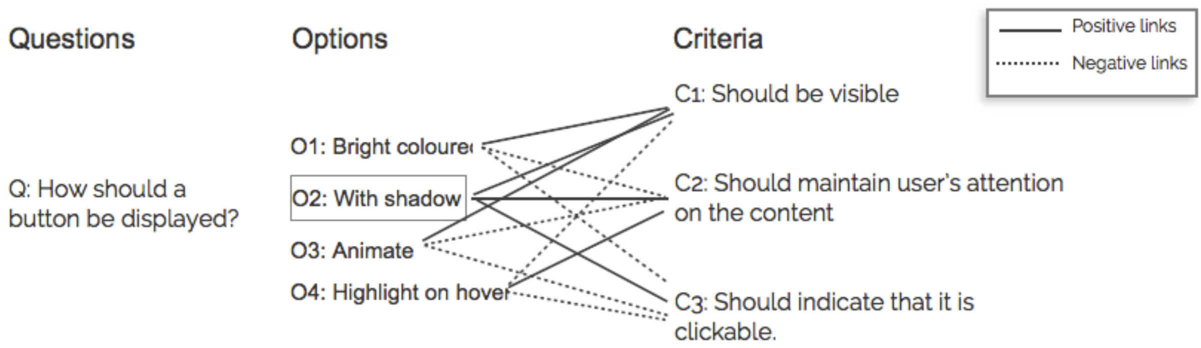


Figure 5. 2: An example of the QOC formulation used

The QOC method was applied in the first and third iterations where design problems were identified and required solving. During the second iteration, this method was not relevant due to the nature of the activities that were carried out. In the third iteration, some modifications to the initial set of QOC were made in order to accommodate the design input from users after the usability testing. All the QOC diagrams created for each question were made easy and simple to manage and understand. A full presentation of these diagrams created is found in the *Appendix 2a*.

5.2.3 Parallel Design Process

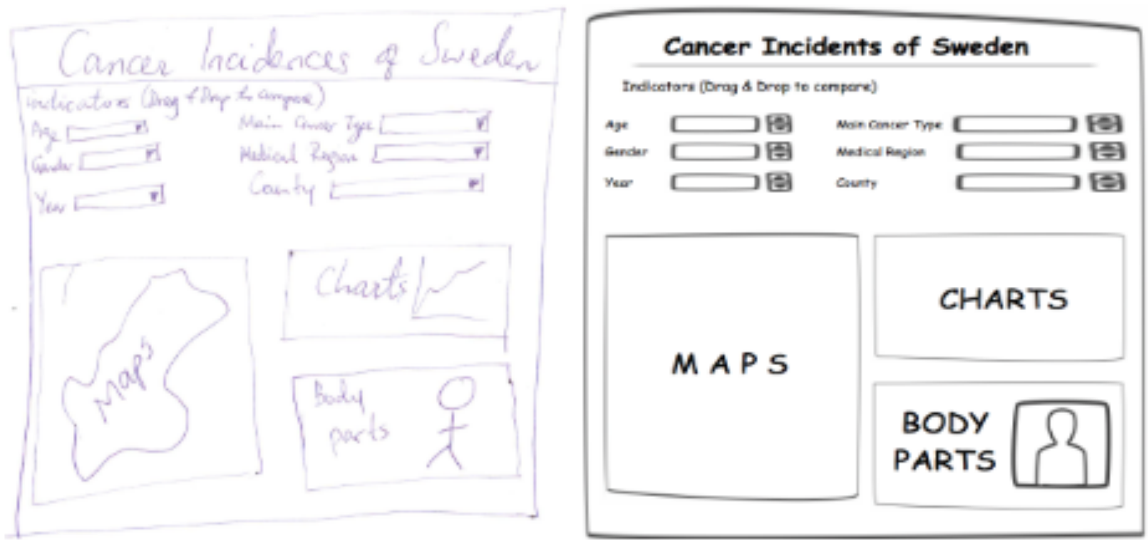
Interaction design process created three different design alternatives as suggested by Rogers et al. (2011),

More than one set of sketches was made in order to present the solution from different angles. These design alternatives were made keeping in mind that the task here was to create something new and not to redesign. During this iteration, emphasis was put on the top features of the designed alternatives. This was to avoid spending too much time on the design solutions that would be revised at the end of the next iteration.

5.2.4 Initial Sketches

Figures 5.3a, 5.4a and 5.5a below present the appearances of the first pages from the three hand-sketched alternatives and *Figures 5.3b, 5.4b and 5.5b* shows the electronic sketch equivalences.

Design alternative 1:



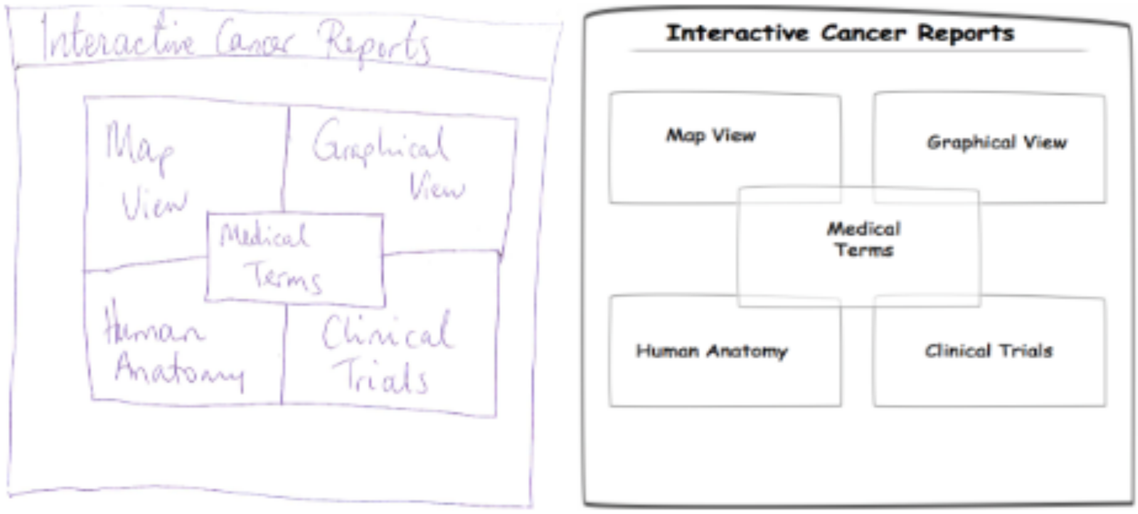
(a) (b)

Figure 5. 3: Hand-drawn and electronic sketches of design alternative 1

This first design portrays a dashboard theme. The top part of the page just below the heading contained the list of variables and the lower part contains the three viewing options; maps, charts and body parts.

The idea behind this design alternative was to allow users to be able to drag and drop the variables from the list onto one of the three viewing modes. Data of the specified variable would be instantly displayed as the user drops the specifications onto the viewing section.

Design alternative 2:

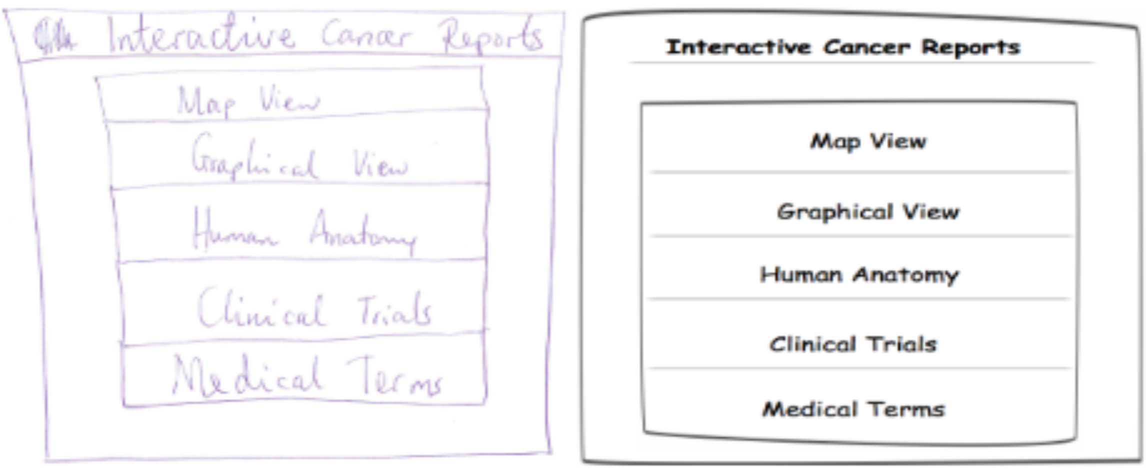


(a) (b)

Figure 5. 4: Hand-drawn and electronic sketches of design alternative 2

This design alternative depicts a slightly different approach of making the user first decide how they want to view the data before giving them a chance to select the variables.
The page was divided into five boxes, which represented the different ways through which data could be viewed. For example, by selecting the map view, the user will be provided with all the possible variables that could be viewed on from a map.

Design alternative 3.



(a) (b)

Figure 5. 5: Hand-drawn and electronic sketches of design alternative 3

More minimalist than the first two alternatives, the third design alternative presents the five different options of viewing the data on horizontally arranged containers across the page. This alternative, like the second one, limited users from selecting the variables before selecting the way in which they want to view this data. Variables provided on the next page, after selecting one of the five options, would depend on the view mode selected.

These sketches only show the intended buttons and functions and their positioning and labelling; but none of the functions or buttons works at this early stage. *Appendix* 1 shows the two key pages for each of the three alternative designs that resulted from this iteration.

5.3 Second Iteration

During this iteration, some interactivity was added to the sketches developed in the previous iteration and usability testing was carried out on all the three design alternatives. Users were invited to interact with the artefact and provided feedback, which was utilized in the iteration that followed, in order to adjust the design features of the final prototype. After analysing the results obtained from the users, some good elements from the three design alternatives were merged into one design concept. In these two last iterations, Komodo Edit and Justinmind prototyper tools were used to further evolve the sketches brought from the first iteration.

5.3.1 Interface Designing

Simple and common interaction elements such as the tooltips and hover states were added to the sketches of all the three design alternatives. Links between pages and between different design elements were also established and some background images and colours were added during this iteration.
At this stage, the prototypes gave more impressions and began to take the shape of what could have been of the finished products, than the sketches alone suggested.

5.3.2 Parallel Prototypes

Figure 5.6 below present the top pages of the three parallel prototypes resulting from the initial sketches in the *Figures 5.3, 5.4, and 5.5* above after they underwent a design transformation. *Appendix 3* shows more screenshots that were developed for each of the design alternative during this iteration.

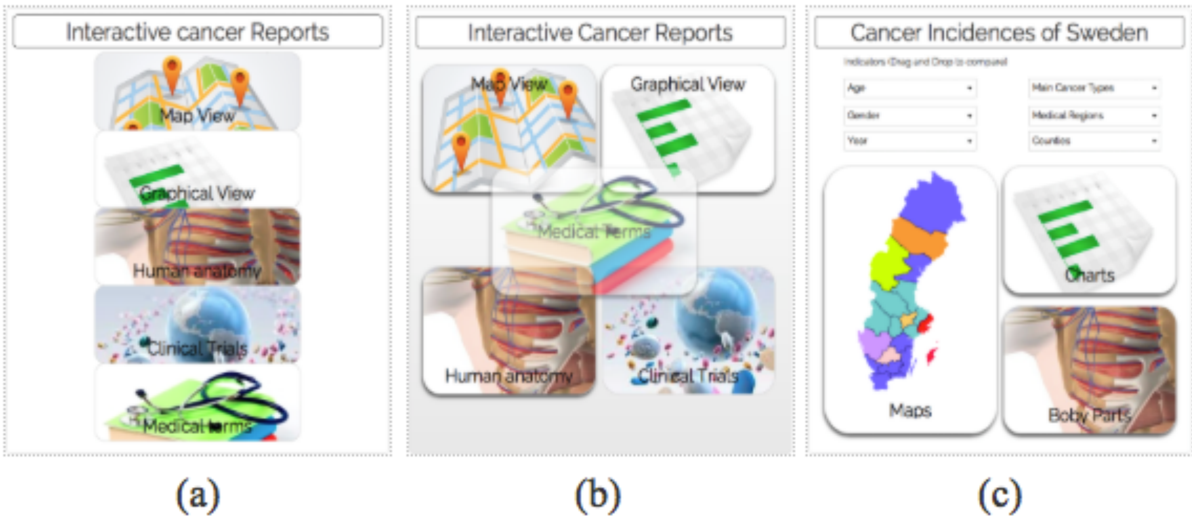


Figure 5. 6: Design alternative 1(a); Design alternative 2(b) and Design alternative 3(c)

5.3.3 Usability Testing

In this case, the usability testing did not aim at selecting a winner design alternative out of the three discussed. It rather aimed at getting a feedback from users regarding what they found as good or bad design features, which is highly relevant for the design refinement. This is the time when users were involved as evaluators of the designed artefacts whose valuable feedback was to be implemented in the final prototype.

The usability testing took about one hour for each participant and was performed in a quiet room where users were free from disruption. There was a total of five participants who participated one at a time and were equipped with a laptop, a pen and two print outs, a consent form and a set of tasks.

The laptop contained all necessary electronic components of the usability testing, namely, the three designed prototypes and a semi-structured questionnaire to be found in the *Appendix 4a*. The *Appendix 4b* shows the consent form with a brief introduction about this research thesis and its purposes and goals; a brief explanation about the content and purpose of the questionnaire. Additionally it provides information about handling the information and requests for explicit consent. It informs that their participation is entirely voluntary.

The set of tasks was used to familiarise users to the designed alternatives provided before rating the artefacts at different usability aspects is provided in the *Appendix 4c*.

The questionnaire was based on Jakob Nielsen (1994) usability aspects of learnability, ease of use, effectiveness, efficiency, error prevention, utility and memorability , and user's satisfaction. It contained a total of seventeen rating statements and questions. Fourteen of them were rating statements, which were used to assess the prototypes against the usability aspects. The *Table 5.1* summarises the rating statements from the questionnaire with their matching usability aspects. The ratings were: *strongly disagree, disagree, neutral, agree and strongly agree*. The remaining three were open questions, which allowed users to freely express their views regarding the presented designs and to provide ideas on how to improve the interfaces' usability.

Table 5. 1: Rating statements based on Usability aspects

Aspects	Rating statement
Ease of use	-I was able to complete the tasks without difficulties. -Terminologies and icons used were clear and easy to understand. -I think the prototype was consistent, easy to follow how the different parts were connected and easy to navigate through.
Effectiveness	-I was able to complete the tasks without any help.
Efficiency	-The tool was slow and confusing to use.
Utility	-The prototype provided the necessary functionalities to enable me perform the tasks I was assigned.

Learnability	-The simple and intuitive design made it easy to learn how to use the tool. -It will take short time to master the usage of this tool.
Memorability	-It was easy to forget how to use the tool if I do not use it on the daily bases.
Error prevention	-I kept on making errors while performing tasks. -The tool prevented me from selecting undesired variables.
Satisfaction	-The design presents information in an aesthetically pleasing manner -I enjoyed using this tool. -I would like to use this tool in the future.

Since the designed artefact aimed at being used by both experienced and novice users, this testing tried to find out how easy it was for users to accomplish the tasks the first time they come across these interfaces and whether it was easy to learn the artefacts' operations after the initial instructions. This was important for two reasons: first, to keep the designed artefact easy to learn and secondly to spare users the effort of learning everything from the start every time they interact with the system.

Users tested whether the designed artefacts fulfilled the set tasks. For each design alternative, comments were noted where task execution became difficult and where users thought the process did not go as smoothly as expected. In addition, users provided some suggestions on how the artefacts could have been improved in order to enhance their effectiveness..

Users tested whether the artefacts had the appropriate functionalities that were necessary for solving different tasks. For instance, there were five tasks, through which users could assess the utility of the designed artefacts.

After getting to understand the concept behind the three designed artefacts, users assessed how easy it was to establish proficiency. Users were asked whether it would be easy or difficult to relearn using the artefact if they did not use it daily.

Error prevention was another aspect, which was considered during this usability testing. Users checked whether the artefacts were designed in such a way that users would not easily make errors unintentionally. They would, for instance, check if some variables were made inactive in cases where it was not possible to select them. In addition, the number of errors made when performing the tasks was recorded as well.

In order to check for users' satisfaction, they were asked whether they enjoyed executing tasks from the different design alternatives presented to them.

The open questions and the comment option in the semi-structured questionnaire gave the testing participants the possibility to express their design opinions, point out usability problems that they have observed, comment on the best design aspects of the presented design alternatives and comment on how the designs could be improved in the iteration that followed.

The testing itself involved the following: briefing users on how data is currently presented to users and how the novel designs intend to help users view this data; explaining to users that the testing intends to assess the designed artefacts and not their performance; requesting users to read and sign the consent form; letting users explore the artefacts through executing some simple tasks; completing the questionnaire where users rated the three design alternatives and commented on the design aspects; and debriefing and thanking the users for their participation. For comparison purposes, time taken to perform each task for each design alternative was tracked. The results of the usability testing as presented in the *Appendix 4d* were analysed in the sub-section that follows hereunder.

The order in which the three design alternatives were tested was altered. This was because users tend to be fresh and have full concentration on their first attempt especially when the testing involve more than one artefact, which is the case here. A reasonably short questionnaire and few tasks were set up to encourage user participation and maintain the momentum.

The five users who participated in the usability testing were made anonymous and the responses to the questionnaires were not easy to identify any participants in particular. The information collected regarding participants were their gender, age, the frequency of their computer/internet use and whether they have, at some point before, searched the web for medically related issues.

The participants aged between twenty and forty nine years and included four males and one female. They all used computer/internet daily and have previously searched the web in search for medical information.

5.3.4 Analysis of the Results

The main reason for performing the usability testing on all the three design alternatives was to obtain users' input and to apply it to the final prototype and by that improve its usability. Therefore, the testing resulted in a hybrid of the three prototypes, which contained the best design aspects from each prototype in addition to the suggestions obtained from users' input. These results showed that all the three design alternatives managed to include the usability aspects to some extent. In order to thoroughly explore the users' contributions, the testing results were categorised following the same usability aspects shown in the *Table 5.1*.

Ease of use: The results from design alternative one (D1) and design alternative two (D2) showed that the interfaces were not the easiest to use. Even though it was possible to complete the tasks, some users experienced some difficulties. On the other hand, design alternative three (D3) seemed to have been the easiest of all. Users managed to complete the tasks smoothly.

Effectiveness: Since users had experienced some difficulties in D1 and D2, effectiveness was not experienced as high in these two design alternatives. Users thought D2 could have a successful execution.

Efficiency: The results showed that users implied that all the three design alternatives were efficient enough for the kind of tasks they had carried out and regardless of the tiny difficulties experienced in D1 and D2. For comparison purposes, time taken for completing each task for each design alternative was recorded and detailed in the *Appendix 4e*.

Table 5. 2: Average time taken to perform a set of tasks during the Usability testing

	Design alternative 1 (D1)	Design alternative 2 (D2)	Design alternative 3 (D3)
Average time	**04:25 min.**	**03:50 min.**	**02:44 min**

Table 5.2 shows the average time in minutes taken for performing all tasks for each individual design alternative. The table shows that D3 was efficiently performed than D2 and D1. There was a relatively noticeable time difference amongst the three designs but this is could be explained by user-friendliness, which was in favour of D3.

Utility: There were mixed responses regarding the functionalities provided for D1 and D2. Some users implied that the two design alternatives did not provide the necessary functionalities to enable them perform the assigned tasks. This was expressed by the negating the rating statement given on the questionnaire while others implied that all necessary functionalities were provided since they managed to accomplish the tasks. On the other hand, results showed that users experienced that D3 provided the right kinds of functionalities that greatly helped them to complete the set of tasks.

Learnability: Results strongly suggested that all the three design alternatives were easy to learn and that the learning process would not require excessive time.

Memorability: Results suggested that all design alternatives would be easy to remember after a period when a user would actually not be using them.

Error prevention: Even though there were no errors recorded for any of the design alternatives, results suggested that D1 possibly had a loophole that would have led users into making errors when executing tasks. D2 and D3 seemed to be resistant from errors.

Satisfaction: For each design alternative, results strongly suggest that users were pleased with the interfaces. It also seemed that users would consider using the artefact in the future. Even though the prototypes presented had appealing interfaces and fulfilled the usability aspects, there were several faults identified in each design alternative. *Table 5.3* summarises user comments and suggestions, which resulted from testing the three design alternatives with users.

Table 5. 3: A summary of the comments obtained from users

Design alternatives	**Comments**
Design alternative one (D1)	**What design features do you think made this design appealing?** -The dashboard was good but only for some tasks. -The background images fitted well. -Simple design - Dashboard.

	-The possibility to choose to look at the data from the map or chart or the body diagram. **What design features do you think should be included in order to improve this interface?** -"GO" button to press when i finish selecting what data i want to see. - more data options (categories eg. deaths) - Search function -Possibility to select more than one item from the same group eg. to select more than one age group. **Please add any comments or recommendations regarding the overall design.** - I could not select more than one variable eg. not possible to choose males and females. - More variables to make the selection more specific, eg. there was no where to select the "Deaths". -I think the dashboard makes the prototype seem incomplete; it will be good if this was not the only option.
Design alternative two (D2)	**What design features do you think made this design appealing?** -The map and the information that comes up when hovering. -The map feature was very handy **What design features do you think should be included in order to improve this interface?** -A better arrangement of the content. -The consistency of the design should be improved. **Please add any comments or recommendations regarding the overall design.** -not easy to find the medical region especially on the map, only the city comes up -Putting clinical trials on the top page is not a good idea. -Medical regions are too hidden. -Could not choose more than one variable. -The map on the map page was not in the centre of the page.
Design alternative three (D3)	**What design features do you think made this design appealing?** -It was easy to find what i was looking for. -I liked the clean design with calm colours. -The design was clean and it had more things to choose from than the first design. -The images used on the top page fitted very well. -The way the different parts were arranged on the top page was neat and easy to select one. **What design features do you think should be included in order to improve this interface?** -Possibility to see the data shown on the graphs -The functions were very good, but some small things should be done to clean up the

	looks. -The search function. **Please add any comments or recommendations regarding the overall design.** -I did not see why the clinical trial is put on the top page. -I think the heading is not so interesting. -Components of the top page are too squeezed together. Removing clinical trials will make space for the remaining ones. -Some parts were consistent but not all. -The heading was not the best. -I thought the medical terms were some kind of a medical dictionary. Maybe another name for that label. -Not very consistent when it comes to the position of the map, human body and the graph. Human body and the graph sections were consistent but not the map.

The overall testing results for design alternatives 1, 2 and 3 are found in the *Appendix 4d. 10b*. The usability testing results indicated that each design alternative contained some good design aspects that were worth carrying forward to the final prototype. The best design aspects from the three parallel versions were merged together and formed what became the final prototype, which was further developed in the third iteration. The Figure 5.7 summarises the parallel design process that begun with three simultaneous designs and resulted in one containing the best design aspects.

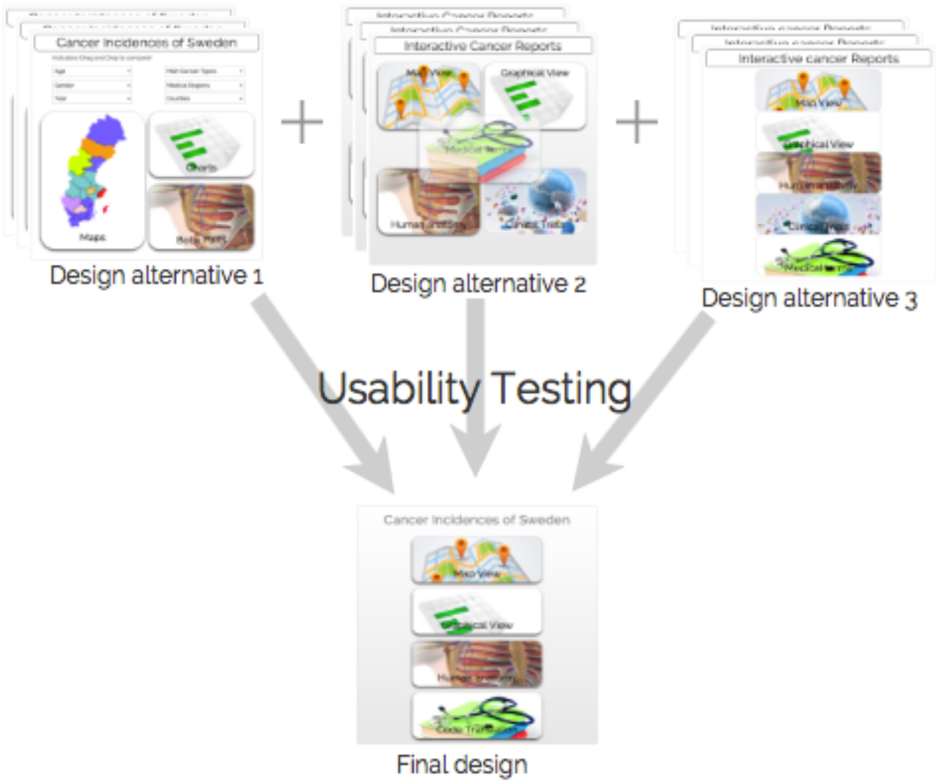


Figure 5. 7: Merging the parallel designs into one

5.4 Third Iteration

This last design iteration focused on developing one final prototype, which incorporated users design opinions and the best design aspects from those three design alternatives (D1, D2 and D3). In order to best use the design space and still come up with a good design solution, the QOC method was used. More complicated interactions were added to the interfaces and the design was fine-tuned with more consistent flow of design elements than the three alternative prototypes discussed in the previous iterations.

5.4.1 Final Interface Designing

The designing process in this iteration began with exploring the feedback from users. This was done by noting down best design features from the parallel designs, and by pointing out the new features that were suggested during the usability testing and were in additional found technically feasible. The design weaknesses were reviewed to find out which design alternatives required the least amount of alterations after which they were made a starting point in developing the final prototype.

During this iteration, the QOC model that was used in the first iteration was modified to incorporate the suggested design ideas. More questions, options and criteria were added and some of the existing components of the model were extended. For example, the *Figure 5.8* points out the extended section of one of the initial QOC presentation of the design space where the *list* was initially not specific enough regarding the presentation details. The modified section is the area inside the circle. In this case, a new set of QOC was created in order to solve the design issue, which originated from one option of another QOC set. Modification for this particular one was necessary because the initial question focused only on how the variables should be presented without considering that the presentation mode required further specifications. *Appendix 2b* presents the modified version of the QOC model, which was used in addition to the initial one in the A*ppendix 2a*.

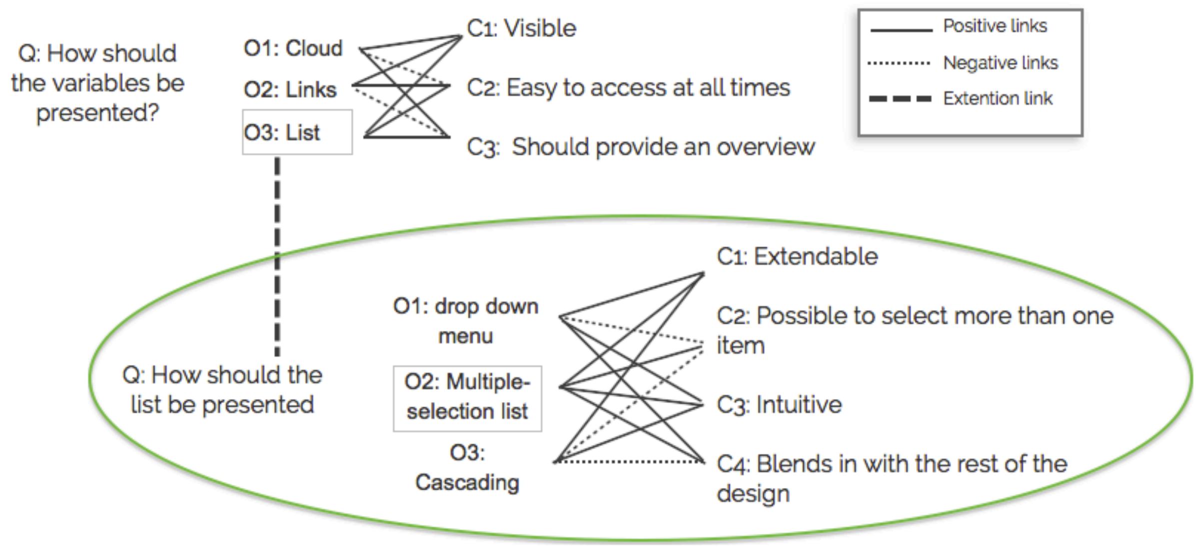


Figure 5. 8: An extended section of the QOC used in the third iteration

Hand-drawn sketches were made after the evaluation suggestions to get the final prototyping process started. Once the design suggestions, alterations and the best design ideas from the three alternative prototypes were incorporated into sketches of the final design, the sketches were made electronic. From the electronic version, page links were created and the necessary interactivity was added to the interface.

The dashboard, map, human body anatomy, and the graph were few of the best design features and aspects appreciated by users that were transferred into the final prototype.

In addition to these features, the testing results also indicated that users were quite positive towards the appearances of the presented design alternatives. They, for instance, commented on the clean designs, the nicely fitting images, calm colours that were used. Such comments contributed to the designing of the final prototype, which was later evaluated by expert.

The search function was among the suggested features that users lacked from the design alternatives they previously tested. This too was incorporated into the final design in such a way that users would also have the possibility to perform an advanced search with more options for fine-tuning search.

There were suggestions regarding the possibility to be view the data behind the graphical presentation. None of the design alternatives had such an option. In the final prototype, this was added as a *'show data'* tab, which could also be accessed from the map mode.

Since the testing was performed on one design alternative at a time, some suggested features were already taken into consideration in one of other design alternatives. This was still useful for the final design process because such features were considered as good features and were carried on to final prototype. For example, the possibility to have a *'GO'* button to suggest that a certain variable selection process has been completed and the nest can start. This feature describing the status of the actions was not included in the first design alternatives. Another example was the possibility for users to select more than just one variable when they were specifying what data they want to view, for instance, selecting age groups 30-34 and 35-39. This feature was not present in design alternatives one and two, but it was provided in the design alternative three. Therefore, the final prototype aimed at including all such features in addition to the newly suggested ones during the testing.

When designing the final prototype, faults pointed out during the testing were considered in order to avoid making the same design errors again. Some faults were discovered only in one or two of the design alternatives, but not in all the three alternatives. Finding solutions to such faults became easy since the design solutions were already available from one of the other design alternative and therefore, only required small adjustments to fit the final prototype.

More faults that were discovered by users during the testing included the inconsistent flow of the design elements of the interfaces; badly positioned interface elements and misleading names and titles on the labels such as '*medical terms'* which was later changed to '*Code Translation'* and *'interactive cancer reports'*, which became 'Cancer Incidences of Sweden. Another interesting weakness was having '*Clinical Trials'* on the top page of design alternatives two and three. This seemed to be confusing to some users. During this final iteration, this was not completely removed from the prototype, but it was moves under the '*detailed facts'.* There it could be accessed when users have already selected a particular type of cancer that would be connected to it. Solving this error in this manner was logical since the typical users would normally not begin by searching for the clinical trials but they would first search for the cancer type. All the design features of the final prototype are found in the *Appendix 5*.

5.4.2 Final Prototype

This sub-section presents a walkthrough of the final prototype, which was a result of the third iteration. In spite of the design trade-offs that had to be made along the way, as pointed out in the QOC formulations earlier, this hybrid prototype has managed to incorporate most of the users' design input in addition to the best design elements from the three alternative designs discussed in the previous iterations.

Figure 5.9 briefly introduces the structure of the prototype. This simple navigation map show the four suggested categories that will assist users to visualise data mined from the registries.

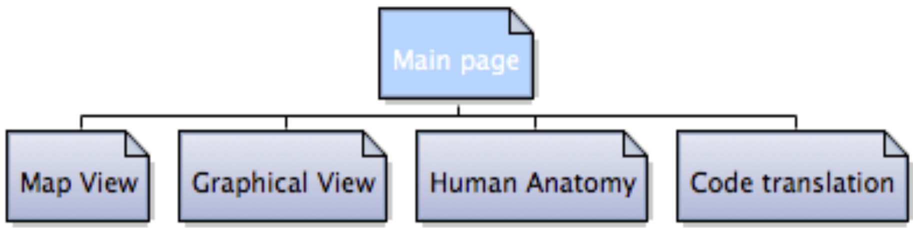


Figure 5. 9: Navigation map

With this design solution, the cancer registries can be presented through the map, graphical and human anatomy visualisation modes. The code translation mode presents the updated translations of the code standards for the anatomical locations of the different types of cancers. It was necessary to categorise this particular type of data in this manner in order to cater for the different needs of the intended user groups.

The design was made simple and kept neat in order to help users navigate smoothly and execute their tasks without difficulties as opposed to what is currently available. With the main page, shown in the *Figure 5.10,* containing few elements, users can finally stop worrying about how to quickly find their way through the data behind the never-ending rows and columns typically seen in the annually reports. Natural language was used to identify the different elements of the interfaces. In some occasions, a combination of text and images was used in order to improve the visibility and the perceived affordances of the design features used.

Figure 5. 10: The prototype's main page

From the different visualisation modes provided, users will be able to view the data in the reports that is relevant for that particular visualisation mode. This constrains the way in which the mined information can be perceived but at the same time, it prevents a possible information loss caused by poor data presentation.

An overview of the four categories: map view, graphical view, human anatomy and code translation; is presented in the *Figures 5.11, 5.12, 5.13 and 5.14*. Depending on the users' preferences, these categories can be accessed from both, the main page and from the individual pages making it easy for users to freely switch from one mode to another as they opt to.

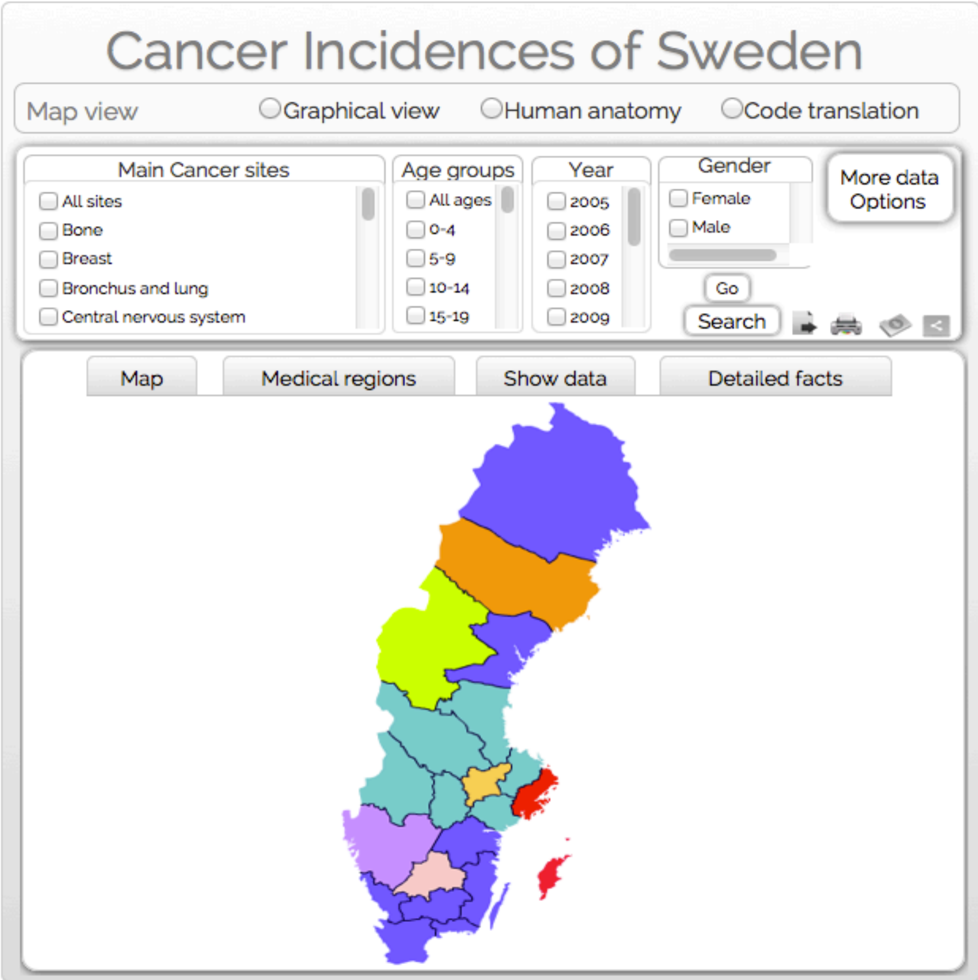


Figure 5. 11: The map view

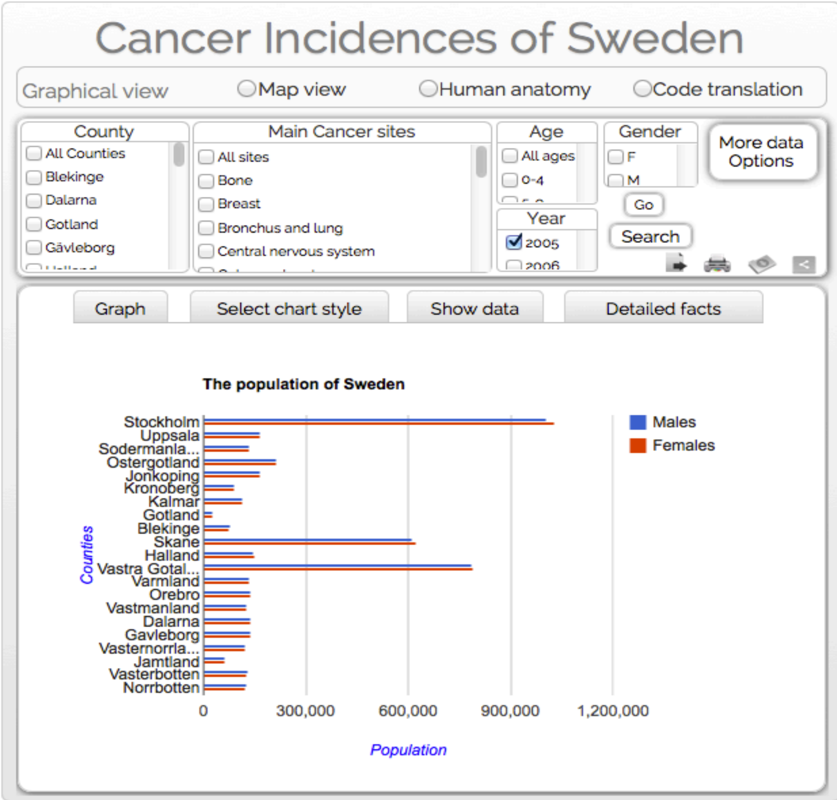


Figure 5. 12: The graphical view

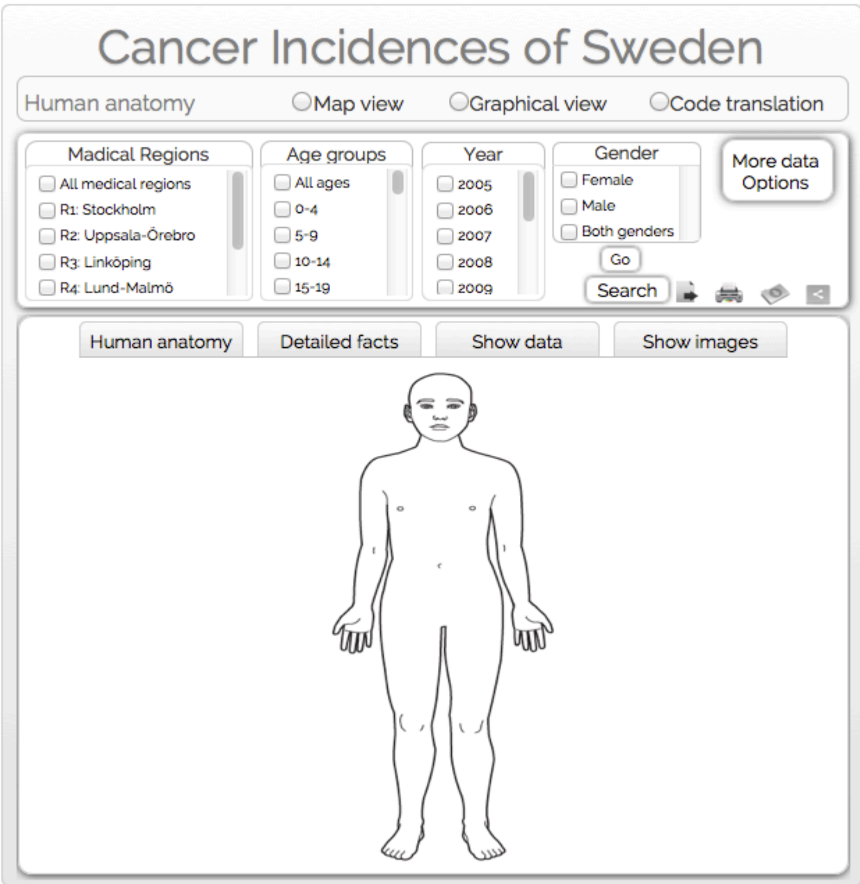
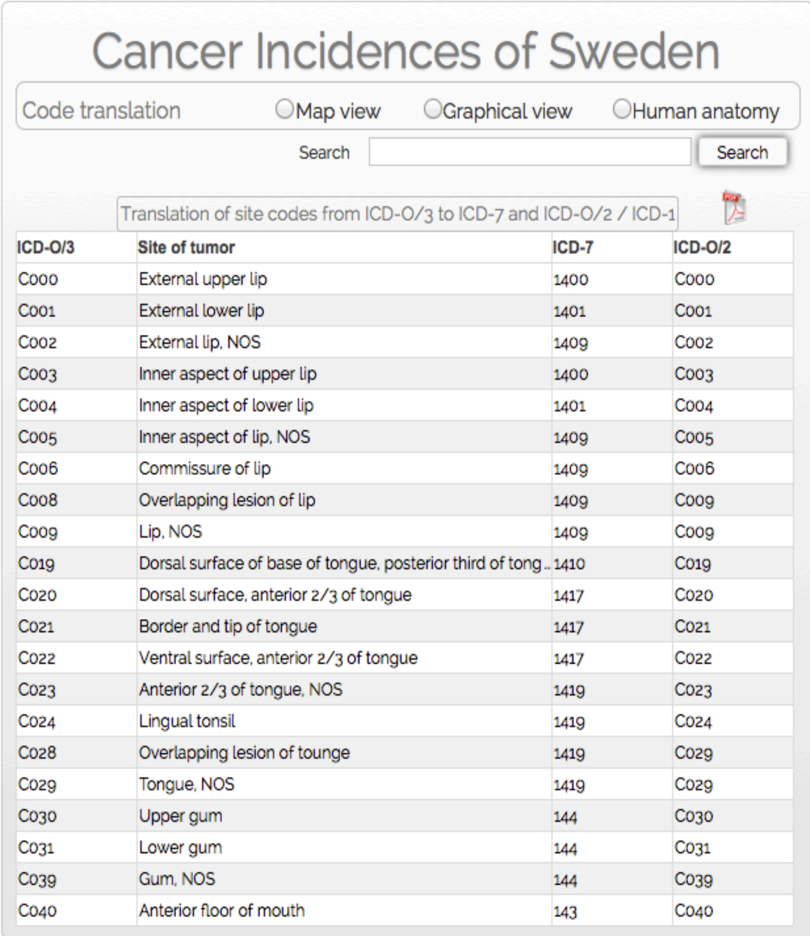


Figure 5. 13: The human anatomy view



ICD-O/3	Site of tumor	ICD-7	ICD-O/2
C000	External upper lip	1400	C000
C001	External lower lip	1401	C001
C002	External lip, NOS	1409	C002
C003	Inner aspect of upper lip	1400	C003
C004	Inner aspect of lower lip	1401	C004
C005	Inner aspect of lip, NOS	1409	C005
C006	Commissure of lip	1409	C006
C008	Overlapping lesion of lip	1409	C009
C009	Lip, NOS	1409	C009
C019	Dorsal surface of base of tongue, posterior third of tong..	1410	C019
C020	Dorsal surface, anterior 2/3 of tongue	1417	C020
C021	Border and tip of tongue	1417	C021
C022	Ventral surface, anterior 2/3 of tongue	1417	C022
C023	Anterior 2/3 of tongue, NOS	1419	C023
C024	Lingual tonsil	1419	C024
C028	Overlapping lesion of tounge	1419	C029
C029	Tongue, NOS	1419	C029
C030	Upper gum	144	C030
C031	Lower gum	144	C031
C039	Gum, NOS	144	C039
C040	Anterior floor of mouth	143	C040

Figure 5. 14: The code translation view

The interface pages consist of the title, navigation menu and a content container. For the map, graphical and human anatomy visualisation modes, the content container was further divided into two: the top container containing the different lists of variables for users to choose from and the bottom container containing the visualisation space where the specified data can be displayed. On the other hand, the code translation mode maintains the one content container due to the nature of the information to be displayed. Therefore, the title and the navigation menu consistently remain in the same position throughout the prototype page and the two content containers are consistently applied to all the subpages of the map, graphical and human anatomy visualisation modes. In this way, users can easily learn and use the interfaces without having to adapt to different positioning of design elements on every page they come across.

Some of the features that come with the designed prototype include the possibility to generate reports. In addition to the annual reports, users should be able to generate custom reports, a feature not available currently. This is especially useful to clinicians who often need to tailor their reports to specific variables following the different research works they are involved in. There were other important features that were included in the design even though they were not obtained from these particular registries directly. These include the possibility to view clinical trials related to the specific cancer types and the possibility to view the pathological images of the related kinds of cancers. These two possibilities were included in this design in order to improve the richness of the data that the typical users are interested in and to make use of data resources that are publically available through other channels that are not very straight forward to users. A good example is that of the clinical trials, where users would normally have to perform a separate search through some random web data sources without having the continuity of the search results. This design, provides a continuation of the type of cancer, which the user has already been looking at and connects it to the data of the related clinical trials. Data displayed from the clinical trials is obtained from reliable publicly available web-based data sources such as the clinicaltrials.gov. This kind of search within the same system will help users to save time and be more efficient in executing their tasks. Another example is that of the pathological images from the different cancers that are not normally presented in the annual reports due to their bulk nature and the fact that there might not be mandatory to convey the knowledge. The designed interfaces provide the possibility for users to view these stored pathological images of the different cancer cells in connection with the statistical data creating a rich combination that will enhance the mined data from these registries especially in an e-learning setting.

This design also provides several functionalities that are included in the different visualisation modes according to their relevance in relation to the specified data, which is presented. Some of the popular ones that have been used in all the visualisation modes include the search, hover and click functionalities, which help users to quickly view the data they want without having to browse through several report pages. For example, in a scenario where one would like to quickly have an overview of the new cases of a certain type of cancer in a particular region. By simply specify the variables from the provided lists and thereafter hover the mouse over that region and a small description box will appear to briefly display the specified data. Depending on the visualisation mode selected and the variables specified, such small description boxes can be displayed by hovering the mouse over the map, graph or over the different body parts under the human anatomy mode.

A detailed screen-by-screen presentation of the interface pages and the functionalities available in this prototype is provided in the Appendix 5.

5.4.3 Prototype Evaluation

In order to address the research questions the final prototype was qualitatively evaluated by experts who followed the ten earlier mentioned heuristics developed by (Jacob Nielsen & Molich, 1990).

Experts examined the interfaces of the prototype and suggested possible problems that users might encounter when interacting with the system. These experts were a combination of clinicians who were skilful experts and master students who role-played the public and medical student user groups. This combination was necessary in order to cater for all the intended three user groups. In addition to a good knowledge about the information systems in the healthcare field, the clinician experts have several years of experience working at the regional oncology centre with clinical registries particularly the cancer data resources. The student experts have a strong background in HCI, have good knowledge about information systems, and have experience in practicing usability methods.

Experts were chosen to evaluate the final prototype in this case as opposed to users because of limited time and geographical locations, which made them not easily accessible. Therefore, involving users from all the target user groups would have been too expensive and time consuming and beyond the scope of this thesis.

There were thirteen experts who evaluated the prototype one at a time. They were equipped with a laptop and two print outs consisting of a set of tasks and a list of the heuristics briefly elaborated. The laptop contained the designed prototype and the evaluation form, which is presented in the *Appendix 6a*. As shown in *appendices 6b and 6c*, the set of tasks was used to familiarise the experts to the different pages of the prototype before they got to analyse it; and the list of the heuristics was used as a reference and basis of the evaluation.

5.4.3.1 The Evaluation Process

The evaluation process started with a briefing session where the experts were welcomed to the evaluation session, introduced to what the research entails and got informed about the evaluation procedures. Experts were reminded to be specific and detailed in addition to clearly record the issues they discovered when interacting with the prototype.

All experts were provided with the same introduction to the session in order to make sure that they all received the same briefing especially because the different evaluation sessions were performed at different times.

A list of tasks was given to the experts so as to explore the different design elements of the prototype. There were ten tasks of which the first five did not involve using the newly suggested features and the last five involved using the new features of the prototype. The first five questions were therefore solved twice: by following the current practises, and by using the designed prototype. Time taken by the experts to perform the first five tasks was tracked for comparison purposes. *Appendix 6d* contains the details of how long each task took for the individual experts. The experts were thereafter, provided with an evaluation form based on Nielsen's heuristics listed earlier in the Section 3.5.2. On the form, they were asked to rate the prototype using a scale starting with 1-10 where 1=Very poor and 10 = Very good. There was also a possibility to add comments after each heuristic measure on the evaluation form. When the evaluation forms were completed, the results were directly saved to Google Drive (CITE), where the data was filtered and categorised according to the order of questions and comments in the form. This simplified the analysis and interpretation processes.

Oral feedback from the experts was provided in the debriefing session. This was done in a discussion with the experts where they elaborated their experiences of interacting with the prototype which they just evaluated and they were asked to suggested solutions for they have faults found. The whole evaluation session took forty five minutes.

The evaluation forms used to gather experts' feedback consisted of closed rating options and comment fields, which resulted into both qualitative and quantitative raw data. In order to gain a good overview of the evaluation results before the analysis and interpretation, the raw data was compiled, cleaned up and summarised into the ten heuristic categories. A detailed presentation of the results is given in the Chapter Six.

5.5 Chapter Summary

This chapter has provided a detailed description of the prototype development process and shown how this process was managed. The three iterations are also detailed presented showing what was achieved in each one of them. In addition, it also described how the usability testing and the expert evaluations were carried out.

Chapter 6: Results and Discussion

6.1 Results

6.1.1 Heuristics Dimensions

This section explores the results obtained from the evaluation of the final prototype carried out by experts. Table 6.1 summarises the results obtained from the closed rating options representing the ten heuristics, where H=Heuristic and E=Expert.

Table 6. 1: Evaluation from the closed rating options

Heuristics	E 1	E 2	E 3	E 4	E 5	E6	E7	E8	E9	E10	E11	E12	E13	Average
H1	10	10	10	9	10	8	8	9	9	9	9	6	7	**8.3**
H2	10	10	10	10	10	8	9	10	10	10	10	10	8	**9.6**
H3	10	9	9	10	10	7	9	10	10	10	7	9	5	**8.8**
H4	10	10	10	10	10	9	10	10	10	10	9	8	7	**9.4**
H5	10	10	10	9	10	10	10	10	10	10	9	10	10	**9.8**
H6	10	10	10	10	10	8	8	8	-	10	9	10	9	**9.3**
H7	10	9	10	10	10	7	10	10	10	8	10	8	5	**9**
H8	-	10	10	10	10	7	10	10	10	9	8	7	7	**9**
H9	10	10	10	9	9	-	10	10	10	10	8	-	5	**9.1**
H10	9	10	10	9	9	-	10	9	10	5	10	7	5	**8.5**

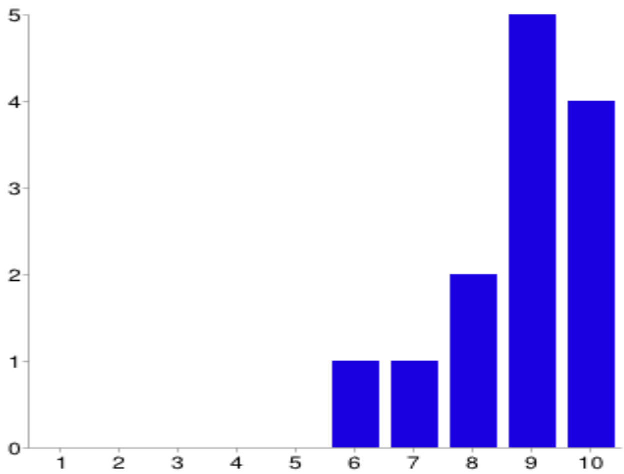
From the closed assessment options, each user's rating and an average value obtained for each heuristic was plotted on the table in order to have an overview of the scores. Experts' comments containing both good and poor design aspects were categorised following the ten heuristics and were explored in detail hereunder. A detailed presentation of the results obtained from the experts' evaluation is found in the *Appendix 6e.*

H1: Visibility of system status.

Table 6. 2: Evaluation comments for H1 – Visibility of system status

H1: Visibility of system status.	
Good design aspects	-The system had good information about what pages you were on.
Poor design aspects	-Not easy to know all the names of counties. -Some operations took a long time and no progress indicator was shown. This left me unsure as to whether or not anything was happening with the system or if I should try to click the element again. -Sometimes you don't know if the system is responding eg. After clicking- go- it is not easy to know when to hover the mouse, it the selection has been updated or not. -Dashboard hidden
Suggestions	-Make a small hint by text or highlight on the map or that human body when the data is updated and ready. -Names of counties on – off in map could be helpful

(a)



1	0	0%
2	0	0%
3	0	0%
4	0	0%
5	0	0%
6	1	8%
7	1	8%
8	2	15%
9	5	38%
10	4	31%

Figure 6. 1: H1 – Visibility of system status

(b)

Table 6.2 and Figure 6.1 show experts' evaluation results based on the first heuristic, *visibility of system status*. *Figure 6.1* represents the number of expert participants and their scores for rating the prototype. As presented in the *Table 6.1*, the designed prototype had an average score of 8.3 out of 10 on a scale of 1-10 where 1=Very poor and 10 = Very good. There were mixed response from individual experts when it came to the ratings. There was a variation between 6 to 10 of the possible 10 scores. Out of the 13 experts, 8% gave the score of 6; another 8% gave a score of 7; 15% gave a score of 8; 38% gave a score of 9; and 31% gave a score of 10. These results are quite good, since there were no scores under 5.

From the experts' comment, some poor aspects as well as good ones were pointed out. There were also two suggestions for improvements. The four poor aspects that were pointed out were: lack of progress indicator that would give feedback to users responding to their actions; operations responded slowly; a dashboard which was not very easy to locate; and lack of county names on the map.

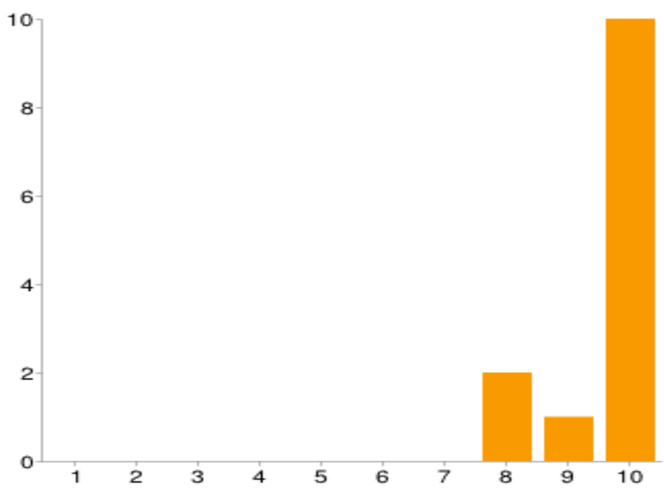
As possible solutions to these design weaknesses, suggestions were: creating some sort of a hint on, for instance, the map, to indicate that the user's actions are being executed; and labelling the different counties on the map especially for those who are not very familiar with the Swedish geography.

H2: Match between system and the real world

Table 6. 3: Evaluation comments for H2 – Match between system and the real world

H2: Match between system and the real world	
Good design aspects	-After looking at the pdf-document this system seems to fix all the problems related to searching and stuff. -Good use of icons and images -The map, body and graph are easy to use. -Not much to fault here, understood mostly what the different menus lead to. -It is a good idea that users can choose to view the data on the map or graph or that human body. -It becomes simple to find the different kinds of cancers. -Easy to understand. -But even though I did not know anything about the field of study, I felt I understood the most.
Poor design aspects	-Of course some of the words were hard to understand for me, who don't have any knowledge about the cancer and cancer cases.
Suggestions	-Skeleton ok for bone cancers. Maybe 3D body for the rest?

(a)



1	0	0%
2	0	0%
3	0	0%
4	0	0%
5	0	0%
6	0	0%
7	0	0%
8	2	15%
9	1	8%
10	10	77%

Figure 6. 2: H2 – Match between system and the real world

(b)

Results, as summarised in the *Table 6.3*, suggest that the designed prototype managed to successfully match the real world. Experts' comments show that the icons and images corresponded well to real world. There was also one expert's comment that compared how searches were performed in the designed prototype to the way it is currently done. This comment indicated that search-related problems, which were experienced when from performing tasks following the current practises, were resolved.

There was only one comment regarding poor design aspect.. This involved the terms used that were not very familiar to an expert who does not have a strong knowledge about cancer cases. It has been proposed as a remedy to use the skeleton image when displaying bone cancers and a 3D-body image could be used for the rest of the cancer types.

Figure 6.2, shows the number of experts who participated and the percentages of their resulting in an average scoring of 9.6 out of 10.

H3: User control and freedom

Table 6. 4: Evaluation comments for H3 – User freedom and control

H3: User control and freedom	
Good design aspects	-Nice search under Code Translation. -It was good that the different views are always available to switch between regardless of what information is showed in the lower part.
Poor design aspects	-The way to return to the main page was a bit hard, and should maybe be a bit more clear. Even though I thought it was easy to "escape" most of the situations.
Suggestions	-I think the icons should have been placed far from the main features since these functions are not used very often.

(a)

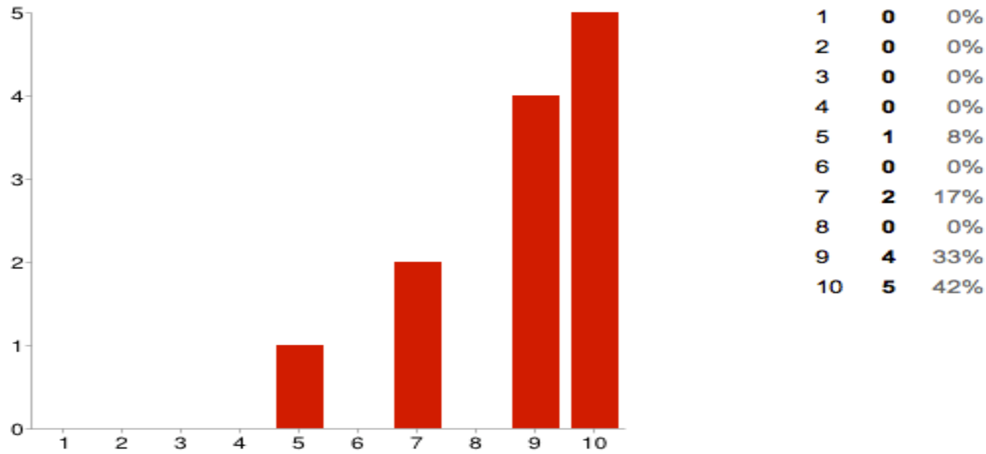


Figure 6. 3: H3 – User control and freedom

(b)

Even though the rating scores varied from 5 to 10 out of 10 as seen in the *Figure 6.3*, the majority of the thirteen expert evaluators, 42%, gave a score of 10 out of 10. 33% gave 9, 17% gave7; and 8% gave 5 against user control and freedom. The average score was 8.8 out of 10.

The search function under the code translation and the possibility to switch between the different views seemed to have impressed some experts according to their comments as shown in the *Table 6.4*. On the other hand, one expert thought it was not easy to return to the main page.

One of the experts meant that the icons for print, export, download, and share; were on the way for the main features and therefore he suggested placing them somewhere else, away from the often used features.

H4: Consistency and standards

Table 6. 5: Evaluation comments for H4 – Consistency and standards

H4: Consistency and standards	
Good design aspects	-Good! -Consistent positioning of the navigation bar was very good. -East to know where you are. -Good use of standard web widgets such as tabs, combo boxes and radio buttons

	made me able to navigate easily without any prerequisite knowledge of the system.
Poor design aspects	None
Suggestions	-However, I miss an option to reset the search criteria before starting a new search. It could be cumbersome to remove all the criteria from the last search if there are many of them.

(a)

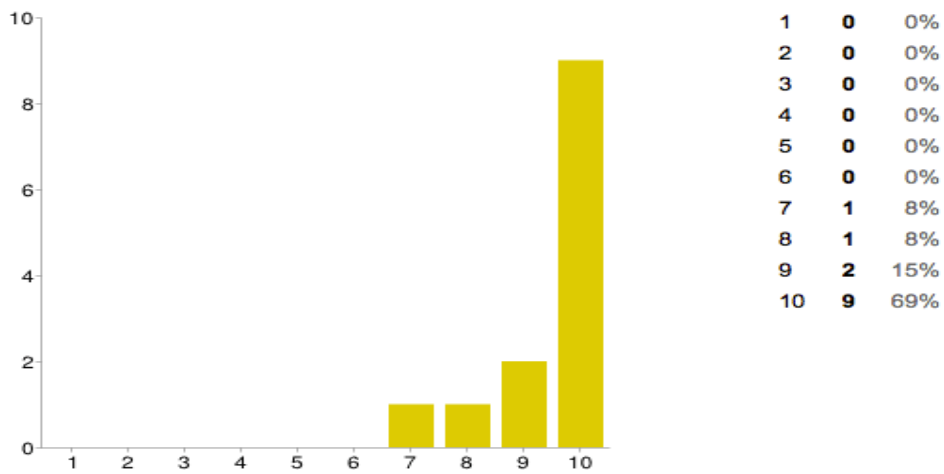


Figure 6. 4: H4 – Consistency and standards

(b)

As summarised in the *Table 6.5*, there were no poor design aspects regarding the consistency and standards dimension. Results show that the prototype demonstrated good use of certain design features that were obvious to users hence simplified navigating around the prototype. However, there was a suggestion of having a *reset* option since it could be difficult to remove previous search criteria in cases where users have performed searches earlier.

As graphically presented in the *Figure 6.4*, the prototype scored 10 out 0f 10 from 69% of the thirteen experts. It scored 9 from 15 %; and 8 from 8% of the participating experts. The average score for this dimension was 9.4 out of 10.

H5: Error prevention

Table 6. 6: Evaluation comments for H5 – Error prevention

H5: Error prevention	
Good design aspects	-No error occurred, No errors at all -Good, not many errors that occurred. -I didn't encounter any errors, so I can only assume that it is good.
Poor design aspects	None
Suggestions	-Maybe there should be some red star next to forms that have to be filled in to get some information.

(a)

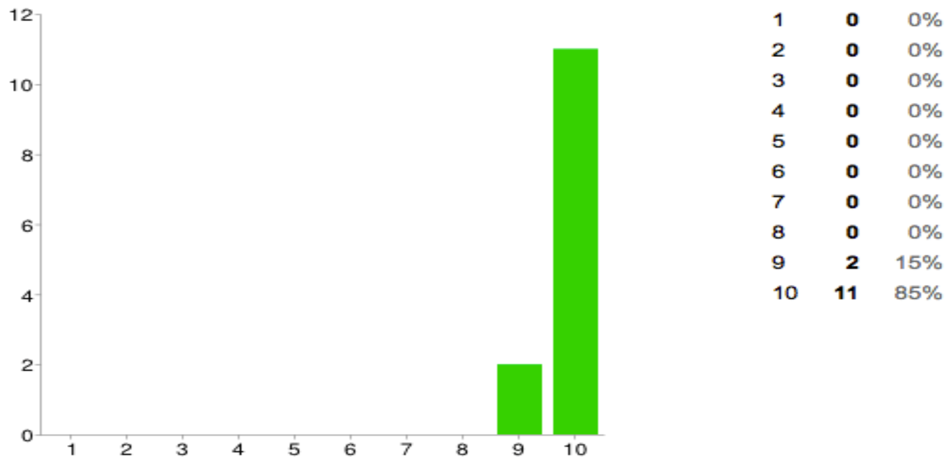


Figure 6. 5: H5 - Error prevention

(b)

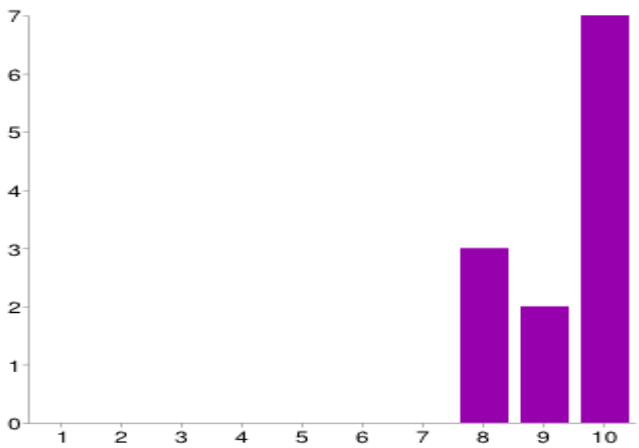
None of the experts encountered any errors while exploring the features of the prototype. This means that all the tasks were executed smoothly. However, the evaluation comment or this dimension shown in the *Table 6.6* shows that there was one suggestion to use a red star to mark the forms that are to be completed. From the *Figure 6.5,* it looks like the prototype scored 10 out of 10 from 85% of the thirteen expert evaluators and scored 9 from 15%. The average score as presented in the *Table 6.1,* was 9.8 out of 10.

H6: Recognition rather than recall

Table 6. 7: Evaluation comments for H6 – Recognition rather than recall

H6: Recognition rather than recall	
Good design aspects	None
Poor design aspects	-Difficult to locate clinical trials, but it was a very interesting feature when i finally located it. -Generate report and Dashboard and clinical trials was difficult to locate -I think the clinical trials were not easy to find, also the dashboard.
Suggestions	-Make them tabs for Generate report and Dashboard and clinical trials.

(a)



1	0	0%
2	0	0%
3	0	0%
4	0	0%
5	0	0%
6	0	0%
7	0	0%
8	3	25%
9	2	17%
10	7	58%

Figure 6. 6: H6 - Recognition rather than recall

(b)

According to the evaluation results as summarised in the *Table 6.7,* some experts found it difficult to locate c*linical trials, generate report* and the *dashboard* functions. One noted that the clinical trials function was an interesting feature once located.

The suggestion for improving this was to create tabs for each of these functions where they could be easily visible to the users.

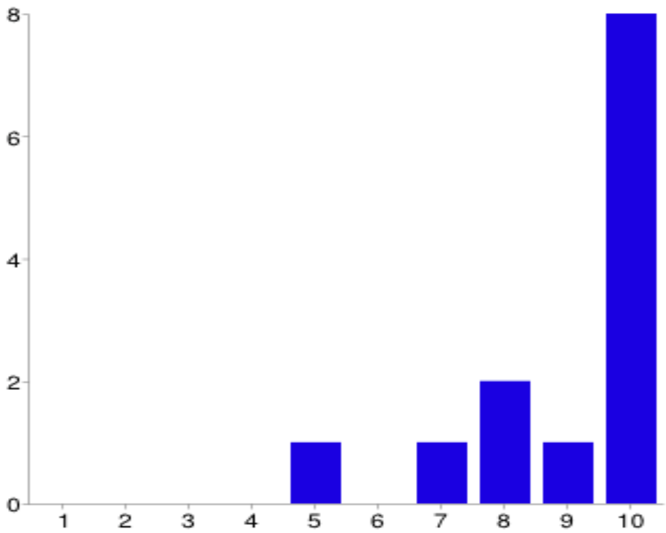
Figure 6.6, displays the distribution of the prototype's scores against recognition rather than recall dimension, and it shows that one of the thirteen experts did not rate the prototype against this dimension. In spite of that, the prototype scored 10 out of 10 from 58% of the twelve participating experts; scored 9 from 17%; and scored 8 from 25% of the twelve evaluators. The average score was 9.3 out of 10.

H7: Flexibility and efficiency of use

Table 6. 8: Evaluation comments for H7 – Flexibility and efficiency use

H7: Flexibility and efficiency of use	
Good design aspects	-When I've done the tasks more then once I easily learned the way of doing it. -If I had done different tasks one more time i think i would be going pretty much faster. So good!
Poor design aspects	-The slowness of the system made it feel a little inefficient in some parts. -There was a search function hidden deep in a menu that could be easier to find to make experts get to that function quicker.
Suggestions	None

(a)



1	0	0%
2	0	0%
3	0	0%
4	0	0%
5	1	8%
6	0	0%
7	1	8%
8	2	15%
9	1	8%
10	8	62%

Figure 6. 7: H7 - Flexibility and efficiency of use

(b)

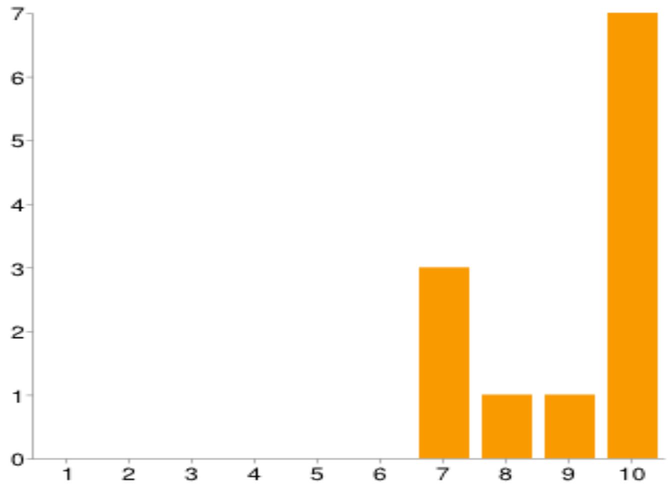
Following comments regarding this dimension, as presented in the *Table 6.8,* there were two poor design aspects pointed out. One of the search functions was not easy to locate and therefore hindered the expert to access the function quickly. One of the experts experienced the prototype as being slow, something that made it seem inefficient at times. These could explain the score differences given by the individual experts as shown in the *Figure 6.7*.
On the other hand, results also show that executing more tasks would facilitate fast learning of how to use the prototype, which in turn would lead to fast and efficient task execution.
The scores presented on the *Figure 6.7,* show that the prototype scored 10 out of 10 for this dimension from 62% of the thirteen experts; 9 from 8%; 8 from 15%; 7 from 8%; and 5 from 8%, respectively. The average score for flexibility and efficiency of use was 9 out of 10.

H8: Aesthetic and minimalist design

Table 6. 9: Evaluation comments for H8 – Aesthetic and minimalist design

H8: Aesthetic and minimalist design	
Good design aspects	-Simple and beautiful design -Very clean design. It made me find the data i wanted to see very fast. -Very minimalist and nice design. -A very sleek and simple design.
Poor design aspects	-Very much information on some pages, but probably needed for those tasks. -I miss clearer visibility of which options are currently selected. for instance, the selected tab is white and the deselected tabs are light grey. It is not that easy to see the difference quickly.
Suggestions	-Nice design, but it could be larger in height and width.

(a)



1	0	0%
2	0	0%
3	0	0%
4	0	0%
5	0	0%
6	0	0%
7	3	25%
8	1	8%
9	1	8%
10	7	58%

Figure 6. 8: H8 - Aesthetic and minimalist design

(b)

From the summarised experts' comments presented in the *Table 6.9* the prototype seemed to have caught some experts' eyes as they were generous in their positive comments expressing how clean, simple, nice, minimalist, sleek, and beautiful the interface design was. They also mentioned that the clean design made it fast to find what they were looking for.

Some poor design aspects were also pointed out. One of them was that some prototype pages were experienced to contain too much information. The comment was followed by the expert's '*possible*' reason for this; that it was probably needed for the some tasks. Another expert pointed out the need for showing the users clearly which options or tabs are selected. The expert implied that this option was available, but it was not emphasised enough. There was a suggestion to increase the heights and widths of the prototype's pages.

One expert did not rate the prototype against this heuristic dimension. As shown in the *Figure 6.8,* the prototype scored 10 out of 10 from 58% of the twelve experts evaluators; 9 from 8%; 8 from 8%; and 7 from 25% of the twelve expert evaluators. The average score was 9 out of 10.

H9: Help users recognize, diagnose, and recover from errors

Table 6. 10: Evaluation comments for H9 – Help users recognise, diagnose, and recover from errors

H9: Help users recognize, diagnose, and recover from errors	
Good design aspects	-Didn't encounter any errors -Did not experience any errors
Poor design aspects	None
Suggestions	None

(a)

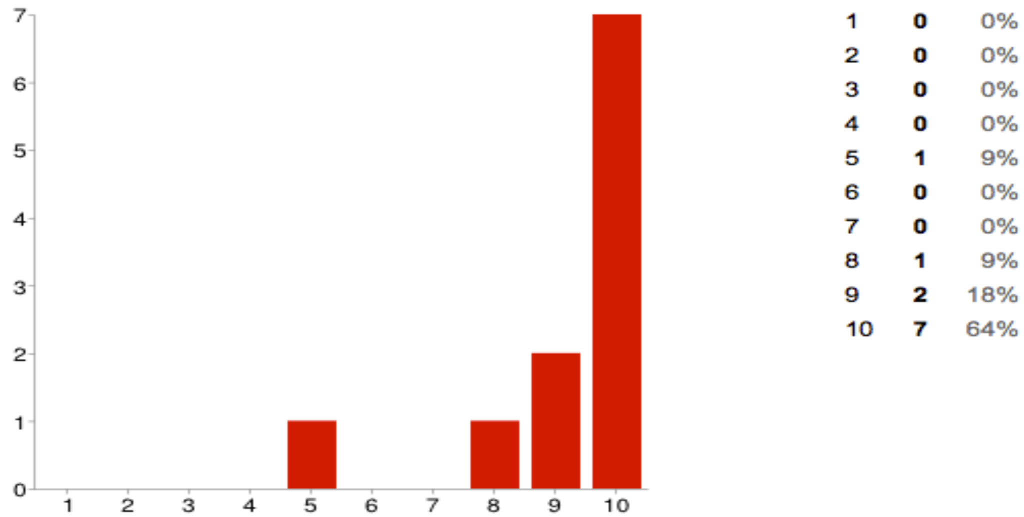


Figure 6. 9: H9 - Help users recognize, diagnose, and recover from errors

(b)

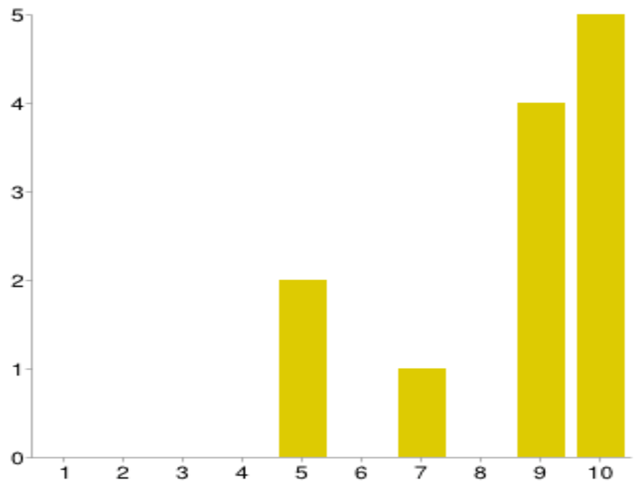
None of the experts experienced errors when performing the tasks as shown on the comments summary in the *Table 6.10.* It was therefore not possible for them to see how the error messages would have been presented to them. Nevertheless, *Figure 6.9* show that some experts still rated the prototype against this dimension, which lead to an average score of 9.1 out of 10. Two out of the thirteen experts did not provide any ratings.

H10: Help and documentation

Table 6. 11: Evaluation comments for H10 – Help and documentation

H10: Help and documentation	
Good design aspects	-It was clear how to use it, so did not need documentation -I don't think extra documentation is needed if the user gets a proper introduction to the system and knows the specific terms for medical stuff. -The documentation and way to understand different codes is much easier in this system than in the old PDF document.
Poor design aspects	-Could not find help and documentation
Suggestions	-I miss having tooltips pop up describing the element when i hover over a button or another control.

(a)



1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**2**	17%
6	**0**	0%
7	**1**	8%
8	**0**	0%
9	**4**	33%
10	**5**	42%

Figure 6. 10: H10 - Help and documentation

(b)

The evaluation results as summarised in the *Table 6.11* show that some of the experts thought the prototype's interface was clear enough and therefore did not necessarily need to have the documentation. In addition to this, some experts though a proper introduction on how to use the prototype would be enough. Therefore, extra documentation would then not be needed.

One expert noted that the documentation was lacking.
A suggestion of providing tooltip pop ups or hover effects with brief descriptions of things was also noted in the form of a function that the expert lacked on the interface.
One expert did not rate the prototype against this dimension. *Figure 6.10* show that the prototype scored 10 from 42% of the twelve evaluators; scored 9 from 33%; 7 from 8%; and 5 from 17% of the twelve experts. The average score was 8.5 out of 10.

6.1.2 Time Dimension

With regard to time, the first five tasks of the evaluation were timed as they were being executed following the current practises and through using the designed prototype. Results shown in the *Appendix 6d,* suggest that there was a relatively big time difference between these two ways of executing the same tasks. As shown in the *Table 6.12*, it took an average of 06:10 minutes to complete five tasks when following the currently available practises (current ways) and it took an average of 02:11 minutes to complete the same tasks when using the designed prototype (the suggested ways).

Table 6. 12: Average time taken by experts to perform 5 tasks during the evaluation

	Current ways	Suggested ways
Average time (in minutes)	**06:10 min**	**02:11 min**

The results suggest that users managed to complete all the tasks without any significant errors and were comfortable with interacting with the prototype, which seemed to have proven to be faster in task execution than what the current practises offer.

6.2 Discussion

This sub-section discusses several design case studies through which the developed artefact was used to answer the research questions. By presenting sample problems and their solutions, these cases show how the mined data is currently presented in real life in comparison to how it could be improved in an efficient and user-friendly manner. In addition to presenting the design study cases, a reflection of the methodologies used, the design process, prototypes reached, the evaluation processes and the findings are also discussed.

6.2.1 Design Case Studies

6.2.1.1 Design Case Study 1

***Problem**:* Finding the number of males and females of all age groups who were newly diagnosed with breast cancer in 2010 in the county of Skåne.

***Solution 1**:* The current solution for doing such a straightforward task involves the following steps; looking through several report pages in search of the categorized statistics then when the appropriate page is found; searching through dense lists with many cancer types in order to find the intended type; then browsing through the rows to find the right number that matches the required region. And finally, the search result is then found in the figures at the intersection between the column where breast cancer meets the column numbered *'12'*, which represents the county of Skåne It is clear that handling data presented this way is a tedious and demanding task for any user and it rather discourage users to do much data exploration. *Figure 6.11* shows how the results to this problem are currently presented to users.

Counties				
01 Stockholm	06 Jönköping	12 Skåne	19 Västmanland	24 Västerbotten
03 Uppsala	07 Kronoberg	13 Halland	20 Dalarna	25 Norrbotten
04 Södermanland	08 Kalmar	14 Västra Götaland	21 Gävleborg	
05 Östergötland	09 Gotland	17 Värmland	22 Västernorrland	
	10 Blekinge	18 Örebro	23 Jämtland	

				Counties																				
Icd7	Sites	Sex	Sweden	01	03	04	05	06	07	08	09	10	12	13	14	17	18	19	20	21	22	23	24	25
160	Nose and nasal sinuses	M	41	13	1	4	4	2	-	1	-	-	4	1	3	1	1	-	1	-	2	-	2	1
		F	25	8	1	-	1	-	-	-	-	-	3	1	7	2	-	-	-	-	-	-	1	1
161	Larynx	M	143	31	2	4	6	7	3	7	-	2	28	3	23	3	3	4	3	2	6	2	3	1
		F	37	7	1	-	3	1	-	-	1	-	11	3	6	1	1	-	1	-	-	-	-	1
162	Trachea, bronchus, lung and	M	1 942	398	72	55	101	57	28	63	14	34	308	69	298	56	75	57	54	83	22	20	28	50
	pleura, primary	F	1 755	377	59	64	91	34	26	34	11	47	309	57	246	59	60	65	55	58	22	22	19	40
163	Lung, not specified as primary	M	4	-	-	-	1	-	-	-	-	-	1	-	2	-	-	-	-	-	-	-	-	-
		F	14	-	-	-	1	-	-	-	-	-	5	2	6	-	-	-	-	-	-	-	-	-
164	Mediastinum	M	5	4	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-
		F	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
170	Breast	M	33	6	1	-	5	1	1	-	-	-	3	-	4	1	-	2	-	3	3	1	1	1
		F	7 917	1 600	272	214	375	326	154	207	39	153	1 262	205	1 271	231	225	215	211	243	184	124	226	180
171	Cervix uteri	F	428	100	23	10	15	12	7	8	1	11	65	9	65	11	12	11	12	14	8	12	14	8
172	Corpus uteri	F	1 351	289	44	45	52	71	32	36	6	21	139	45	217	46	42	45	50	30	47	20	43	31
173	Chorionepithelioma of uterus	F	6	2	1	1	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-
174	Uterus, part unspecified	F	97	7	5	2	2	3	4	4	-	4	16	1	24	3	2	3	3	3	2	5	4	-
175	Ovary, tube and broad ligament	F	749	144	21	29	39	32	11	21	5	10	90	32	114	19	31	25	24	28	20	12	27	15
176	Other female genital organs	F	223	36	6	6	9	12	6	10	2	4	41	11	29	6	-	5	8	8	11	6	4	3
177	Prostate	M	9 697	1 828	303	298	602	382	257	242	46	192	1 140	346	1 714	253	260	280	275	259	267	162	323	268
178	Testis	M	297	48	10	7	12	15	8	5	4	1	44	10	62	8	11	8	8	7	9	9	9	2
179	Other male genital organs	M	73	10	1	6	4	3	1	-	1	2	6	3	20	1	1	3	2	5	-	-	4	-
180	Kidney	M	680	116	19	22	40	25	27	13	3	10	99	22	104	29	13	11	21	26	23	14	29	14
		F	406	63	9	17	18	20	16	8	-	3	59	6	63	14	21	11	19	14	11	14	14	6
181	Urinary organs	M	1 855	276	42	57	85	68	48	52	11	42	289	55	335	71	79	54	57	57	54	21	54	48
		F	673	121	10	17	25	24	10	23	2	18	106	26	127	21	29	16	16	24	20	14	12	12
190	Malignant melanoma of skin	M	1 466	302	40	54	101	44	20	50	6	20	207	52	274	53	47	40	38	37	21	12	17	31
		F	1 351	267	34	46	101	51	23	26	7	25	188	50	274	31	48	33	32	31	28	8	26	22
191	Skin (melanoma excluded)	M	2 912	585	61	77	183	107	75	84	16	78	501	202	485	107	73	37	74	78	20	14	30	25
		F	2 113	406	44	42	122	91	56	56	12	49	344	179	375	82	47	33	31	50	12	15	37	30
192	Eve	M	74	12	3	2	3	2	2	1	2	-	9	6	9	3	2	-	7	3	5	1	-	2

Figure 6. 11: Solution1 - Example of the current way of presenting the results (Ericsson et al., 2011)

Solution2: The interface presented in the developed artefact simplifies the task by allowing users select the variables and click the 'GO' button so that they can view the results by hovering the mouse over the region they want on the map. Shown in the *Figure 6.12* is the practical solution to this problem presenting the results on the map after the user has specified the variables *breast cancer*; *all ages*; year *2010* and *both genders* from the top section of the page marked with circles before hovered the mouse over Skåne county. With this particular example, users can for instance, be encouraged to view the statistics of the occurrence of the same type of cancer in the other counties since it effortlessly involves moving the mouse over whichever county they want.

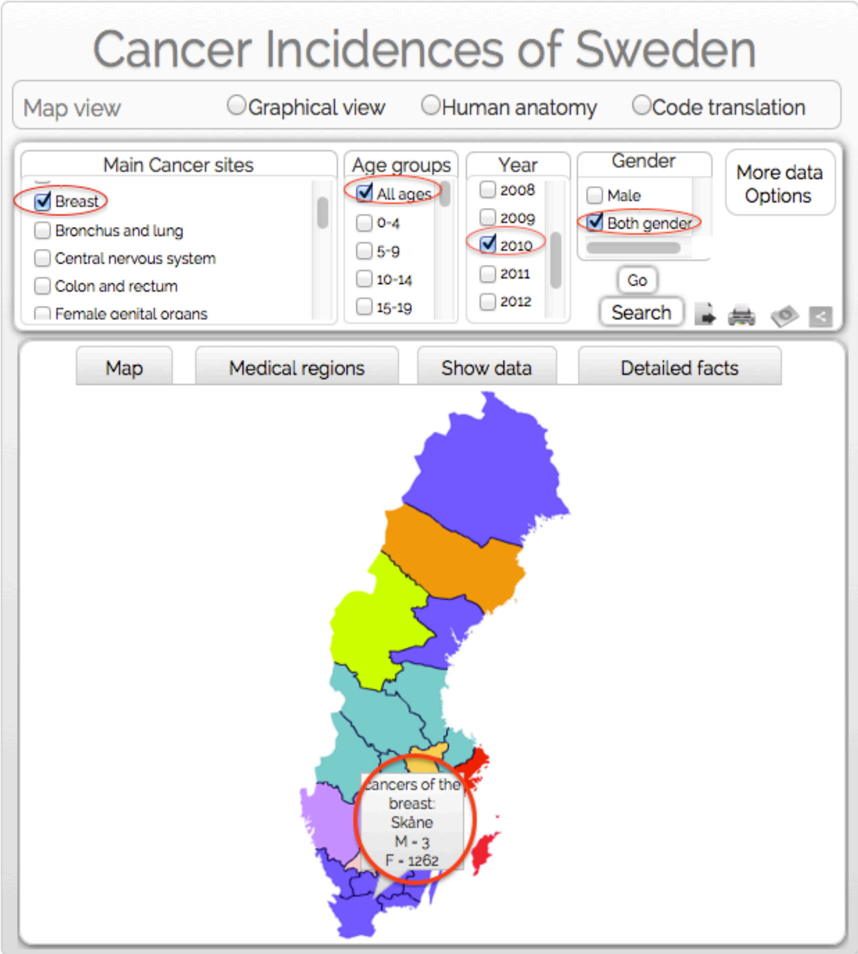


Figure 6. 12: Solution2-the suggested way of presenting the results combining map and statistics

6.2.1.2 Design Case Study 2

Problem: Finding the number of occurrences of cancers of the bone for both genders living in Blekinge in 2010 at the same time, seeing how the same type of cancer has been diagnosed throughout the country in patients aged between 2 – 84 years. In addition, finding how many autopsied cases in the course of death register of 2009 had cancers of the Stomach; Colon and rectum; Pancreas; Bronchus and lung; and Prostate.

Solution 1*:* For such a complicated task, the current solution executes each of these sub-tasks separately following the same steps as described in design case study one above for the first and last part of the given problem, only that this time the steps have to be repeated for each sub-tasks. For the second part of the problem, *'seeing how the same type of cancer has been diagnosed throughout the country in patients aged between 2 – 84 years',* users need to sum up all the individual types of the bone cancers across the country according to the ages at the time of diagnosis in addition to following the steps described in the design case study ***one***. The resulting solution to this problem is seven separate pages containing solutions to each of the sub-tasks.

Executing tasks in this manner ends up in using too much time going through the bulk data, something that is not very pleasant for any user. Extracting information from performing this task may lead to information loss, as the user is not provided with possibilities to look closely at this data from different angles. In addition, it is not very easy for users to have an overview of the variations of, for instance, the cancers of the bones amongst the different age groups as pointed out in this particular problem.

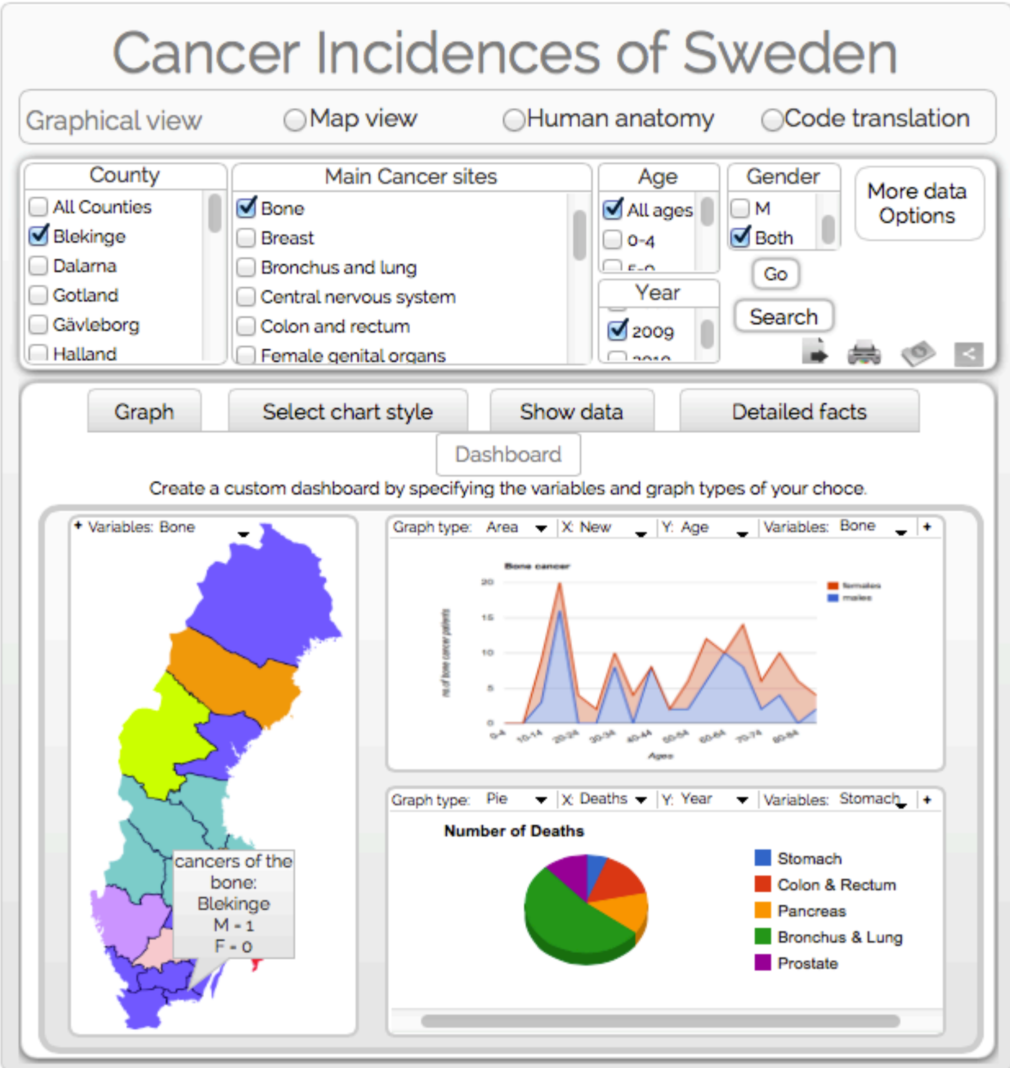


Figure 6. 13: Dashboard solution to design case study 2

Solution 2: Even though this task may seem complicated, the designed prototype provides users with an easy to use dashboard function with possibilities to view all the data from all the sub-tasks on one screen. *Figure 6.13* shows a solution to the stated problem on a dashboard with interactive features allowing users to directly interact with the system. This gives users room to extract hidden information that could not be easily detected otherwise. For example, with this solution, users are able to gain an overview of the different age groups that were diagnosed, in this case, with bone cancer from the area graph on the dashboard.

6.2.1.3 Design Case Study 3

Problem: Finding the trends of the occurrences of cancers of the respiratory tract for the last 10 years, then preparing a custom report for a list of all the new cases of this particular cancer type recorded from medical region 6 -Umeå in 2010. Thereafter, finding recent study results of the related clinical studies for these cancer types that were carried out in the nearest clinical trials centre. And finally, finding and printing an image of bone cancer cells that you came across earlier today and would like to explore it later.

***Solution 1**:* Currently, users are not able to solve this problem without exiting the system and seeking help from external resources such as the Internet or dedicated research stuff. Solving this kind of problem involves browsing through several sites in order to collect pieces of information that would eventually answer all the questions of the given problem. The main issues with this solution are that: it is time consuming, insufficient and the results produced may or may not be from a reliable source.

Solution 2: The designed prototype provides several new features that utilize the available and reliable oncological data resources in order to supply users with the necessary information available from the knowledge base. The new features suggested in this thesis project demonstrate the possibility of solving problems such as this, in an effective and user-friendly manner without exiting the system. The five features which were considered during the design process that became useful towards solving this problem included:
Firstly: to provide the possibility for users to be able to have a broad overview of the trends of the different types of cancers. This would utilize the same data stored in these registries, only

that it would allow users to explore the individual cancer types in details. For example, not only for visualizing trends but also for providing a detailed summary of each specific kind of cancer instead of providing a generalized summary of cancers that belong to the same category, for instance, cancers of the head and neck. Through the '*detailed facts'* page provided on the interface shown in the *Figure 6.14*, the designed prototype suggests a possibility to obtain more information about the different kinds of cancers without having to exit the system. In addition, this provides a solution to the first part of this problem where users can also view trends of different cancer types, the circled area.

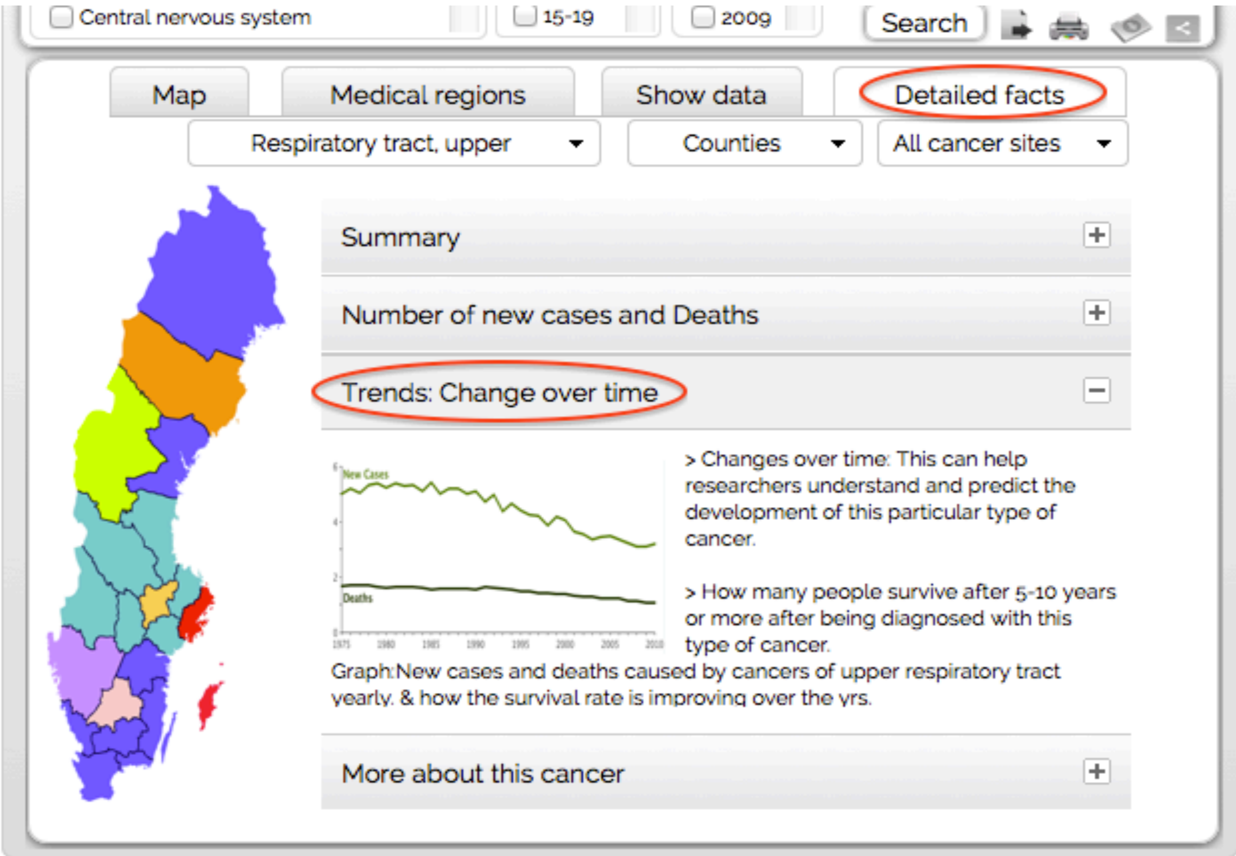


Figure 6. 14: Possibility to view trends of cancer types

Secondly: to provide users with the possibility to generate custom reports of the individual cancer types in addition to generating the yearly reports. Such a possibility would open for more detailed reports that users can more benefit from than from generalized annual reports currently provided as the only option. *Figure 6.15* shows the simple suggested way through which users can generate custom reports in addition to the yearly reports. This option is also accessible from all pages under the '*More data options*' marked with a circle on same figure- *Figure 6.15* This function can be especially useful for clinicians and medical students. This is because these two user groups are the ones most likely to be generating short reports due to the nature of the tasks to be carried out in their work environments.

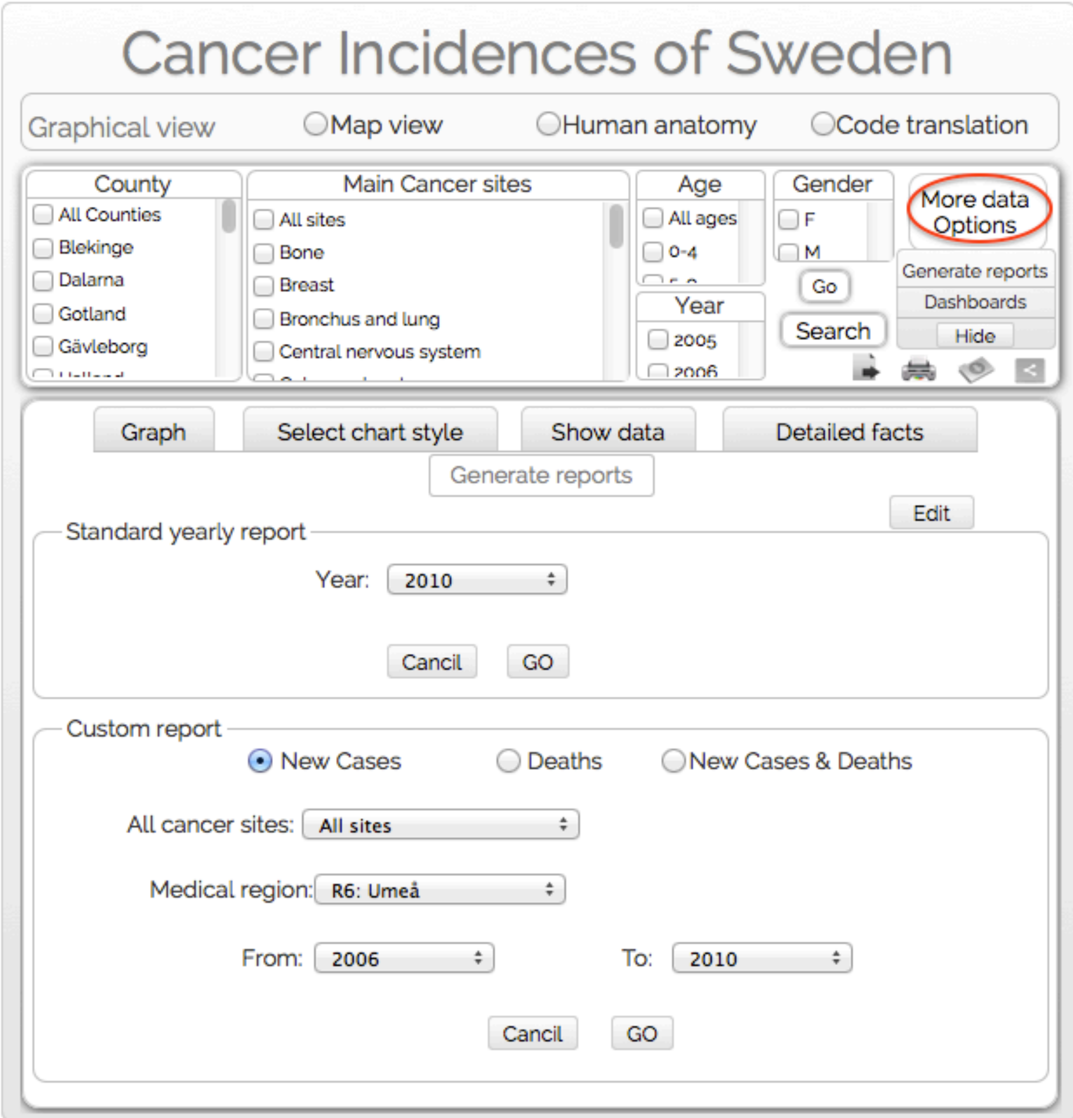


Figure 6. 15: Possibility to generate custom reports

Thirdly: to provide users with the possibility to connect to reliable clinical studies related to the types of cancers users are presented with, using the cancer registries. Clinical trials were selected because of their importance to the development of new treatments of cancers and the nature of the search pattern of the typical users. *Figure 6.16* shows the possibility for users to specify their clinical trials searches. *Figure 6.17,* shows the results of the trial's search pointing out the circled area as the study results that are part of the solution to the above given problem. The sample data used in the prototype to demonstrate this possibility was obtained from clinicaltrials.gov, which is a reliable data resource for clinical trials.

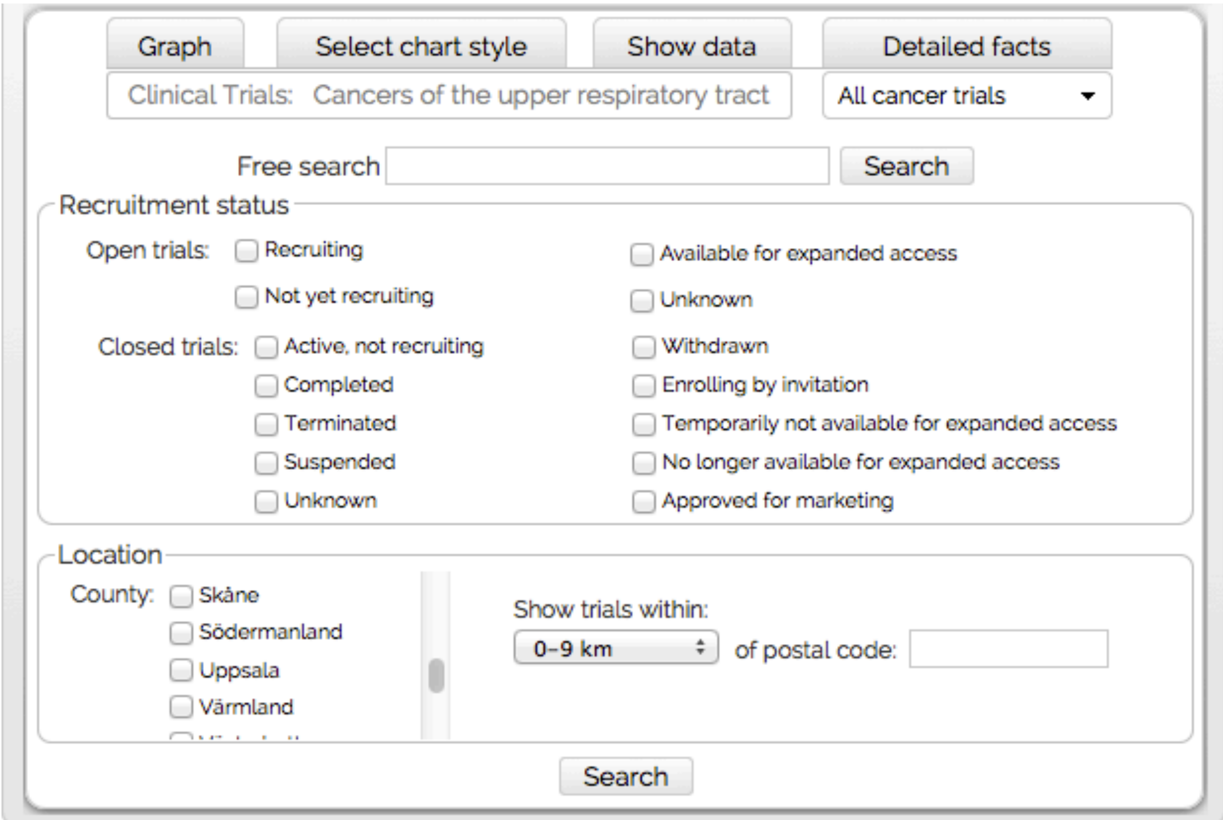


Figure 6. 16: Retrieving information about clinical trials

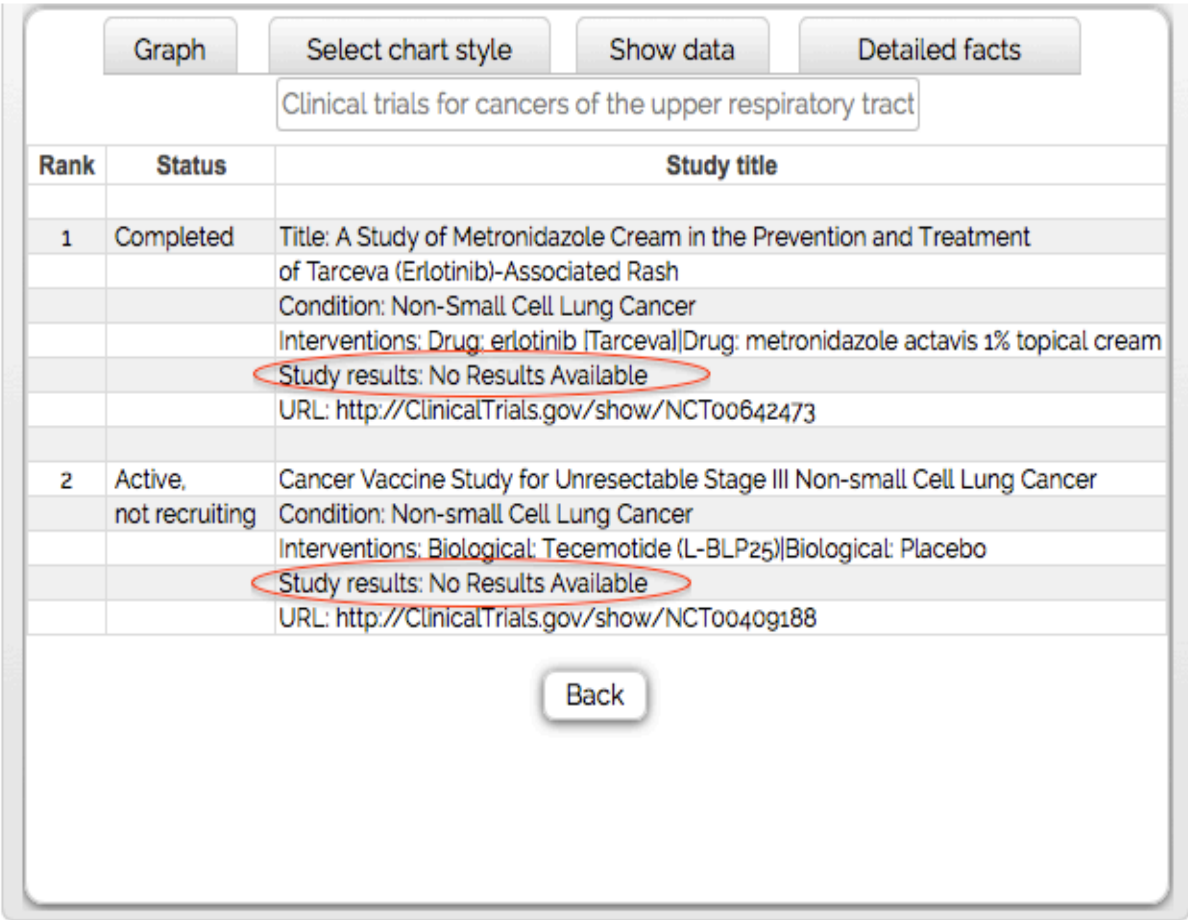


Figure 6. 17: Clinical trials search

Fourthly: to provide the possibility for users to utilize the available image storages containing images of the different cancer cells in order to study the cells closely. *Figure 6.18* shows a page with images of the bone cancer cells as requested in the problem above. This solution is obtained from the '*show image* 'function, which is easily accessible on the prototype, This function can be interesting to all the three user groups, namely, clinicians, medical students and the public.

Clinicians can benefit from this function by using the images to study the changes of these cells in the different research projects. In an e-learning setting it can be used to broaden the students' understanding and perception of the appearances of these cells. The public can, for instance, out of curiosity, have a chance to see how colourful and interesting these deadly cells can be.

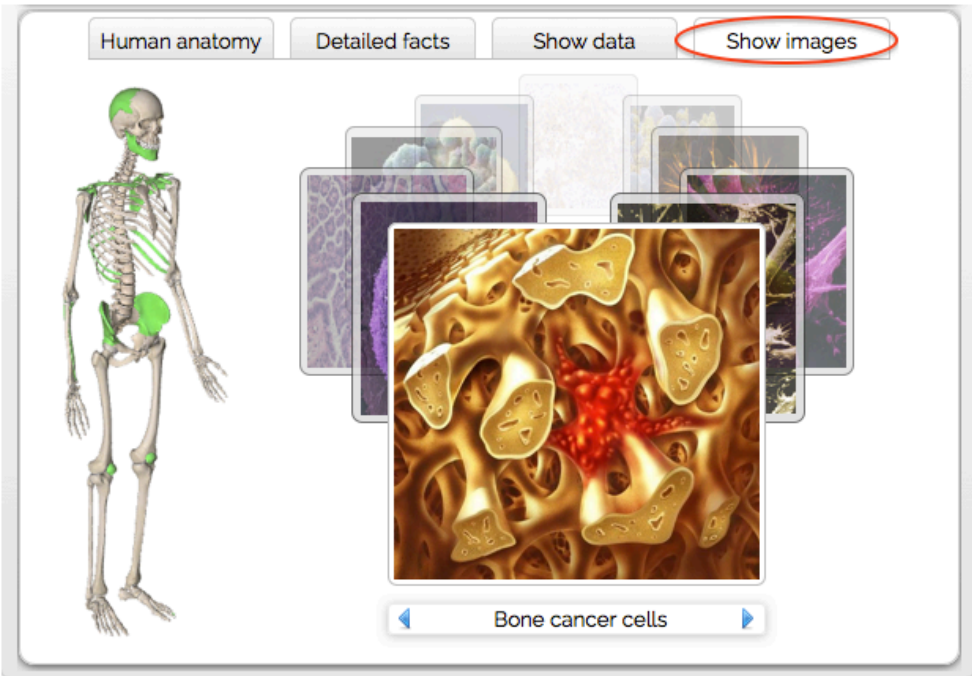


Figure 6. 18: Possibility to view images from the image base.

Fifthly: to provide users with the possibility to print the different parts of the reports as they opt to without being limited to only printing the whole 100-paged yearly report. As shown in the Figure 6.19, this possibility together with the export, download and share are easily accessible for users from the top right corner of the prototype's interfaces.

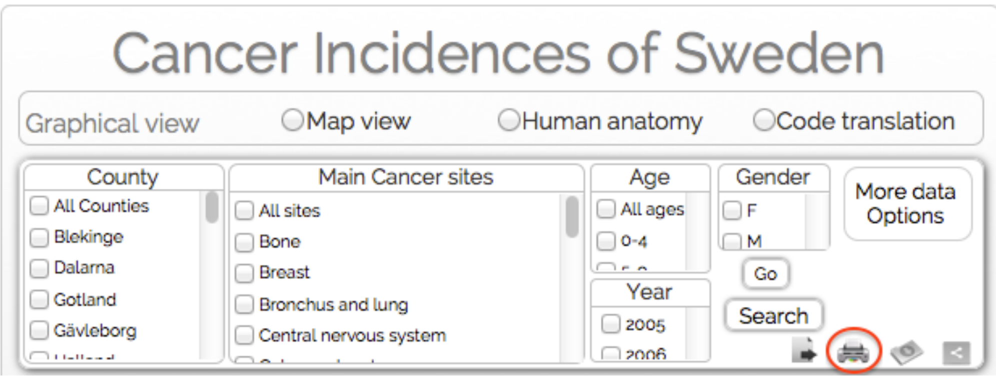


Figure 6. 19: Printing possibility

6.2.2 How the Research Questions were Answered

6.2.2.1 Research Question 1

How can Human Computer Interaction design make interfaces for data visualisation in data mining more efficient and user friendly?

This research question focused on how existing knowledge from Human Computer Interaction design and Information System fields could be utilized in order to create a novel interactive solution for visualising data mined from the cancer registries. In order to answer this question, a prototype was developed to address the challenges faced by the current practises. This was done through three iterations, explored in the *Chapter Five*. Users tested the initial prototypes for usability and experts evaluated the final prototype. The evaluation results suggest that the designed artefact is efficient and user-friendly. This was also demonstrated through all the three design case studies presented in the *Section 6.2.1*. These cases were used to present the real life situations and covering some common tasks and the ways in which they are currently executed, to be compared to the solutions designed by executing the same tasks. Through these cases, it is clear that the application of Human Computer Interaction design knowledge has enhanced the prototype's interfaces and led to interactive, efficient and user-friendly data visualisation for data mined from the cancer registries.

6.2.2.2 Research Question 2

How can the research of this thesis help transform the current state of presenting data from the cancer registries to a better and preferred state?

Answering this research question began by understanding how data mining from the cancer registries is currently conducted, and by exploring the current practises involved in presenting the mined data to users. Thereafter, the question was answered by suggesting ways to improve the current state of presenting data from these registries by designing an interactive artefact through which users could have direct interaction with the system. The designed artefact provided users with several possibilities to deal with the mined data that were currently not an option. It, for example, contains new and enhanced features allowing users to handle the data in a user-friendly manner through the easy to use interfaces. All these features provided in the designed artefact are likely to open potential users' understanding of IT solutions and letting them experience this thesis through seeing the design outcomes as demonstrated through the design case studies, which present a description of how tasks are executed currently and with the help of the designed artefact.

The Design case study 1 demonstrates how simple it is to carry out tasks when using the designed artefact as compared to the current practises, which involve more stages and is stressful for users to follow. Through the Design case study 2, the designed artefact was used to present the possibility for users to have a broad overview of the data they search for and compare the search results on a dashboard. This is something that is currently not an option and users are left to view the data in bulk formats making it impossible to visualise and extract meaning from it. The Design case study 3 explores more of the new functionalities that are not currently offered. For example the possibility to find detailed information of the searched cancer type; generating custom reports; access related clinical trials; access cancer cells image storage and print possibility that are flexible to users' choices.

Therefore, this thesis provide users with what they want without them knowing what exactly they wanted before they were presented with the artefact.

6.2.2.3 Research Question 3

How can smart user interfaces enable good data visualisation for different users groups?

Even though a good amount of time was spent in understanding the data mining processes involved in the cancer registries, the goal was to provide good, smart and user-friendly interfaces for the target user groups. Therefore, by understanding the technical background of the data presented to users from these data resources, it was easy to suggest design solutions, which were technically feasible for implementation. The resulting interactive prototypes contained good interfaces, which were easy to use for users with different levels of IT background. With such interfaces, users do not have to worry about defining queries or understanding the technicalities of the underlying files and methods; but instead, they can concentrate only on the interface design.

In order to answer this research question, a prototype was developed with the different visualisation modes, as explored in the *Section 5.4.2*, and allow the different user groups to decide their own level of search according to the nature of the tasks they wanted to execute. The designed prototype caters for experts such as the clinicians who will have more predefined ideas especially connected to some research works; medical students who will need more defined e-learning tasks to make sense of the data; and the public who will need informative answers to simple and direct answers, which could be part of the reports. Proven through the different tasks carried out during the evaluation and those presented in the Design case studies in the Section 6.2.1, the prototype's user interfaces, could be used by the different user groups (which tailor) by tailoring their searches in order to execute tasks at hand.

6.2.2.4 Research Question 4

Will this kind of research and interfaces help to change the way users look at the registry data?

This kind of registry data has always been presented to users in ways that are not very attractive to extract information from and that are time consuming. With the designed prototype in addition to the theoretical foundations of this thesis, it is very likely that this kind of research and the explored interfaces will change the way users look at the registry data. This is through the user-friendly interfaces that could encourage users to utilize the information found in these registries by effective task execution and the use of the newly provided functionalities to explore the cancer registry data, something that is currently not very easy to do.

The average time difference for performing a set of tasks has shown how users could save time by using the designed prototype when performing the same tasks, which took longer to perform when following the current practises. The simplicity of the features and functionalities explored in all the Design case studies prove the likeliness for users to perceive the registry data in a different way than they do currently. The evaluation results in the Section 6.1 point out the favourable differences in the way that solutions suggested in the thesis could be appreciated. This is demonstrated on the differences shown between the current and the suggested ways of tackling the same tasks.

The suggested solution focuses on the different visual presentations of data that will enhance information extraction from the mined data. Large screens, such as the PC screens, are the main presentational frames. By using the right IT technologies, it is possible to apply the same visualisation concept presented in the designed prototype, so as to accommodate visualisation of the same kind of data in small devices. This is also expected to change the way users will look at the registry data, especially because small devices such as mobile phones and tablets are increasingly becoming popular in clinical practices.

6.2.3 Research Methodologies Used

This thesis was conducted through following the Design science methodology as described by Hevner et al. (2004) and produced an interactive prototype, which was evaluated by expert. The design guidelines presented by Hevner et al. (2004) were followed in order to accomplish designing a meaningful artefact and one that would be useful in providing answers to the research questions. This research methodology fitted well for this research thesis since it focused on developing an innovative solution in order to solve an existing problem regarding the current practises of presenting mined data from the cancer registries.

It also had several stages of the iteration and all influenced by the users. However users were not directly involved in the initial stages of the design process, instead, genius design also known as the innovative design approach was applied in order to present users with tentative design solutions. This approach was chosen because it was necessary to present users with something to compare with what they currently have in order for them to be able to suggest what they thought would have been a good solution. Approaching users in this manner seemed to have worked well as it became easy for them to come up with some useful suggestions after seeing the design possibilities. Presenting the tentative design solutions made them realise what they really wanted in order to improve the current practises, something that was difficult to envision without having an overview of what the current technology can offer. This is also proving the research hypothesis raised by Beck (1999) where users did know what they really wanted until they were shown what they could have.

The QOC method (MacLean et al., 1989) was used for design space analysis when the designing features of the prototypes' interfaces. A set of questions, options and criteria were created in order to address and document the design problems, their alternative design solutions and the justification for the solutions. This method was useful since it was of great assistance when making design decisions about the interfaces' appearances as it provides an overview of the design choices made and reasoning for selected choices. This method was applicable in the first and third iterations and the QOC set used are documented in the *Appendix 2a and 2b*.

6.2.4 Design Patterns and Challenges

The design process was carried out through a combination of iterative and parallel design processes. Personal Kanban (Benson & Barry, 2011) was used to managed all the tasks and activities involved during the design process. There were three iterations in total and the parallel design approach was followed in order to produce alternative prototypes for users to test.

The first iteration resulted in three parallel hand-drawn then electronic sketches of the initial design solutions. This iteration focused on the top interface features and the prototype sketches were not interactive at this stage yet.

The second iteration resulted in three interactive prototypes. Activities involved in this iteration were adding some interaction to the prototypes and performing usability testing with users. Feedback obtained from users was used to create a final prototype.

The third iteration resulted in the final prototype, which was evaluated by experts. During this iteration the three parallel prototypes from the second iteration were merged into one prototype containing the best ideas from the parallel prototypes and the new features suggested by uses. The resulting prototype is presented in the *Appendix 5*. *Figure 6.20* shows a summary of the main design achievements for each iteration.

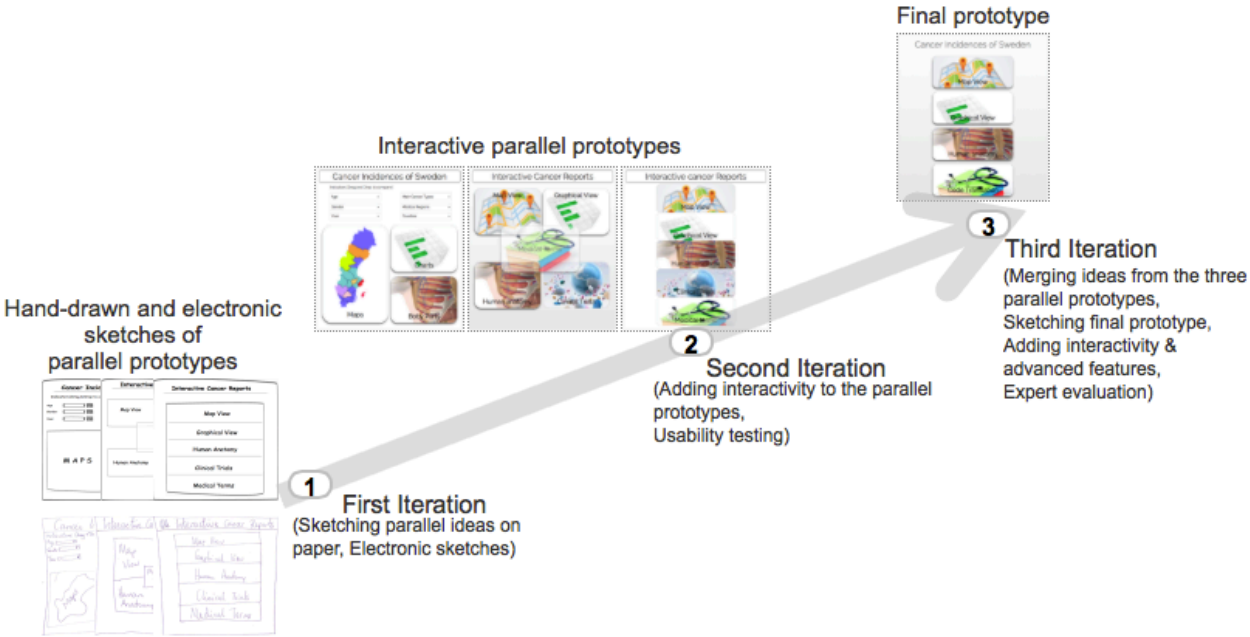


Figure 6. 20: Main design achievements for each iteration

The iterative process helped to enhance the designed prototype by adding some improvements to the interfaces during each iteration. It was simple to follow and flexible to accommodate as many iterations as time allowed.

The parallel design process was useful in providing alternative design solutions, which were helpful in improving the quality of the feedback from users during the testing. Parallel design was easy to adapt and combine with the iterative approach, and it did not require a complete design of all features and pages, but only a few enough to convey the concept of the design idea. Of the limitations, parallel design ended up being time consuming since three parallel prototypes were created during the first and second iterations and then tested with users in the second iteration.

Following Personal Kanban was easy and the approach was useful in providing an overview of the tasks, hence simplified keeping track of what was completed and what was still waiting to be completed. With the WIP limit, earlier mentioned in the Section 5.1, set to five tasks, it encouraged doing the right prioritization and was easy to avoid having too many tasks started half way, something that improved productivity. The Pomodoros were disruptive at the beginning, but after a couple of them, it was easy to fall into the rhythm.

The main design challenge was coming up with a novel design solution for an unfamiliar area, the cancer registries. This was eased after gaining an understanding of the kinds of data involved and the data mining activities from the cancer registries. This knowledge was helpful in deciding what features were technically feasible when dealing with this kind of data. Even though decisions regarding the design features to be included and their positioning was not a simple task, the applied methods and approaches, which were discussed in the *Chapters Three and Five,* provided a lead to addressing such challenges.

6.2.5 Evaluations and Findings

Usability testing with five users was carried out in the second iteration. During this testing users assessed the three parallel prototypes for usability and provided their design input in form of comments. This was through a questionnaire containing rating statements and open questions. The aspects assessed included the prototypes' ease of use, effectiveness, efficiency, utility, learnability, memorability, error prevention and satisfaction. This kind of feedback from users was very useful for the development of the final prototype, especially since innovative design approach was followed at the very beginning. The summary of the resulting feedback from the usability testing is obtained in the *Table 3* of the *Section 5.3.4*. The testing results showed that none of the three parallel prototypes tested were perfect but they all contained some good design elements that were worth reusing in the final prototype. They also presented the users' useful ideas, which were not initially included in any of the parallel prototypes.

Expert evaluation was performed on the final prototype during the third iteration. At this stage, the prototype's utility, quality and efficacy were evaluated. It was important to evaluate this prototype in order to check if the prototype's user interfaces would enable efficient and user-friendly knowledge extraction from registry data.

Some experts' first impression of the prototype was that it was perfect and ready for production. The false impression was because the prototype looked and functioned like a finished product since it was a high fidelity prototype, which allowed interaction.

The student evaluators who role-played experts were mainly young, smart, had a strong IT background and were not tested in a stratified way. Therefore, the results were assumed to still represent the users since the overall experts used were smart and understood how IT technologies can be employed to such domains.

When it came to learning how to use the designed artefact through trying out the different features, it was expected that the physicians would take a bit longer time than the IT students who role-played experts. Nevertheless, it turned out that it was quick for them to learn and understand the design concept relatively fast.

6.3 Chapter Summary

This chapter has presented the evaluation results and the discussion. It has explored several design case studies through which the developed artefact was used to answer the research questions, and thereafter, provided a reflection of this thesis study.

Chapter 7: Future Works and Conclusions

7.1 Future Works

While the designed prototype succeeded in conveying the design concepts that aimed at enhancing visualisation of data mined from the cancer registries, many opportunities for extending the scope of this thesis remain. This reveals the need for long-term and short-term future works.

7.1.1 Short-term Future Works

The evaluation results showed that the prototype could still be improved through increasing the number of iterations. Therefore, the constructive remarks and suggestions of the expert evaluations can be used to evolve the prototype further. Performing more usability testing and expert evaluations will help to fine tune the prototype until there are minimum or no design faults is found.

The work performed in this thesis project provides a basis for future research in several areas within clinical registries. It demonstrated the viability of the design idea that is not limited to only the cancer registries. The idea could also be used in other clinical registries that experience similar inconveniences with presenting their mined data. Such a design solution could open doors for users to be independent and free to explore the available registry data and resources.

Further research on cancer registries may lead to more academic writings within this area which could be interesting since there is a limited number of references published on this combination theme of interaction design and cancer registries.

The use of small devices such as the mobile phones and tablets is becoming more and more popular in working environments. It will therefore be interesting to see how such a design solution can be adapted to small devices.

7.1.2 Long-term Future Works

E-learning is particularly interesting since the same resource could be used for training purposes. This would result in a completely new way of interacting with registry data, which will make students comfortable interacting with large publicly accessible databases and exploring the data for their own research purposes. IT development will mean in this case also paying attention to the smart e-learning trends and patterns that can already be noticed in other domains.

This work could also be applicable when utilising techniques for data mining and following user needs in which visualisation and smart interaction would be of great assistance. Smart visualisation and smart interface designs would also enable data mining in online fashion, which will be appreciated by physicians and other researchers.

7.2 Conclusions

To answer the research questions, this thesis has produced a prototype that demonstrated a design solution for enhancing visualisation of data mined from the cancer registries. Through applying a combination of iterative and parallel design approaches, three hand-drawn sketched design alternatives of low fidelity created in the initial iteration evolved into one rich interactive high fidelity prototype by the end of the third iteration. The low fidelity prototypes were especially useful for quickly presenting and modifying design ideas particularly when working with the hand-drawn sketches. The high fidelity prototype on the other hand, mimicked a finished product that was utilized fully during the evaluation.

In this work, emphasis was also put into understanding and follow-up of the concepts of user experience and usability in order to come up with an appealing design solution. Their aspects together with the design science guidelines were of great assistance when designing for user experience and producing a usable final prototype, which had continuously undergone several design transformations and alterations in order to improve its quality, simplicity, ease of use, and effectivity as it evolved from one iteration to the next.

The resulting prototype has provided users with a choice of several visualization possibilities for data mined from the long-term national cancer registries and dedicated web-databases in a smart, pleasant and user-friendly manner. In the real world, this sounds very easy and possible, but in reality, one would first need to understand and consider all the clinical processes of collecting, storing, sharing and mining such kinds of clinical data. In addition, legal constraints and ethical considerations are also crucial aspects to be considered when dealing with data from clinical registries.

An evaluation following Nielsen's ten heuristics was used to assess the elements of the prototype during the evaluation process. The prototype scored highly on all the ten dimensions: *H1=Visibility of system status scored* 8.3 out of 10; *H2=Match between the system and the real world* scored 9.6; *H3=User control and freedom* scored 8.8; *H4=Consistency and standards* scored 9.4; *H5= Error prevention* scored 9.8; *H6=Recognition rather than recall* scored 9.3*: H7=Flexibility and efficiency of use* scored 9; *H8=Aesthetic and minimalist design* scored 9; *H9=Help users recognize, diagnose, and recover from errors* scored 9.1; and *H10= Help and documentation* scored 8.5.

Time was another measure through which results showed that it was faster to solve tasks when using the developed prototype than when following the current practises. An average of *06:10 minutes* of working in a traditional set-up could be brought down to an average of *02:11 minutes.*

The results suggest that users were comfortable with the interaction and managed to complete all the tasks without any significant errors or impossible to overcome difficulties. All in all, the prototype seemed to have exhausted evaluation possibilities at this level so it can be now taken to clinicians to invite for their further suggestions.

As a contribution to domains with clinical databases and to cancer registries in particular, this solution will simplify the process of extracting information from the registry data, and will enable various user groups, clinicians, medical students and the public, to directly interact with the system and have quick access to the data stored in these registries. It will also bring the knowledge of Human Computer Interaction into the well established clinical domains with growing data bases.

References

Album, Adobe Photoshop. (2004). 2.0 Tutorial.

Apers, Chris, & Paterson, Daniel. (2010). Development Tools *Beginning iPhone and iPad Web Apps* (pp. 3-11): Springer.

Beck, Kent. (1999). Embracing change with extreme programming. *Computer, 32*(10), 70-77.

Benson, Jim, & Barry, Tonianne Demaria. (2011). *Personal Kanban: Mapping Work, Navigating Life*: Modus Cooperandi Press.

Bifet, Albert, Holmes, Geoff, Kirkby, Richard, & Pfahringer, Bernhard. (2010). MOA: Massive Online Analysis. *The Journal of Machine Learning Research, 11*, (1601-1604).

Bifet, Albert, Holmes, Geoff, Pfahringer, Bernhard, Kranen, Philipp, Kremer, Hardy, Jansen, Timm, & Seidl, Thomas. (2010). MOA: Massive Online Analysis, a Framework for Stream Classification and Clustering (pp. (3-16)): University of Waikato, Hamilton, New Zealand,, Aachen University, Germany.

Bryman, Alan. (2012). *Social research methods*: Oxford university press.

Coakes, Sheridan J, & Steed, Lyndall. (2009). *SPSS: Analysis without anguish using SPSS version 14.0 for Windows*: John Wiley & Sons, Inc.

Cook, Diane J, & Holder, Lawrence B. (2006). *Mining graph data*: John Wiley & Sons.

Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med, 34*(2), 113-127. doi: 10.1016/j.artmed.2004.07.002

Dingsøyr, Torgeir, Dybå, Tore, & Moe, Nils Brede. (2010). Agile software development: an introduction and overview *Agile Software Development* (pp. 1-13): Springer.

Dix, Alan. (2009). *Human-computer interaction*: Springer.

Eaton, Chris, Deroos, Dirk, Deutsch, Tom, Lapis, George, & Zikopoulos, Paul. (2012). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data* Retrieved from http://public.dhe.ibm.com/common/ssi/ecm/en/iml14296usen/IML14296USEN.PDF

Ericsson, Jan, Åberg, Anders, Holmberg, Lars, Rozell, Barbro Lundh, Ayoubi, Shiva, Khan, Staffan, & Klint, Åsa. (2011). Cancer Incidence in Sweden 2010 *Statistics – Health and Medical Care* (pp. 1-100).

Few, Stephen. (2013). Data Visualisation for Human Perception. In M. S. a. R. F. Dam (Ed.), Encyclopedia of Human-Computer InteractionThe Encyclopedia of Human-Computer Interaction, (2nd ed.): The Interaction Design Foundation. Retrieved from http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html.

Freeman, Elisabeth, & Freeman, Eric. (2006). *Head First Html With CSS & XHTML*: O'Reilly Media, Inc.

Gardner, Brett S. (2011). Responsive Web Design: Enriching the User Experience. *Connectivity and the User Experience, 13*.

Goodwin, Kim. (2011). *Designing for the digital age: How to create human-centered products and services*: John Wiley & Sons.

Guilizzoni, G. (2010). "Balsamiq Mockups. *Balsamiq, http://www. balsamiq. com*.

Hall, Mark, Frank, Eibe, Holmes, Geoffrey, Pfahringer, Bernhard, Reutemann, Peter, & Witten, Ian H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations, 11*(1), (10-18).

Heer, Jeffrey, Card, Stuart K, & Landay, James A. (2005). *Prefuse: a toolkit for interactive information visualization.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Hevner, Alan, T.March, Salvatore, Park, Jinsoo, & Ram, Sudha. (2004). Design Science in Information Systems Research. *MIS Quarterly, 28*(1), 75-105.

Justinmind. (2014). Justinmind Prototyper. from http://justinmind.com/

Kalil, Tom, Butler, Jeff, Hogle, Mark, Iacono, Suzanne, Loper, Mark, Sankaran, Vish, . . . Wynne, Marc. (2012). Demistyfying Big Data: A Practical Guide To Transforming The Business of Government (pp. 40). Washington, D.C: TechAmerican Foundation.

KanbanFlow. (2014). KanbanFlow - Lean Project Management, Simplified. Retrieved 30.05.2014, from https://kanbanflow.com/

Kanza, G, & Babic, Ankica. (2014). *Data Mining in Cancer Registries: A Case for Design Studies.* Paper presented at the XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013.

Khattree, Ravindra, & Naik, Dayanand N. (2000). *Multivariate data reduction and discrimination with SAS software*: Sas Institute.

Kirkby, Richard, Eibe Frank, and Peter Reutemann. (2007). WEKA Explorer User Guide for Version 3-5-5.

Köhler, Jacob, Baumbach, Jan, Taubert, Jan, Specht, Michael, Skusa, Andre, Rüegg, Alexander, . . . Philippi, Stephan. (2006). Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics, 22*(11), 1383-1390.

MacLean, Allan, Young, Richard M., & Moran, Thomas P. (1989). Design Rationale: The argument behind the artifact. *SIGCHI Bull., 20*, 247-252. doi: 10.1145/67450.67497

Manyika, James, Chui, Michael, Brown, Brad, Bughin, Jacques, Dobbs, Richard, Roxburgh, Charles, & Byers, Angela Hung. (2011). Big data: The next frontier for innovation, competition, and productivity *McKinsey Global Institute* (pp. 1-143).

March, Salvatore T, & Smith, Gerald F. (1995). Design and natural science research on information technology. *Decision support systems, 15*(4), 251-266.

Nielsen, Jacob, & Molich, Rolf. (1990). Heuristic evaluation of user interfaces. *CHI '90*, 249-256. doi: 10.1145/97243.97281

Nielsen, Jakob. (1994). *Usability engineering*: Elsevier.

Nunamaker Jr, Jay F, Chen, Minder , & Purdin, Titus D.M. (1990). Systems development in information systems research. *Journal of Management Information Systems, 3*, (89-106).

Persidis, Aris. (2000). Data Mining. *Nature Biotechnology, 18*, (IT35 - IT36). doi: 10.1038/80073

Pickard, Alison Jane, & Childs, Sue. (2007). *Research methods in information*: Facet London.

Rogers, Yvonne, Sharp, Helen, & Preece, Jenny. (2011). *Interaction design: beyond human-computer interaction* (3rd ed.). UK: John Wiley & Sons Ltd.

Saffer, Dan. (2010). *Designing for interaction: creating innovative applications and devices*: New Riders.

Shaw, Michael J., Subramaniam, Chandrasekar, Tan, Gek Woo, & Welge, Michael E. (2001). Knowledge management and data mining for marketing. *Decision support systems, 31*(1), (127-137).

Šmite, Darja, Moe, Nils Brede, & Ågerfalk, Pär J. (2010). *Agility Across Time and Space: Implementing Agile Methods in Global Software Projects*: Springer.

Sommerville, Ian, & Sawyer, Pete. (1997). *Requirements engineering: a good practice guide*: John Wiley & Sons, Inc.

Tidwell, Jenifer. (2010). *Designing interfaces*: " O'Reilly Media, Inc.".

Walliman, Nicholas. (2006). *Social research methods*: Sage.

Ware, Colin. (2000). *Information Visualization: Using perception to think* (Vol. 2). San Francisco: Morgan Kaufmann

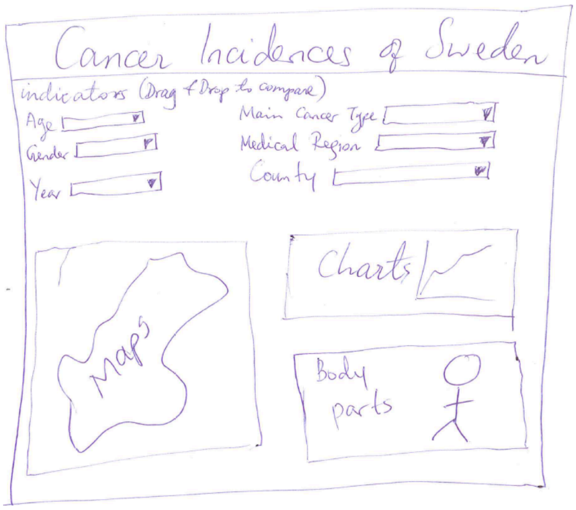
Williams, Laurie. (2012). What agile teams think of agile principles. *Communications of the ACM, 55*(4), 71-76.

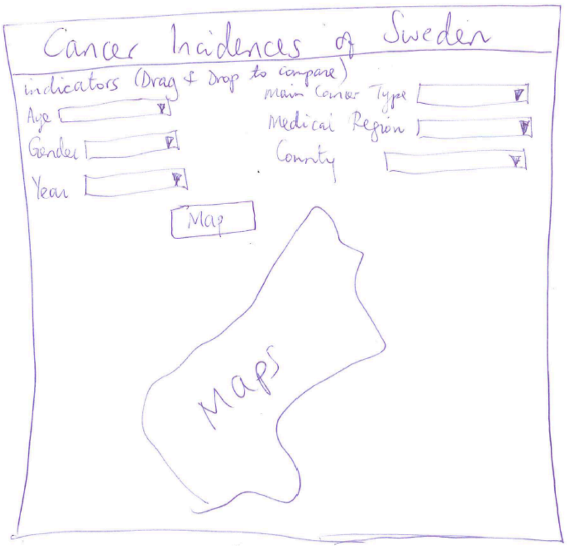
Wilson, Andrew M., Thabane, Lehana, & Holbrook, Anne. (2003). Application of data mining techniques in pharmacovigilance. *British Journal of Clinical Pharmacology, 57*(2), (127-134). doi: 10.1046/j.1365-2125.2003.01968.x

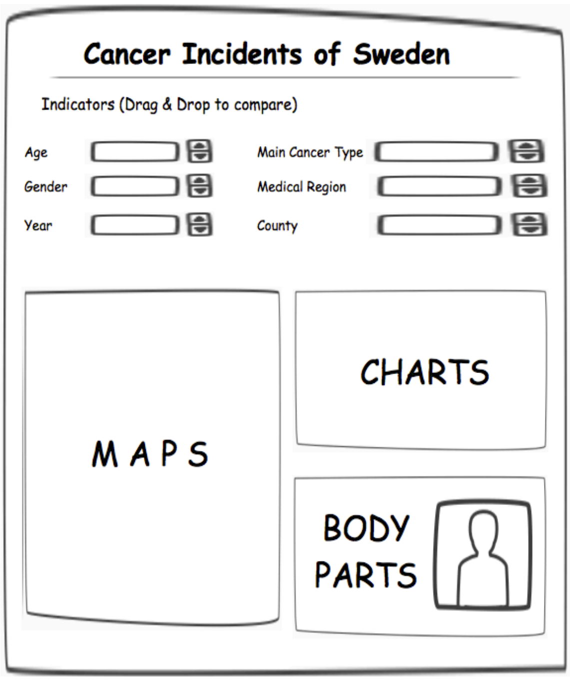
Appendices

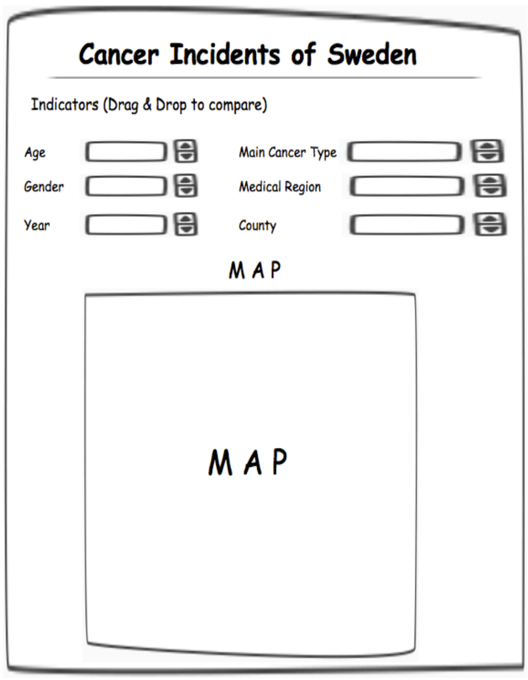
Appendix 1: Hand-drawn & electronic sketches from first iteration

Design alternative 1:

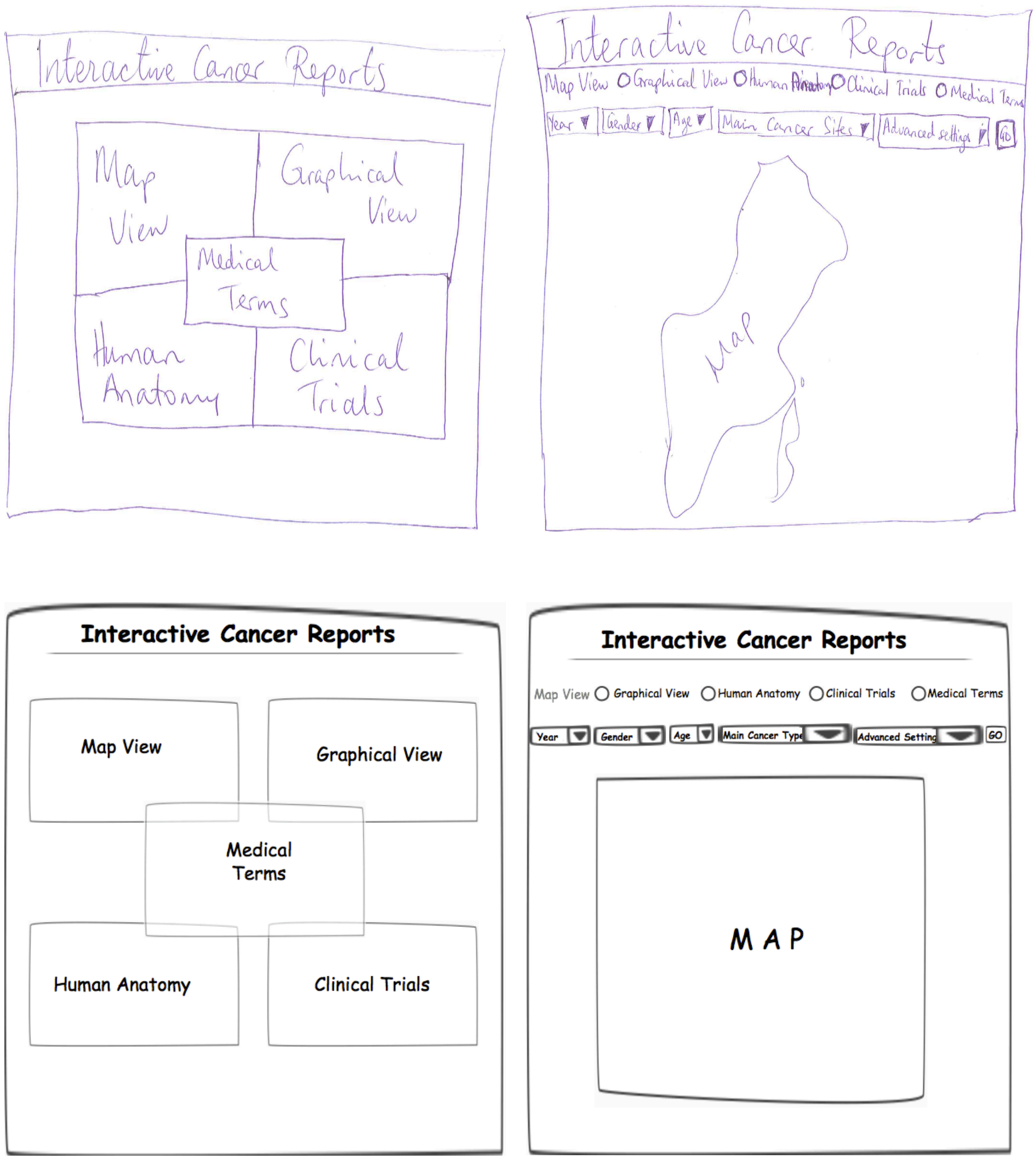




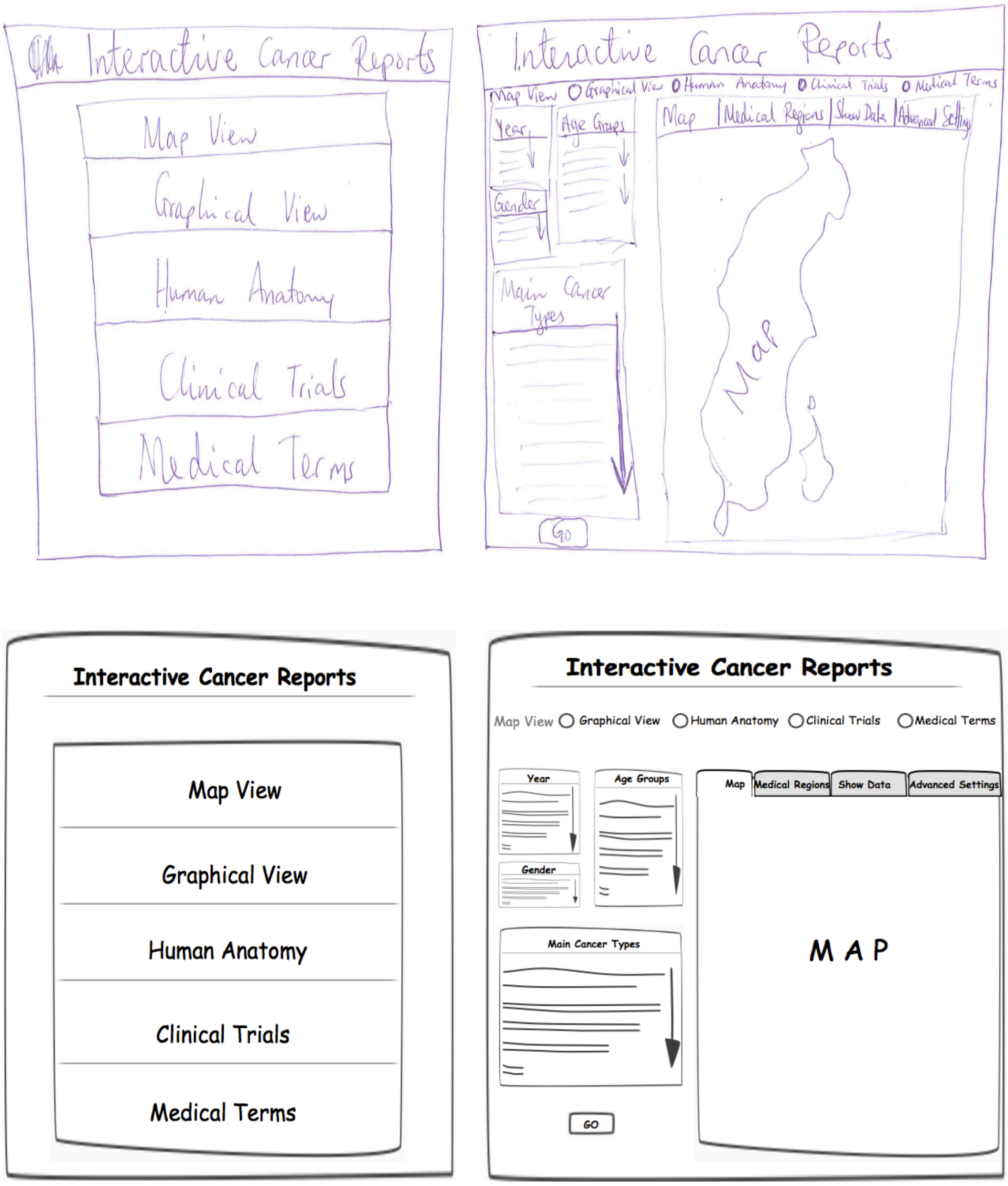




Design alternative 2:



Design alternative 3:


Interactive Cancer Reports
Map View
Graphical View
Human Anatomy
Clinical Trials
Medical Terms
Interactive Cancer Reports
Map View
Graphical View
Human Anatomy
Clinical Trials
Medical Terms
Year
Age Groups
Gender
Main Cancer Types
Go
Map
Medical Regions
Show Data
Advanced Settings
Map
Interactive Cancer Reports
Map View
Graphical View
Human Anatomy
Clinical Trials
Medical Terms
Interactive Cancer Reports
Map View
Graphical View
Human Anatomy
Clinical Trials
Medical Terms
Year
Age Groups
Gender
Main Cancer Types
GO
Map
Medical Regions
Show Data
Advanced Settings
M A P

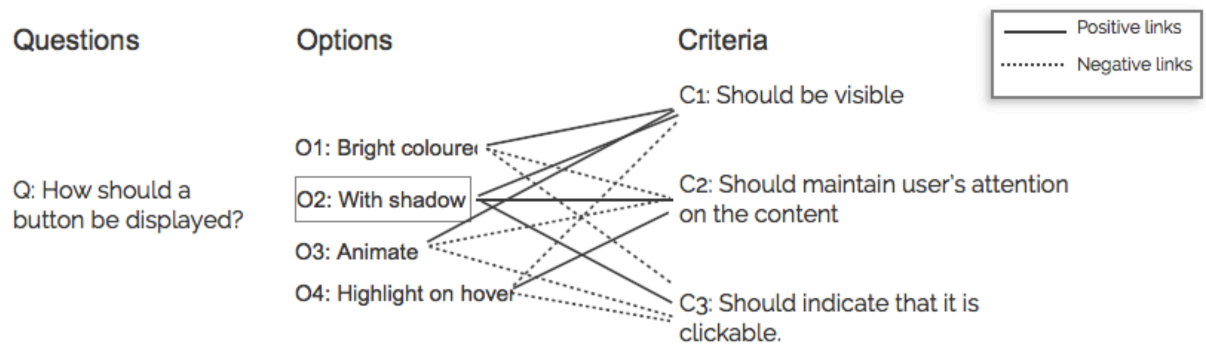
Appendix 2: Question Option Criteria (QOC) Formulation

The Questions, Options and Criteria were first constructed in a table before matching the options against the criteria. Therefore, Appendix 2a presents first each QOC set in a table, which is followed by its diagrammatical presentation.

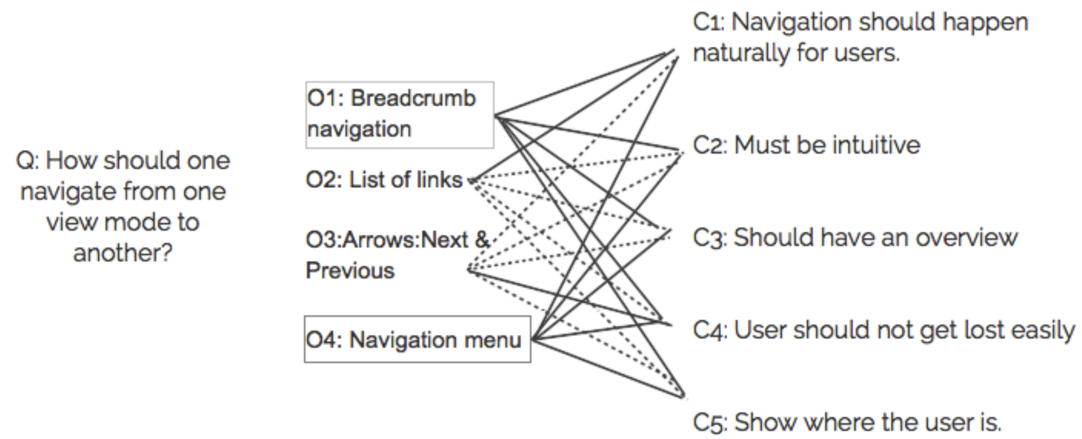
Appendix 2a: First Iteration: QOC set

This QOC set was used in the design process during the first iteration.

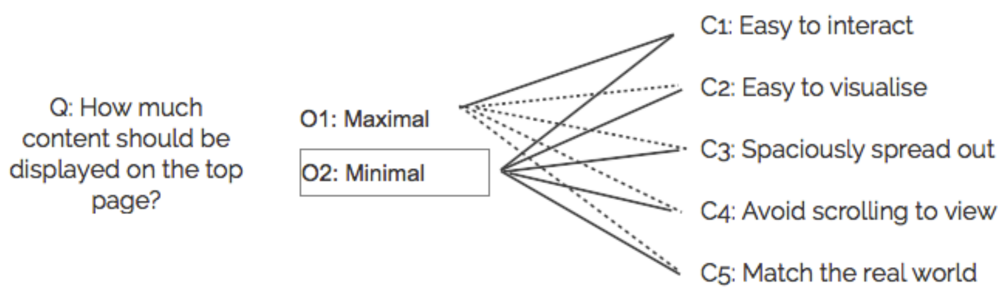
Q1: How should a button be displayed?	O1: Bright coloured O2: With shadow O3: Animate O4: Highlight on hover	C1: Should be visible C2: Should maintain user's attention on the content C3: Should indicate that it is clickable.



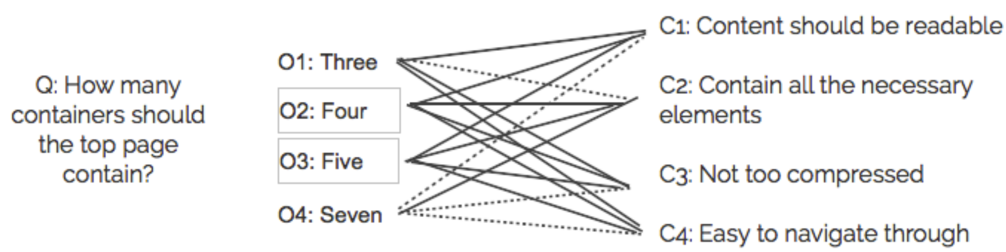
Q2: How should one navigate from one view mode to another?	O1: Breadcrumb navigation O2:List of links O3:Arrows:Next & Previous O4: Navigation menu	C1: Navigation should happen naturally for users. C2: Must be intuitive C3: Should have an overview C4: User should not get lost easily C5: Show where the user is.



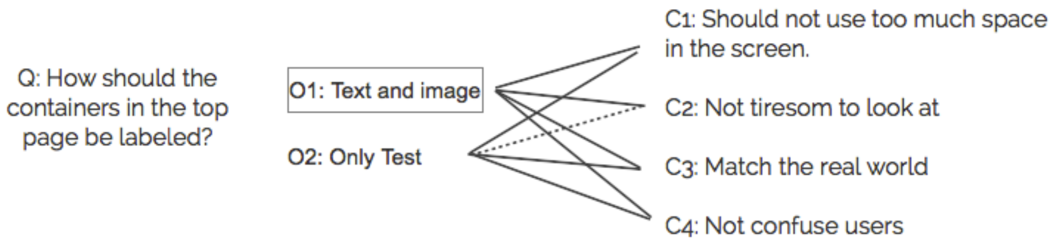
Q3: How much content should be displayed on the top page?	O1: Maximal O2: Minimal	C1: Easy to interact C2: Easy to visualise C3: Spaciously spread out. C4: Avoid scrolling to view C5: Match the real world



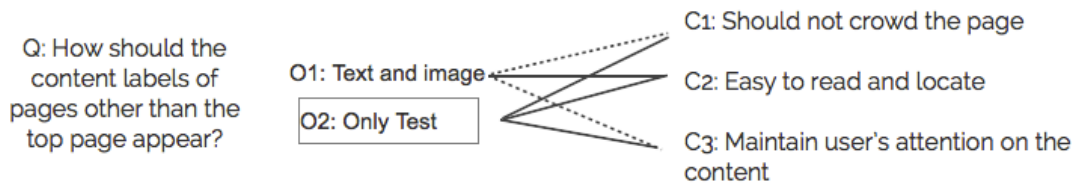
Q4: How many containers should the top page contain?	O1: Three O2: Four O3: Five O4: Six O5: Seven	C1: Content should be readable C2: Contain all the necessary elements C3:Not too compressed C4: Easy to navigate through



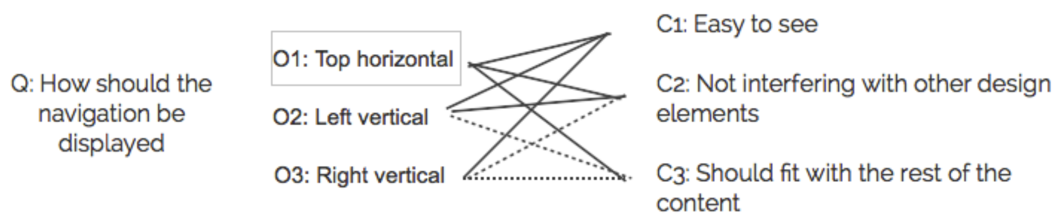
Q5: How should the containers in the top page be labeled?	O1: Text and image O2: Only Test	C1: Should not use too much space in the screen. C2: Demonstrate affordance C3: Match the real world C4: Not confuse users



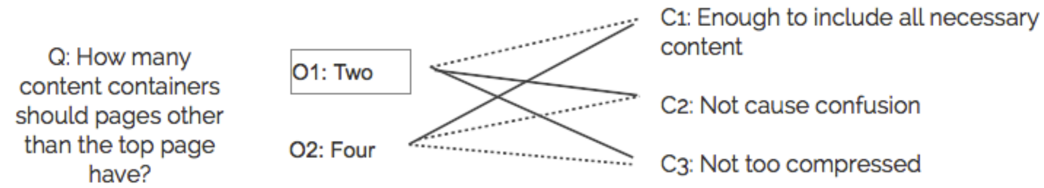
Q6: How should the content labels of pages other than the top page appear?	O1: Text and images O2: Only Test	C1: Should not crowd the page C2: Easy to read C3: Easy to locate C4: Maintain user's attention on the content



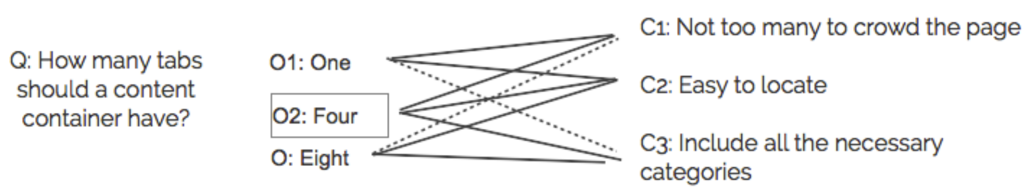
Q7: How should the navigation be displayed	O1: Top horizontal O2: Left vertical O3: Right vertical	C1: Easy to see C2: Not interfering with other design elements C3: Should fit with the rest of the content



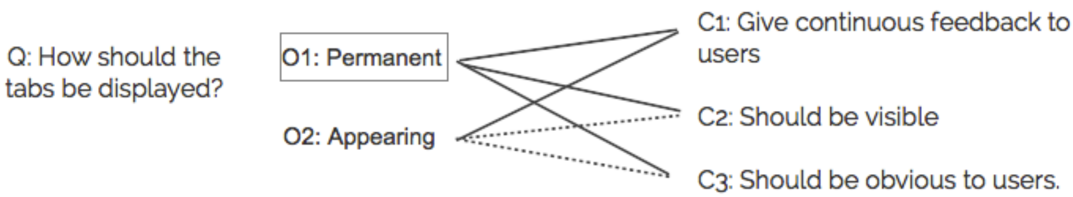
Q8: How many content containers should pages other than the top page have?	O1: Two O2: Four	C1: Enough to include all necessary content C2: Not cause confusion C3: Not too compressed



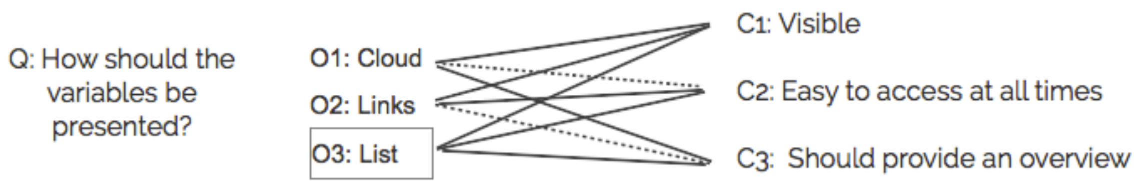
Q9: How many tabs should a content container have?	O1: One O2: Four O3: Eight	C1: Not too many to crowd the page C2: Easy to locate C3: Include all the necessary categories



Q10: How should the tabs be displayed?	O1: Permanent O2: Appearing	C1: Give continuous feedback to users C2: Should be visible C3: Should be obvious to users.



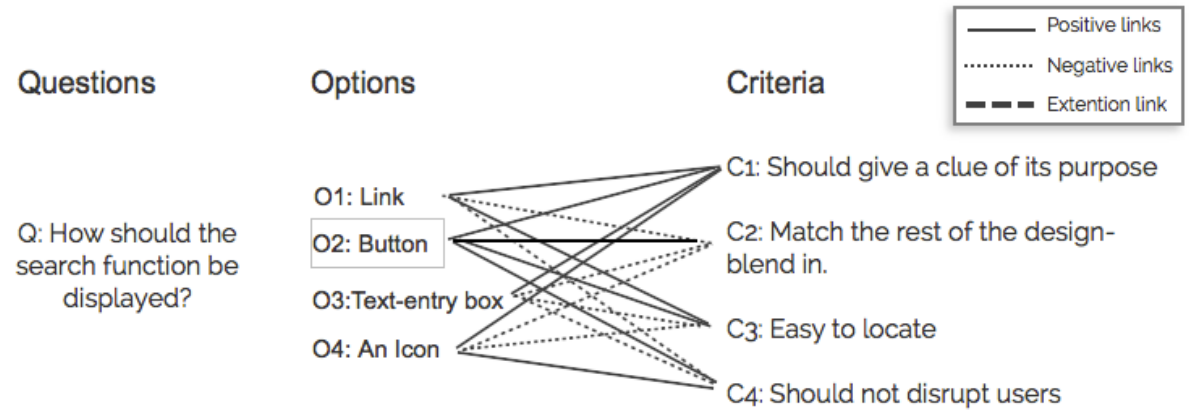
Q11: How should the variables be presented?	O1: List O2: Cloud O3: Links	C1: Visible C2: Easy to access at all times C3: Should provide an overview



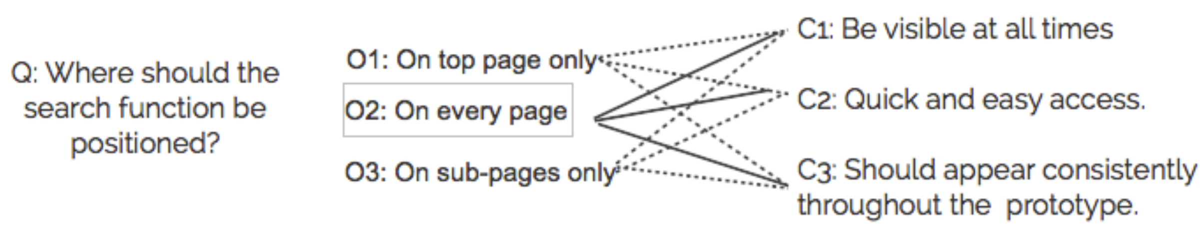
Appendix 2b: Third Iteration: QOC set

This QOC set together with the set presented in the Appendix 2a were used for designing the final prototype during the first iteration.

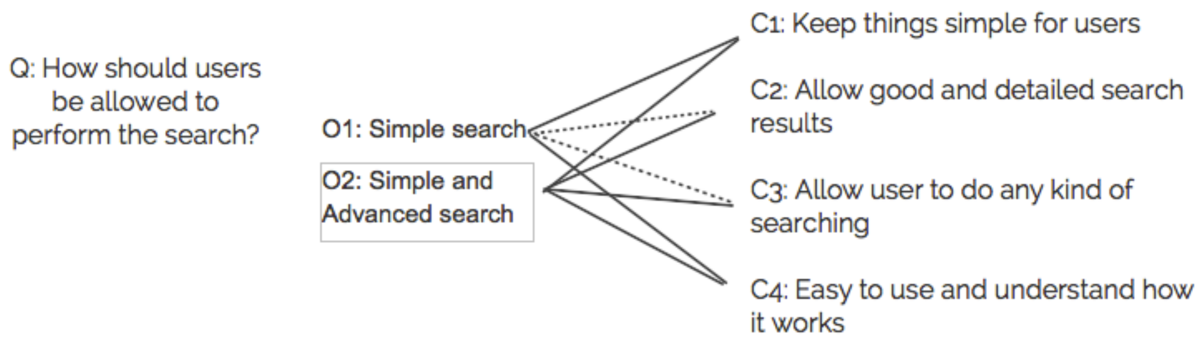
Q1: How should the search function be displayed?	O1: Link O2: Button O3:Text-entry box O4: An Icon	C1: Should give a clue of its purpose C2: Match the rest of the design-blend in. C3: Easy to locate C4: Should not disrupt users



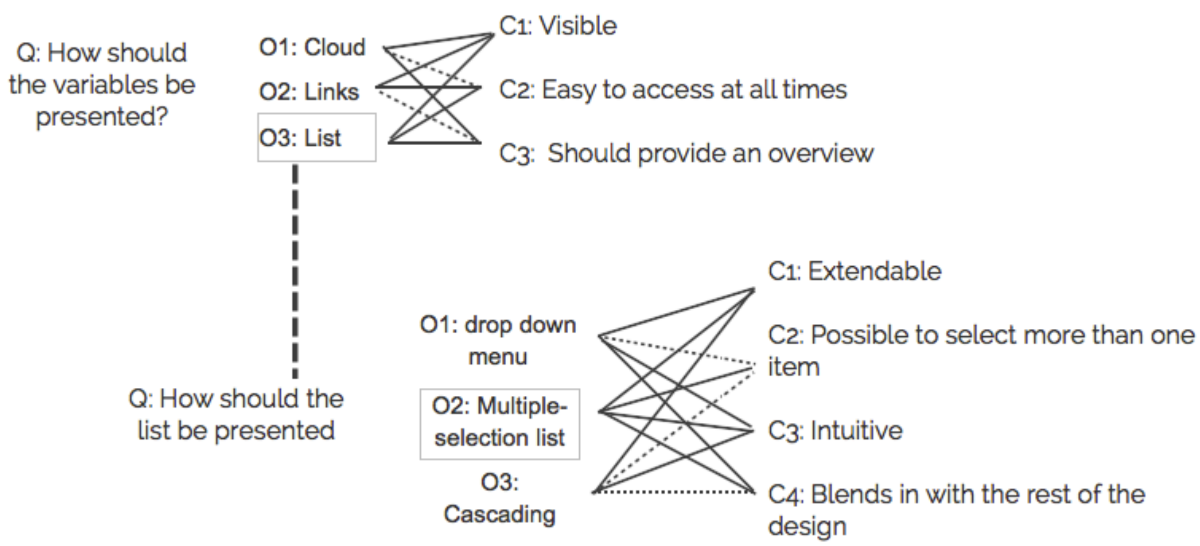
Q2: Where should the search function be positioned?	O1: On top page only O2: On every page O3: On sub-pages only	C1: Be visible at all times C2: Quick and easy access. C3: Should appear consistently throughout the prototype.



Q3: How should users be allowed to perform the search?	O1: Simple search O2: Simple and Advanced search	C1: Keep things simple for users C2: Allow good and detailed search results C3: allow user to do any kind of searching C4: Easy to use and understand how it works



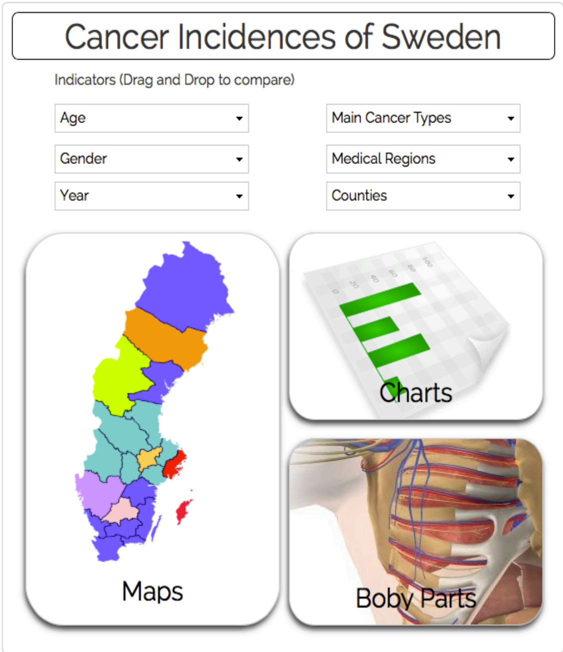
**Q4: How should the variables be presented? (Extension to the last Qn)	O1: List O2: Cloud O3: Links	C1: Visible C2: Easy to access at all times C3: Should appear consistently throughout the prototype.
Q5: How should the list be presented	O1: drop down menu O2: Multiple-selection list O3: Cascading menus	C1: Extendable C2: Possible to select more than one item C3: Intuitive C4: Blends in with the rest of the design

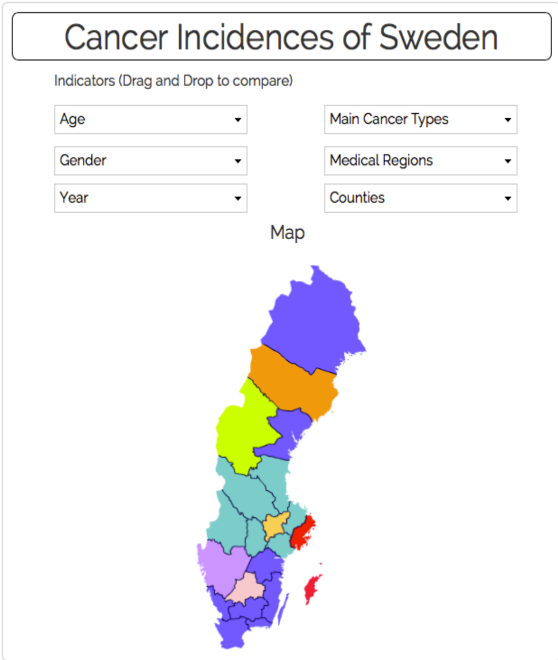


Appendix 3: Interactive prototype from second iteration

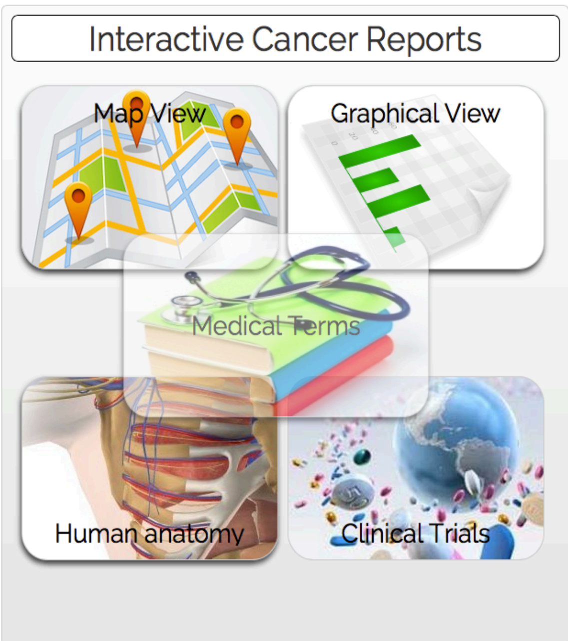
Screenshots of the parallel prototypes with interactivity created during the second iteration.

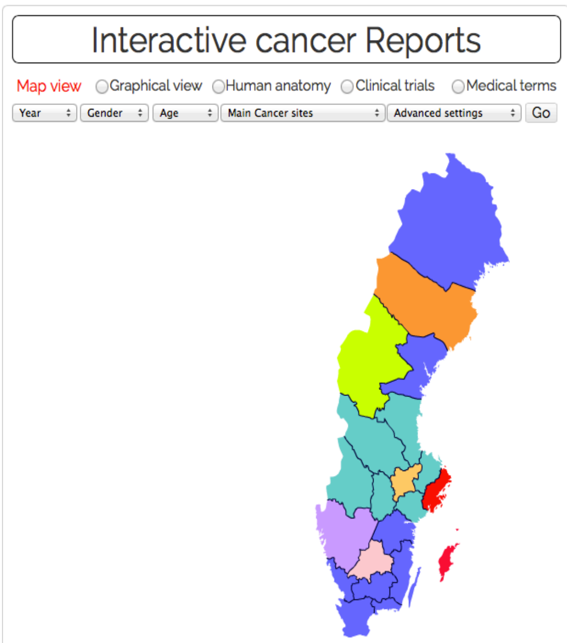
Design alternative 1:



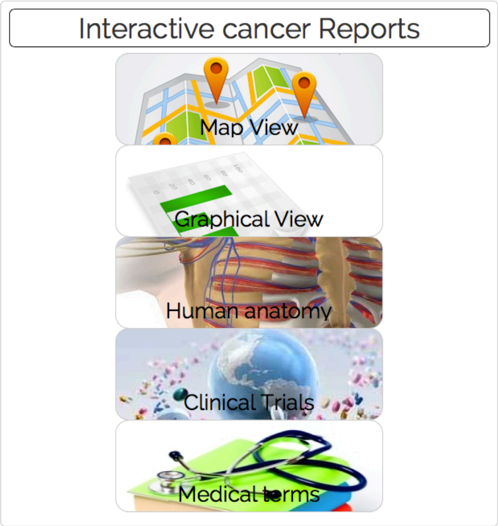


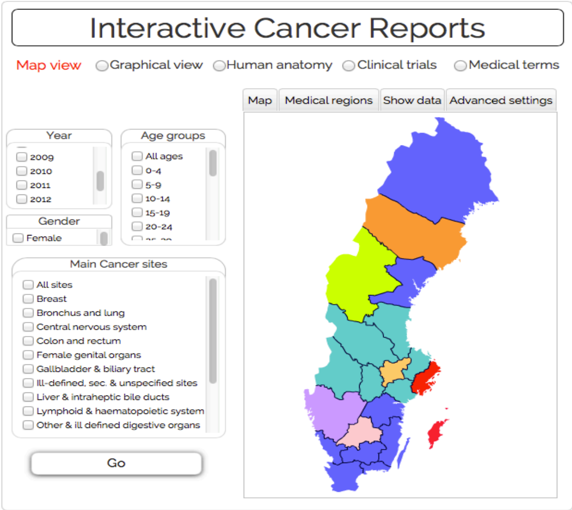
Design alternative 2:





Design alternative 3:


Interactive cancer Reports
Map View
Graphical View
Human anatomy
Clinical Trials
Medical terms


Interactive Cancer Reports
Map view
Graphical view
Human anatomy
Clinical trials
Medical terms
Map
Medical regions
Show data
Advanced settings
Year
2009
2010
2011
2012
Age groups
All ages
0-4
5-9
10-14
15-19
20-24
Gender
Female
Main Cancer sites
All sites
Breast
Bronchus and lung
Central nervous system
Colon and rectum
Female genital organs
Gallbladder & biliary tract
Ill-defined, sec. & unspecified sites
Liver & intrahepatic bile ducts
Lymphoid & haematopoietic system
Other & ill defined digestive organs
Go

Appendix 4: Usability Testing

Appendix 4a: Semi-structured questionnaire for the Usability Testing

This semi-structured questionnaire was completed online from the following links:

For Design Alternative 1:
https://docs.google.com/forms/d/1zGDYwkb0Tp5KEVs9ui2TfJLcDrPhs-7yOOJL5s2SSY0/viewform?usp=send_form

For Design Alternative 2:
https://docs.google.com/forms/d/1nzrhiGFszpniprJRSaP006N--hmqTECQgNQM_A-BaOA/viewform?usp=send_form

For Design Alternative 3:
https://docs.google.com/forms/d/1kNljpq1bZn4IZTrQGpdgdHEbFDix_eEUvP8zBqF7QMo/viewform?usp=send_form

USABILITY TESTING: QUESTIONNAIRE

Please circle the most appropriate selection:

Gender:	Male	Female				
Age range:	20-29	30-39	40-49	50-59	60-69	70+
Computer / Internet use	Daily	Weekly	Monthly	Yearly	Never	
Searched the web for information about medical issue(s) before:		Yes	No	Don't remember		

Please rate the designed prototype using the ratings: Strongly disagree, Disagree, Neutral, Agree or Strongly agree.

	Question	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	I was able to complete the tasks without difficulties.					
2	I was able to complete the tasks without any help.					

3	Terminologies and icons used were clear and easy to understand.					
4	The tool was slow and confusing to use.					
5	The prototype provided the necessary functionalities to enable me perform the tasks I was assigned.					
6	The simple and intuitive design made it easy to learn how to use the tool.					
7	It was easy to forget how to use the tool if I do not use it in the daily bases.					
8	It will take short time to master the usage of this tool.					
9	The design presents information in an aesthetically pleasing manner					
10	I kept on making errors while performing tasks.					
11	The tool prevented me from selecting irrelevant variables.					
12	I think the prototype was consistent, easy to follow how the different parts were connected and easy to navigate through.					
13	I enjoyed using this tool.					
14	I would like to use this tool in the future.					

15. What design features do you think made this design appealing?

____

____

____

16. What design features do you think should be included in order to improve this interface?

____

____

____

17. Please add any comments or recommendations regarding the overall design.

____

____

____

Appendix 4b: Consent Form

Consent form

Request for participation in the research project

“Human Computer Interaction Design for Data Mining In Cancer Registries”

Investigators:

I am a master student in Information Science at the University of Bergen writing a master’s thesis about “Human Computer Interaction Design for Data Mining In Cancer Registries”. My supervisor is Ankica Babic from the Department of information science and Media Studies at the University of Bergen and the Department of Biomedical Engineering in Linkoping University.

Background and purpose of the thesis:

This is a request for you to participate in this research project that aims at providing different design solutions that will enable efficient and user-friendly visualisation of data mined from the cancer registries.
The research study will address the following questions:

- How can Human Computer Interaction make interfaces for data visualisation in data mining more efficient and user friendly?
- How can the research of this thesis help transform the current state of presenting data from the cancer registries to a better and preferred state?
- How can smart user interfaces enable good data visualisation for different users groups?
- Will this kind of research and interfaces help to change the way users look at the registry data?

This research thesis is planned to complete on 1st June 2014.

What does the interview entail?

During this written semi-structured interview, participants will be asked to perform some simple tasks on the new design and then afterwards asked to rate the design by agreeing or disagreeing to some statements connected to the tasks and the performance of the designed prototypes.
These statements will focus on the usability aspects, functionalities and interaction between the user and the designed prototypes while the open questions will allow participants to

express their design opinions and comment on how the design can be improved in the next iterations.
The target user groups clinicians, medical students and the public. The estimated time for performing the tasks and completing the questionnaire is estimated to be about one hour.

What will happen to the interview content and the information about you?

Notes and descriptions of the reflections from the interviews carried will be anonymous and the information obtained from the interviews will remain confidential and protected from any access by third parties. Any academic publications resulting from this research thesis upon the completion of the project will not reveal any individual interviewees in any ways.

Voluntary participation:

Participation in this interview is voluntary. You can withdraw your consent to participate in this interview at any time without stating any particular reason.
Therefore, before the interview begins, I ask you to consent to participation by signing that you have read and understood the information about what this thesis project is about and would like to participate in the interviews.

Consent for participation in the interview:

I (name)____have read and understood the information above and give my consent to participate willingly in the interview.

_____________________________ ______________ __________________
(Signature) (date) (place)

Appendix 4c: Usability Testing Tasks

Tasks carried out by users during the Usability Testing

TASKS FOR USABILITY TESTING

These simple tasks aim at familiarising the testing participants with the designed prototypes before they start answering the questionnaire. The same tasks will be used for each of the three design alternatives.

TASK 1:
Find the Female population in Stockholm and the Male population in Norrbotten for the year 2010.

TASK 2:
Find out how many men and women living in Uppsala were diagnosed with cancer of the colon in 2010.

TASK 3:
Find out how many people from Blekinge aged between 25-60 years were diagnosed with cancers of the lip in 2010.

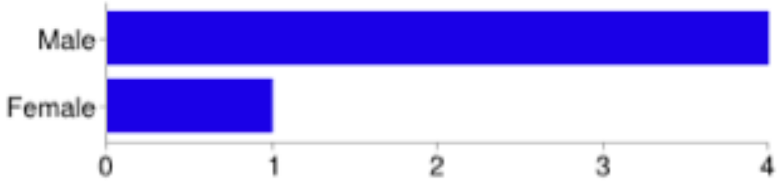
TASK 4:
Find out how many men and women from Stockholm medical region were diagnosed with cancers of the stomach.

TASK 5:
Find out how many people from Skåne died of breast cancer in 2010.

Appendix 4d: Results from Usability Testing

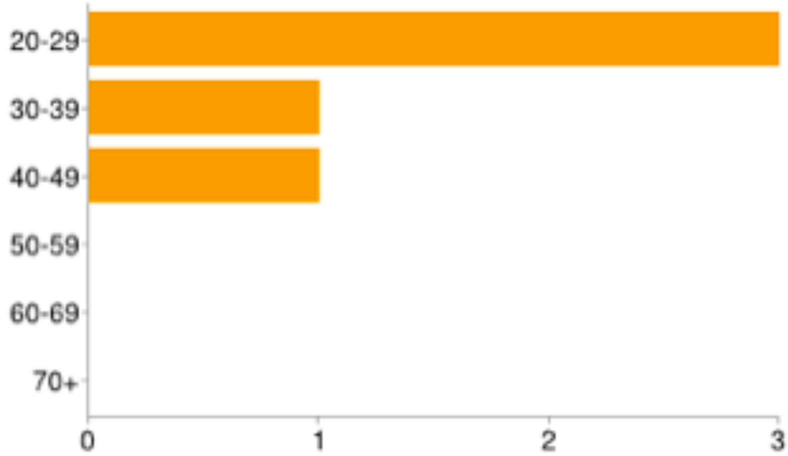
The following information about the five participating users was obtained: gender, age, frequency of computer/Internet use, and whether they had searched the web for medical-related issues.

Participation by gender:



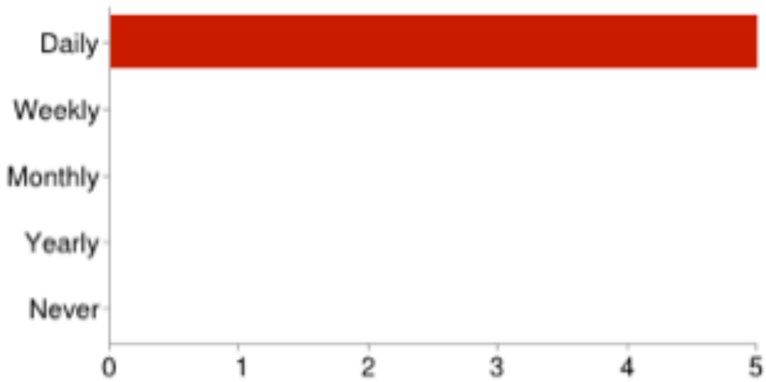
Male	4	80%
Female	1	20%

Age:



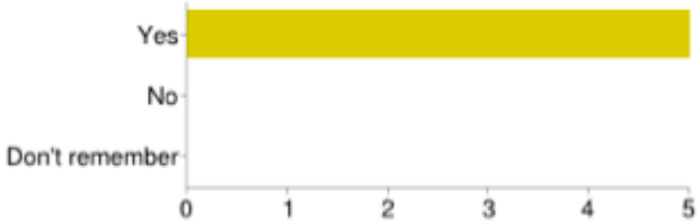
20-29	3	60%
30-39	1	20%
40-49	1	20%
50-59	0	0%
60-69	0	0%
70+	0	0%

Frequency of computer /Internet use



Daily	5	100%
Weekly	0	0%
Monthly	0	0%
Yearly	0	0%
Never	0	0%

Searched the web for information about medical issue(s) before:

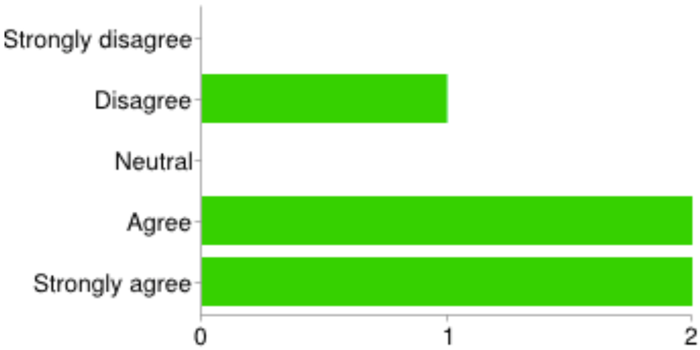


Yes	**5**	100%
No	**0**	0%
Don't remember	**0**	0%

Design alternative 1

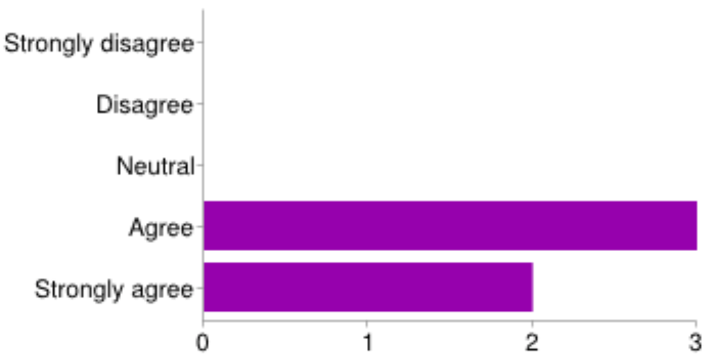
Responses from the Questionnaire.

QN 1:I was able to complete the tasks without difficulties.



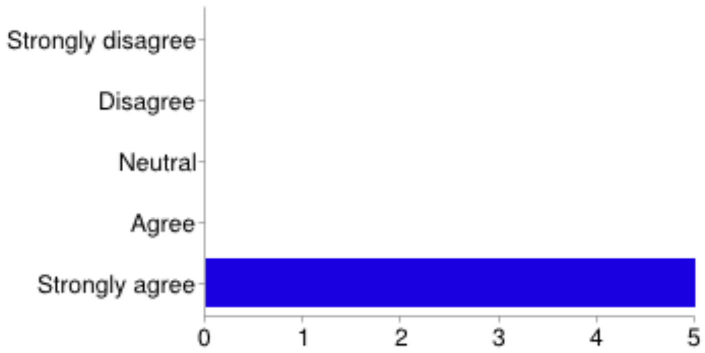
Strongly disagree	**0**	0%
Disagree	**1**	20%
Neutral	**0**	0%
Agree	**2**	40%
Strongly agree	**2**	40%

QN 2: I was able to complete the tasks without any help.



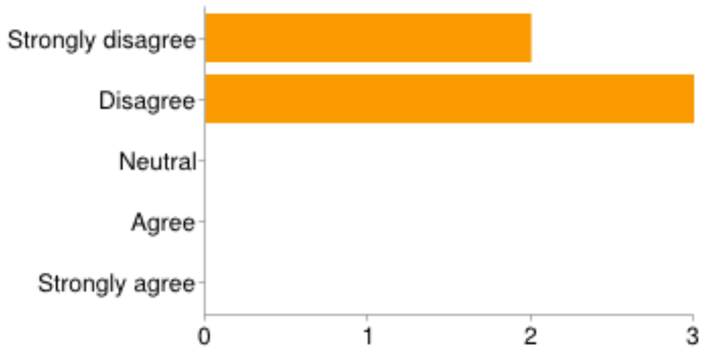
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**3**	60%
Strongly agree	**2**	40%

Qn 3:Terminologies and icons used were clear and easy to understand.



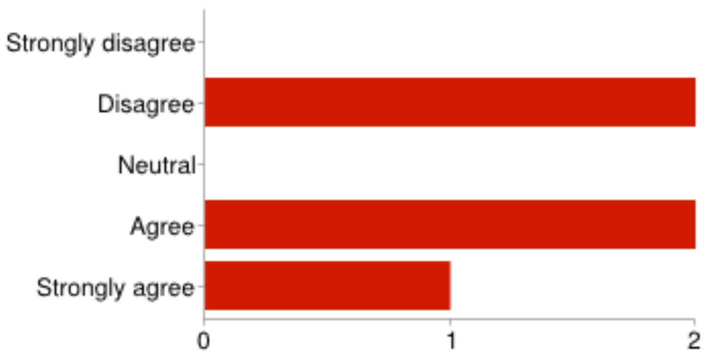
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 4:The tool was slow and confusing to use.



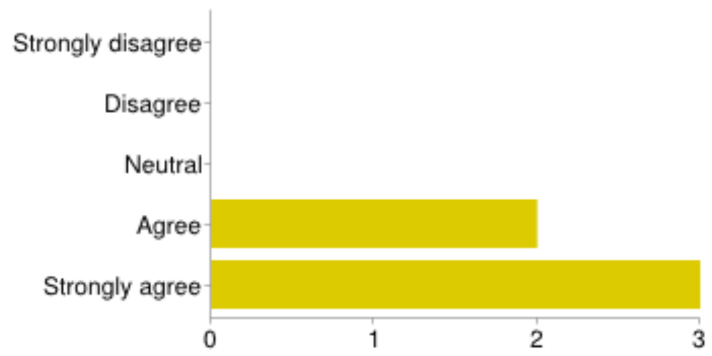
Strongly disagree	**2**	40%
Disagree	**3**	60%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**0**	0%

Qn 5:The prototype provided the necessary functionalities to enable me perform the tasks I was assigned.



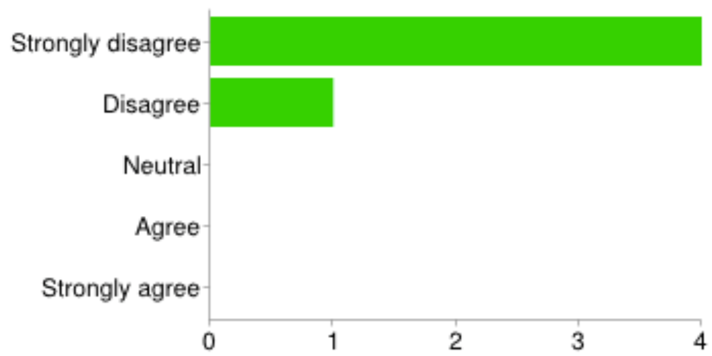
Strongly disagree	**0**	0%
Disagree	**2**	40%
Neutral	**0**	0%
Agree	**2**	40%
Strongly agree	**1**	20%

Qn 6:The simple and intuitive design made it easy to learn how to use the tool.



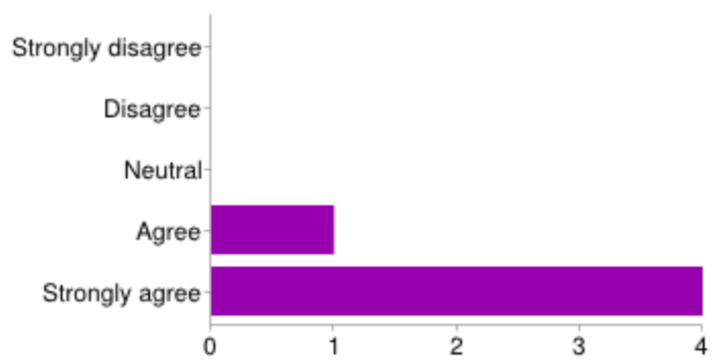
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**2**	40%
Strongly agree	**3**	60%

Qn 7:It was easy to forget how to use the tool if I do not use it in the daily bases.



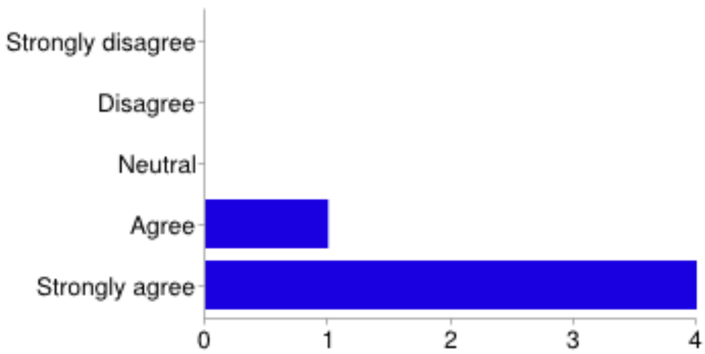
Strongly disagree	**4**	80%
Disagree	**1**	20%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**0**	0%

Qn 8:It will take short time to master the usage of this tool.



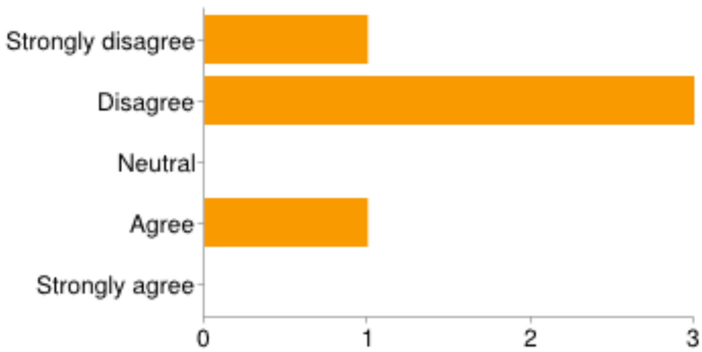
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**1**	20%
Strongly agree	**4**	80%

Qn 9:The design presents information in an aesthetically pleasing manner



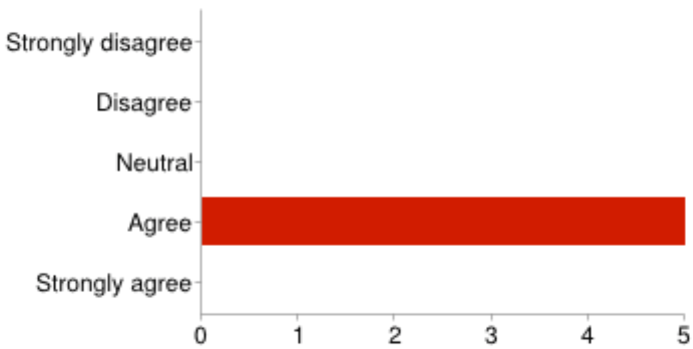
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**1**	20%
Strongly agree	**4**	80%

Qn 10:I kept on making errors while performing tasks.



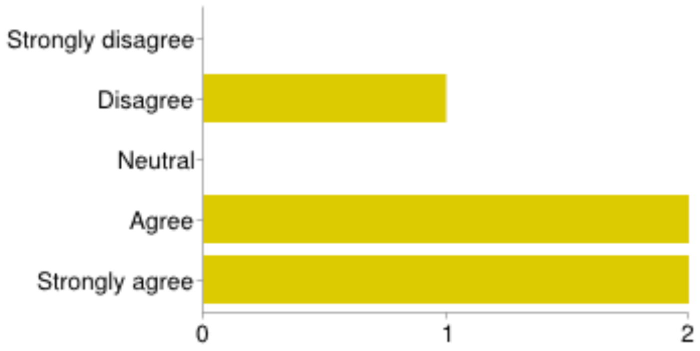
Strongly disagree	**1**	20%
Disagree	**3**	60%
Neutral	**0**	0%
Agree	**1**	20%
Strongly agree	**0**	0%

Qn 11:The tool prevented me from selecting irrelevant variables.



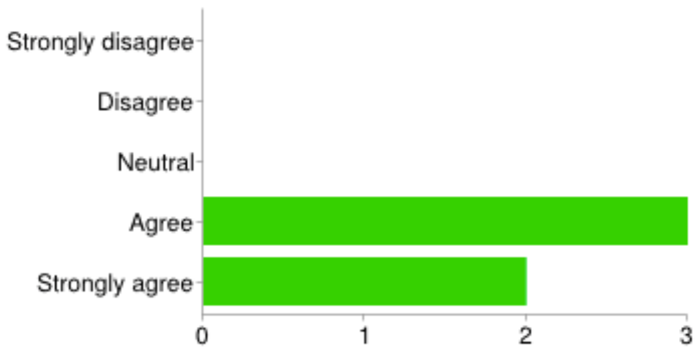
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**5**	100%
Strongly agree	**0**	0%

Qn 12:I think the prototype was consistent, easy to follow how the different parts were connected and easy to navigate through.



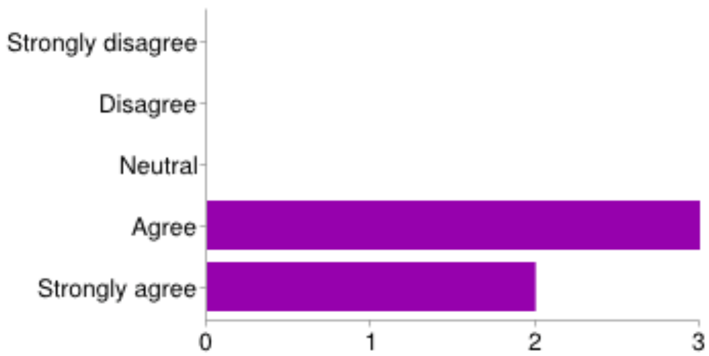
Strongly disagree	0	0%
Disagree	1	20%
Neutral	0	0%
Agree	2	40%
Strongly agree	2	40%

Qn 13:I enjoyed using this tool.



Strongly disagree	0	0%
Disagree	0	0%
Neutral	0	0%
Agree	3	60%
Strongly agree	2	40%

Qn 14:I would like to use this tool in the future.



Strongly disagree	0	0%
Disagree	0	0%
Neutral	0	0%
Agree	3	60%
Strongly agree	2	40%

Qn 15: What design features do you think made this design appealing?

- The dashboard was good but only for some tasks.
- The background images fitted well.
- Simple design - Dashboard
- The possibility to choose to look at the data from the map or chart or the body diagram.

Qn 16: What design features do you think should be included in order to improve this interface?

- GO" button to press when i finish selecting what data i want to see- more data options (categories eg. deaths)
- Search function -Possibility to select more than one item from the same group eg. to select more than one age group.

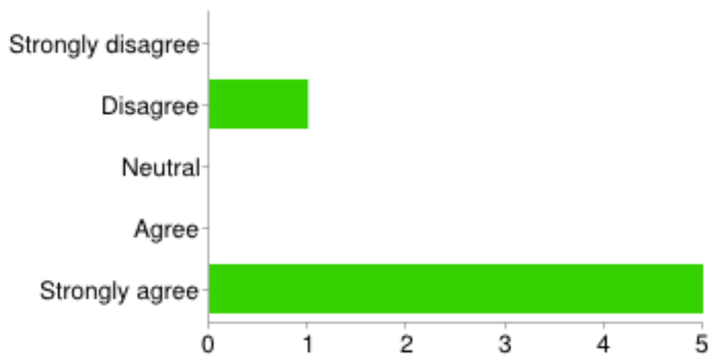
Comment: Please add any comments or recommendations regarding the overall design.

-I could not select more than one variable eg. not possible to choose males and females.
-More variables to make the selection more specific, eg. there was no where to select the "Deaths".
-I think the dashboard makes the prototype seem incomplete, it will be good if this was not the only option

Design alternative 2

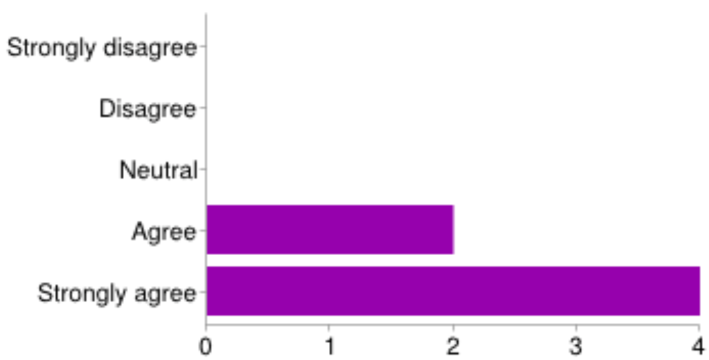
Responses from the Questionnaire.

QN 1:I was able to complete the tasks without difficulties.



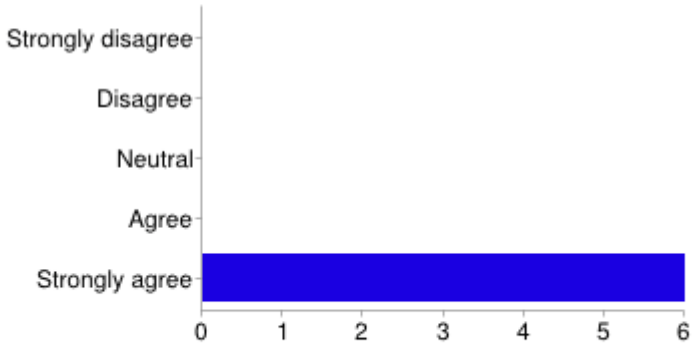
Strongly disagree	**0**	0%
Disagree	**1**	17%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	83%

QN 2: I was able to complete the tasks without any help.



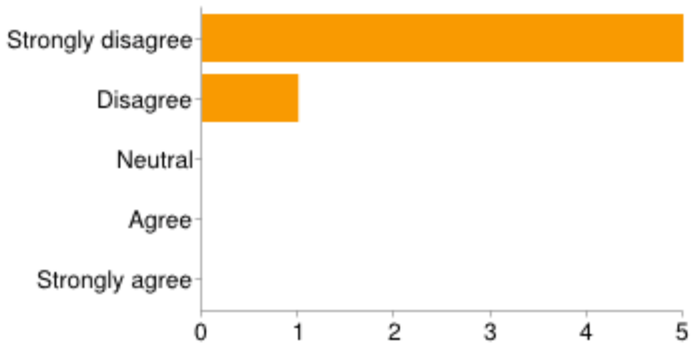
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**2**	33%
Strongly agree	**4**	67%

Qn 3:Terminologies and icons used were clear and easy to understand.



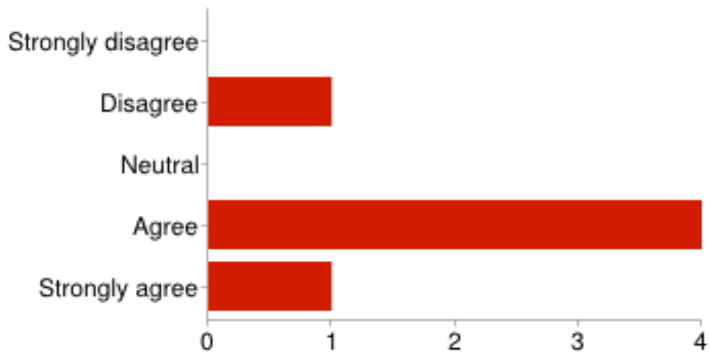
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**6**	100%

Qn 4:The tool was slow and confusing to use.



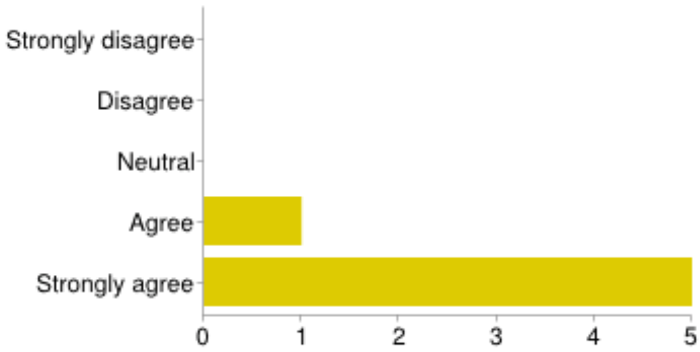
Strongly disagree	**5**	83%
Disagree	**1**	17%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**0**	0%

Qn 5:The prototype provided the necessary functionalities to enable me perform the tasks I was assigned.



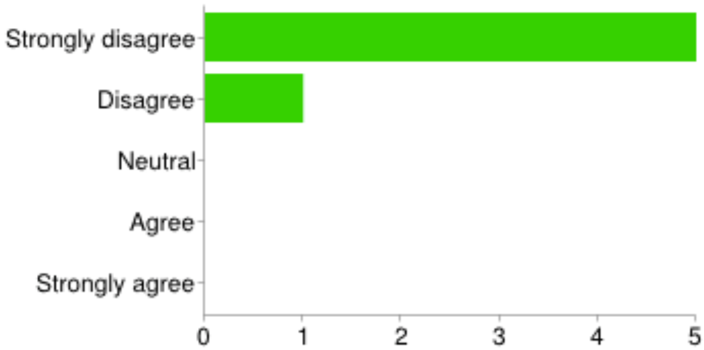
Strongly disagree	**0**	0%
Disagree	**1**	17%
Neutral	**0**	0%
Agree	**4**	67%
Strongly agree	**1**	17%

Qn 6:The simple and intuitive design made it easy to learn how to use the tool.



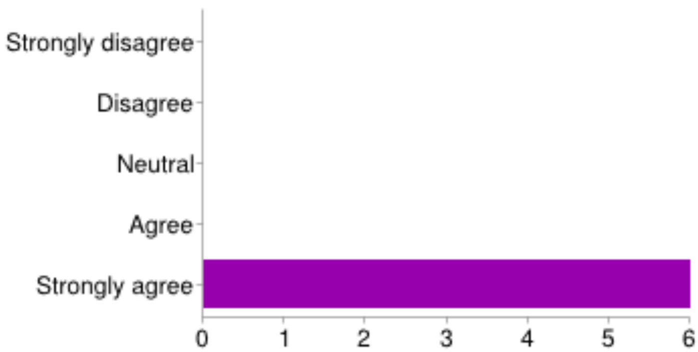
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**1**	17%
Strongly agree	**5**	83%

Qn 7:It was easy to forget how to use the tool if I do not use it in the daily bases.



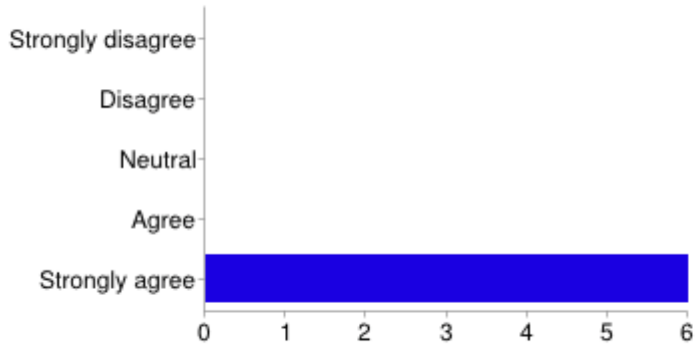
Strongly disagree	**5**	83%
Disagree	**1**	17%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**0**	0%

Qn 8:It will take short time to master the usage of this tool.



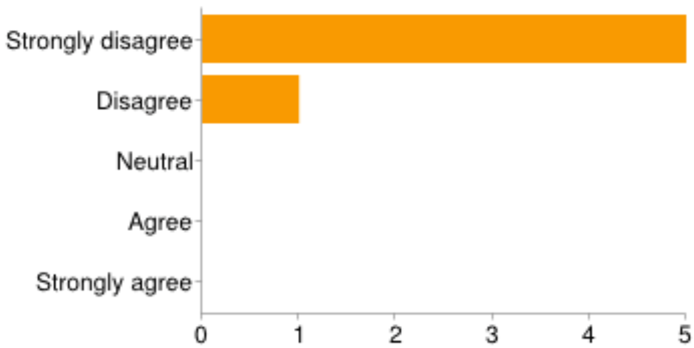
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**6**	100%

Qn 9:The design presents information in an aesthetically pleasing manner



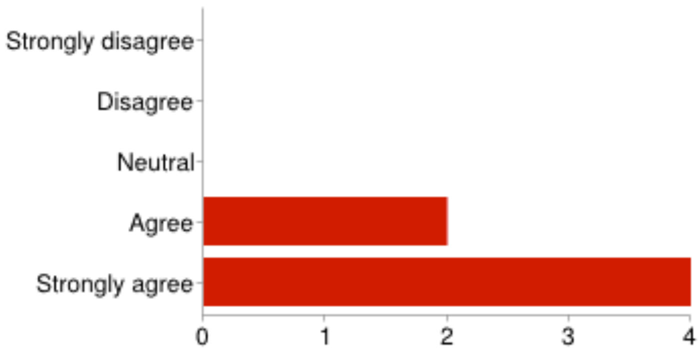
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**6**	100%

Qn 10:I kept on making errors while performing tasks.



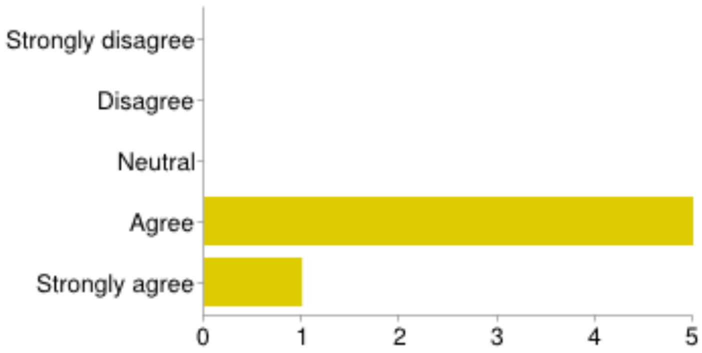
Strongly disagree	**5**	83%
Disagree	**1**	17%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**0**	0%

Qn 11:The tool prevented me from selecting irrelevant variables.



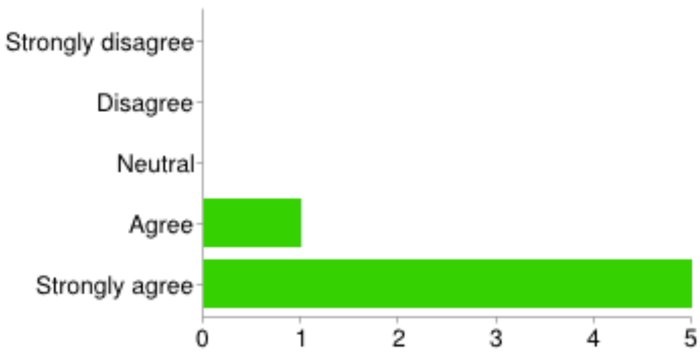
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**2**	33%
Strongly agree	**4**	67%

Qn 12:I think the prototype was consistent, easy to follow how the different parts were connected and easy to navigate through.



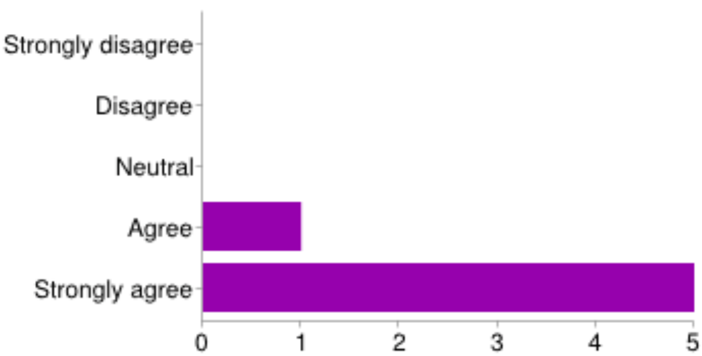
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**5**	83%
Strongly agree	**1**	17%

Qn 13:I enjoyed using this tool.



Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**1**	17%
Strongly agree	**5**	83%

Qn 14:I would like to use this tool in the future.



Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**1**	17%
Strongly agree	**5**	83%

Qn 15: What design features do you think made this design appealing?

- The map and the information that comes up when hovering
- The map feature was very handy.

Qn 16: What design features do you think should be included in order to improve this interface?

- A better arrangement of the content.
- The consistency of the design should be improved.

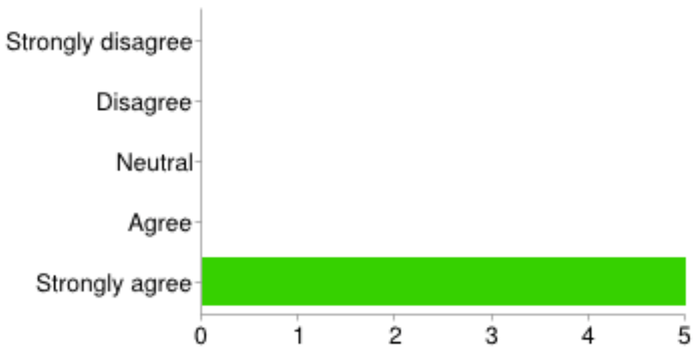
Comment: Please add any comments or recommendations regarding the overall design.

- The consistency of the design should be improved.
- not easy to find the medical region especially on the map, only the city comes up.
- Could not choose more than one variable.
- The map on the map page was not in the center of the page.
- Putting clinical trials on the top page is not a good idea.
- Medical regions are too hidden

Design alternative 3

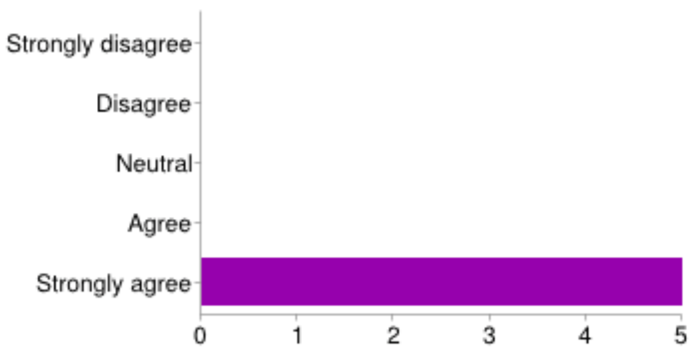
Responses from the Questionnaire.

QN 1:I was able to complete the tasks without difficulties.



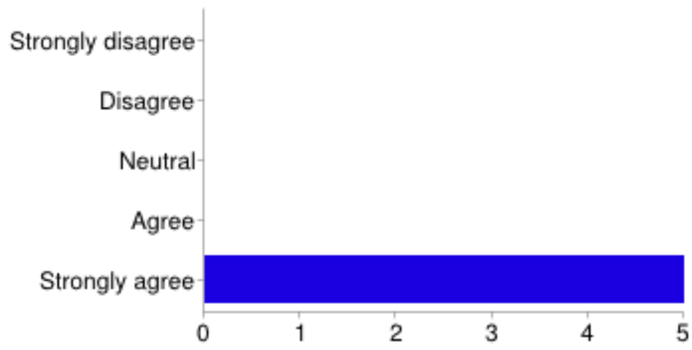
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

QN 2: I was able to complete the tasks without any help.



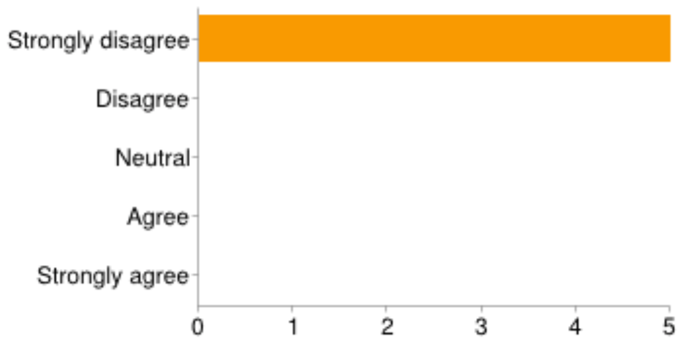
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 3:Terminologies and icons used were clear and easy to understand.



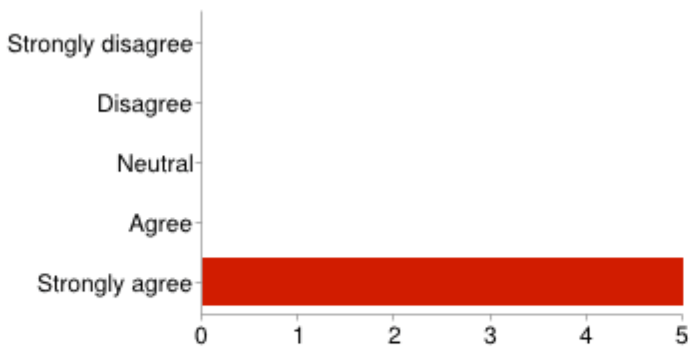
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 4:The tool was slow and confusing to use.



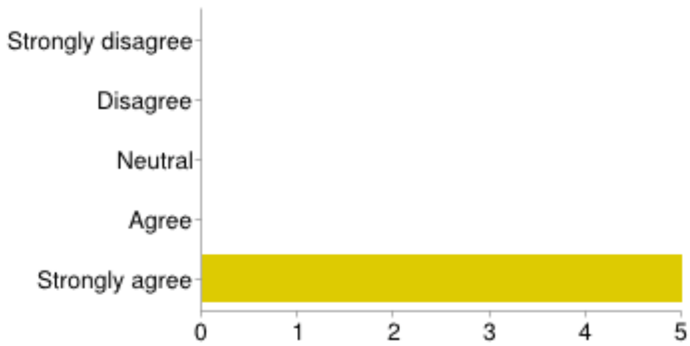
Strongly disagree	**5**	100%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**0**	0%

Qn 5:The prototype provided the necessary functionalities to enable me perform the tasks I was assigned.



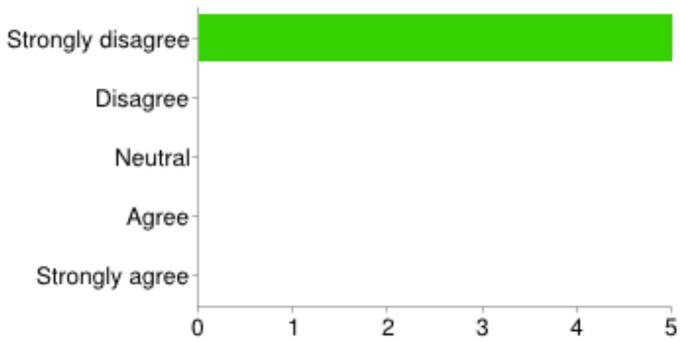
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 6:The simple and intuitive design made it easy to learn how to use the tool.



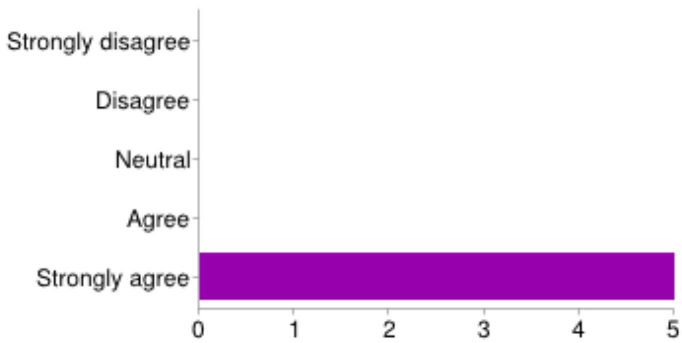
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 7:It was easy to forget how to use the tool if I do not use it in the daily bases.



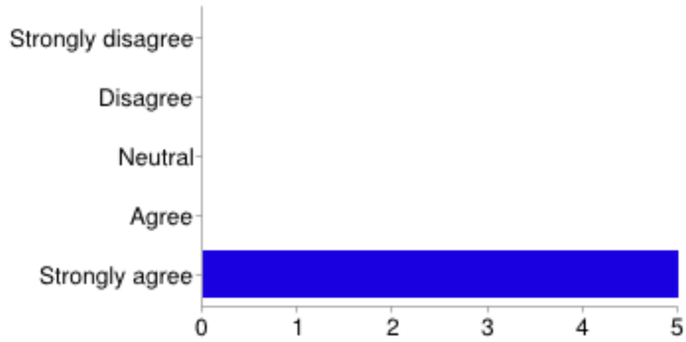
Strongly disagree	**5**	100%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**0**	0%

Qn 8:It will take short time to master the usage of this tool.



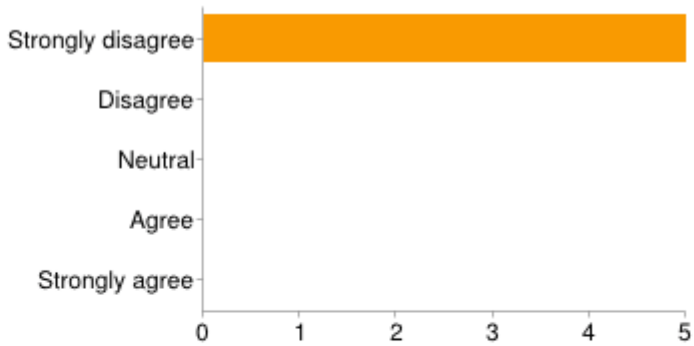
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 9:The design presents information in an aesthetically pleasing manner



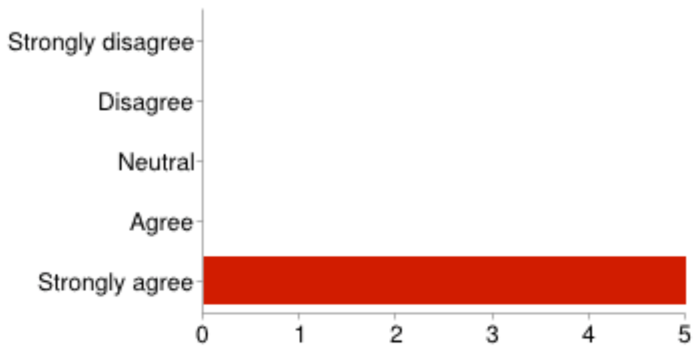
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 10:I kept on making errors while performing tasks.



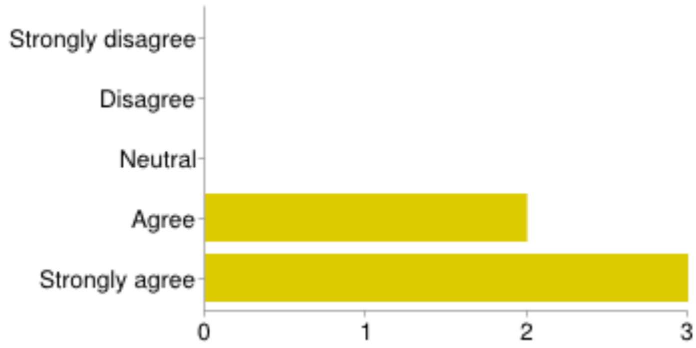
Strongly disagree	**5**	100%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**0**	0%

Qn 11:The tool prevented me from selecting irrelevant variables.



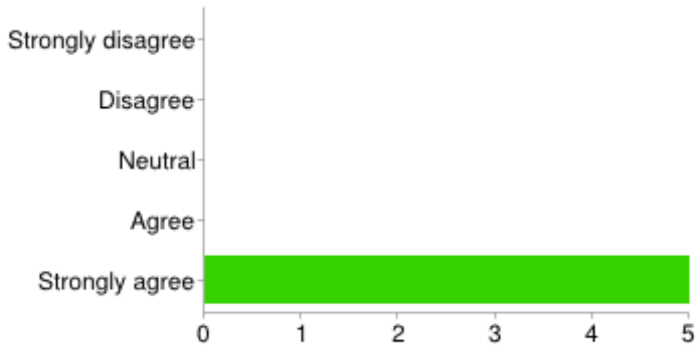
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 12:I think the prototype was consistent, easy to follow how the different parts were connected and easy to navigate through.



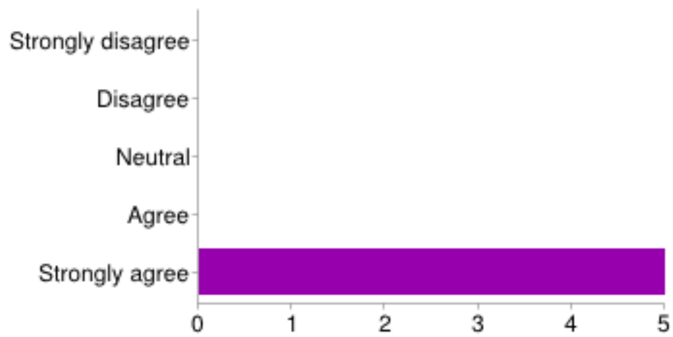
Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**2**	40%
Strongly agree	**3**	60%

Qn 13:I enjoyed using this tool.



Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 14:I would like to use this tool in the future.



Strongly disagree	**0**	0%
Disagree	**0**	0%
Neutral	**0**	0%
Agree	**0**	0%
Strongly agree	**5**	100%

Qn 15: What design features do you think made this design appealing?

- It was easy to find what i was looking for
- I liked the clean design with calm colors.
- The design was clean and it had more things to choose from than the first design
- The images used on the top page fitted very well.
- The way the different parts were arranged on the top page was neat and easy to select on.

Qn 16: What design features do you think should be included in order to improve this interface?

- Possibility to see the data shown on the graphs.
- The functions were very good, but some small things should be done to clean up the looks.
- The search function.

Comment: Please add any comments or recommendations regarding the overall design.

- I did not see why the clinical trial is put on the top page.
- I think the heading is not so interesting.
- Components of the top page are too squeezed together. Removing clinical trials will make space for the remaining ones.
- Some parts were consistent but not all.
- The heading was not the best
- I thought the medical terms was some kind of a medical dictionary. Maybe another name for that label.
- Not very consistent when it comes to the position of the map, human body and the graph. Human body and the graph sections were consistent but not the map.

Appendix 4e: Time tracking - Usability Testing

(Time, in minutes, taken for users to complete each task for each design alternative.)

Below are tables with the times taken for users to perform the set tasks for each of the three parallel prototypes. Each table shows how long each task took for a user to complete.

USER 1

Tasks	Design**1**	Design**2**	Design**3**
Task 1	:01:06	:00:48	:00:32
Task 2	:00:46	:00:45	:00:25
Task 3	:01:01	:01:03	:00:30
Task 4	:00:59	:00:50	:00:44
Task 5	:01:01	:00:43	:00:42
Total	:04:53	:04:09	:02:53
Average	**:00:59**	**:00:50**	**:00:35**

USER 2

Tasks	Design**1**	Design **2**	Design**3**
Task 1	:00:59	:00:53	:00:42
Task 2	:00:52	:00:41	:00:31
Task 3	:01:09	:00:59	:00:31
Task 4	:00:58	:00:55	:00:40
Task 5	:01:01	:00:44	:00:41
Total	:04:59	:04:12	:03:05
Average	**:01:00**	**:00:50**	**:00:37**

USER 3

Tasks	Design**1**	Design**2**	Design**3**
Task 1	:00:42	:00:36	:00:34
Task 2	:00:32	:00:34	:00:24
Task 3	:00:56	:00:59	:00:31
Task 4	:00:44	:01:01	:00:29
Task 5	:00:58	:00:41	:00:36
Total	:03:52	:03:51	:02:34
Average	**:00:46**	**:00:46**	**:00:31**

USER 4

Tasks	Design1	Design2	Design3
Task 1	:00:40	:00:41	:00:29
Task 2	:00:31	:00:38	:00:22
Task 3	:00:57	:00:58	:00:36
Task 4	:00:45	:00:40	:00:30
Task 5	:00:59	:00:40	:00:31
Total	:03:52	:03:37	:02:28
Average	**:00:46**	**:00:43**	**:00:30**

USER 5

Tasks	Design**1**	Design**2**	Design**3**
Task 1	:00:59	:00:51	:00:34
Task 2	:00:42	:00:49	:00:30
Task 3	:01:02	:01:03	:00:32
Task 4	:00:46	:00:52	:00:30
Task 5	:01:01	:00:48	:00:32
Total	:04:30	:04:23	:02:38
Average	**:00:54**	**:00:53**	**:00:32**

AVERAGE TIME USED

User	Design**1**	Design**2**	Design**3**
U1	:04:53	:04:09	:02:53
U2	:04:59	:04:12	:03:05
U3	:03:52	:03:51	:02:34
U4	:03:52	:02:37	:02:28
U5	:04:30	:04:23	:02:38
Total	:22:06	:19:12	:13:38
Average	**:04:25**	**:03:50**	**:02:44**

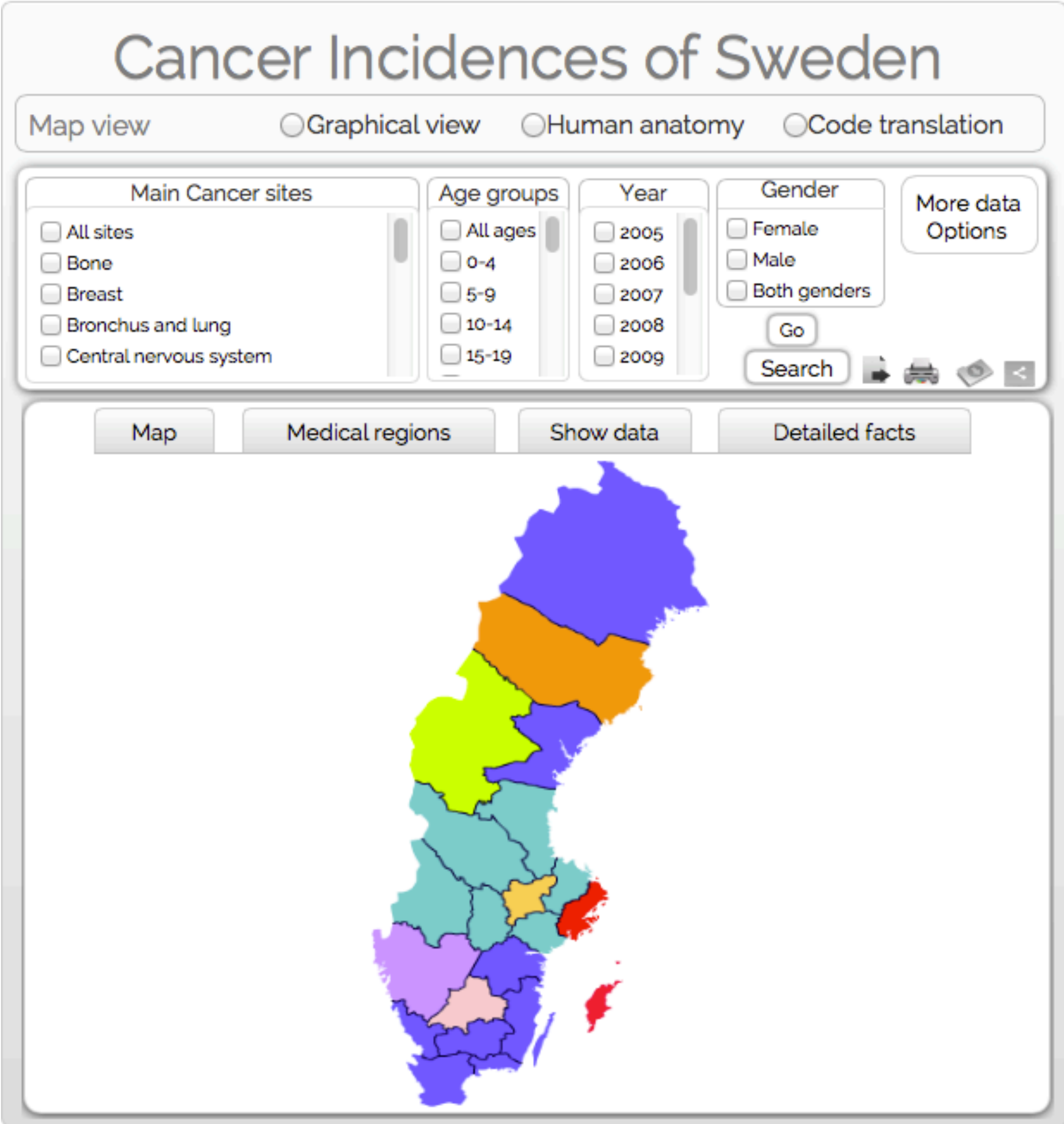
Appendix 5: Final prototype

The design features of the final prototype.

1. The Top page: presents four different options of viewing the data.
From a selected view, users will be able to view the data in the reports that is relevant for that particular view. The selection of the variable is on the top content box, and it remains constant throughout the selected view. The search, hover and click functionalities help users to quickly view the data they want without having to browse through several report pages.

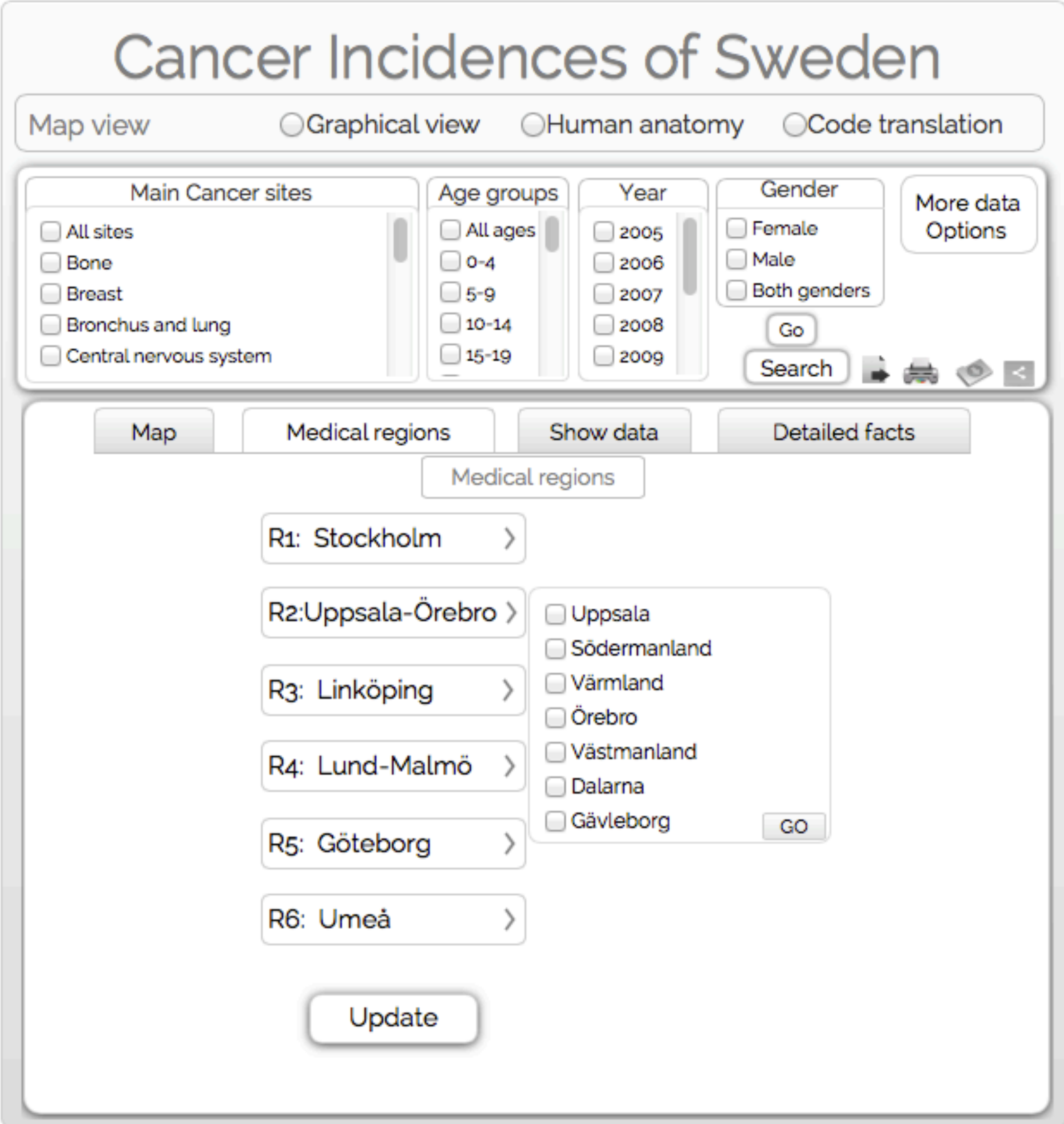
Top/Main page

2. Map View:

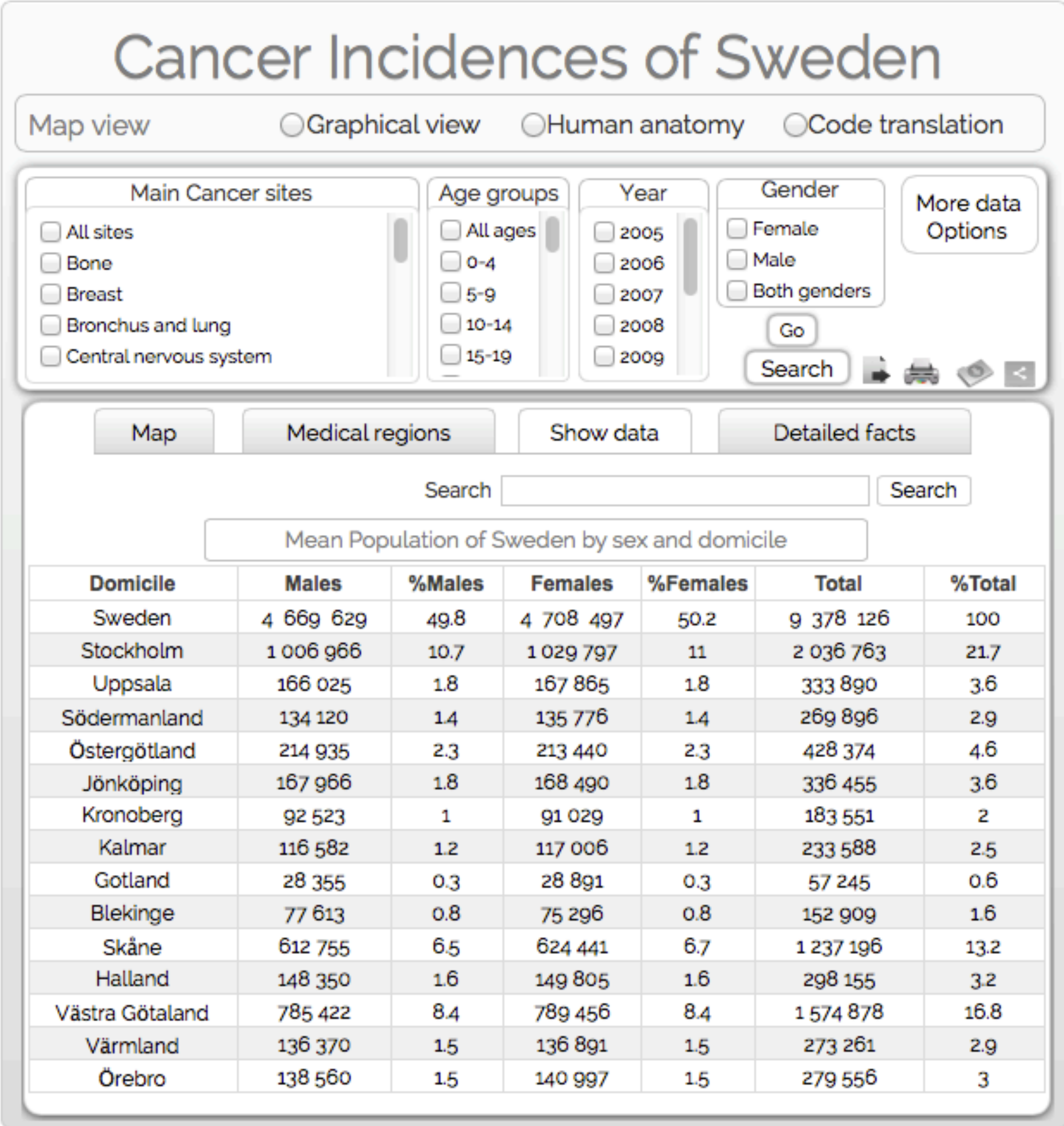


Map View

The sub-pages found under the '*Map view*' are Medical regions, Show data and Detailed facts.

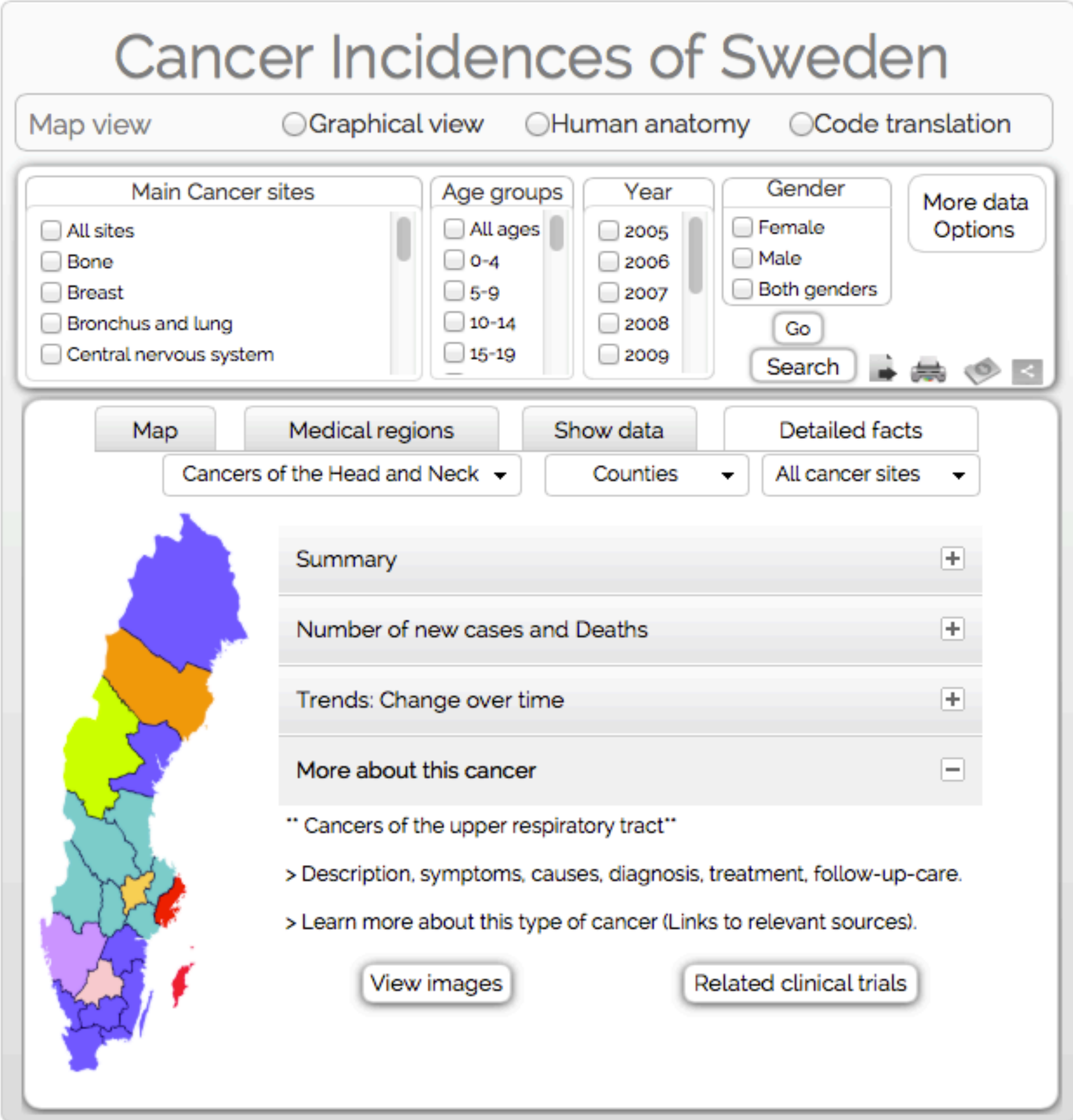


Medical regions



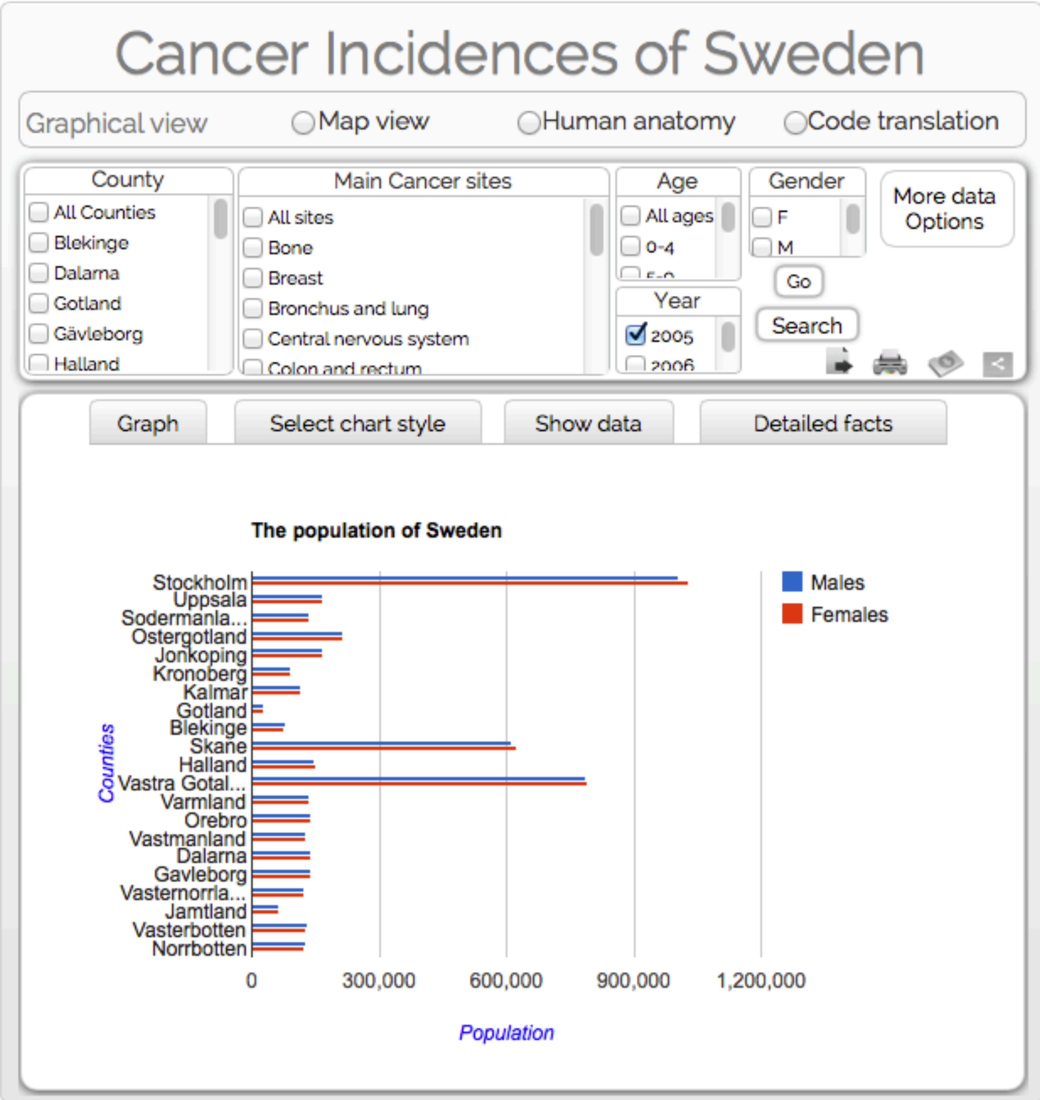
Domicile	Males	%Males	Females	%Females	Total	%Total
Sweden	4 669 629	49.8	4 708 497	50.2	9 378 126	100
Stockholm	1 006 966	10.7	1 029 797	11	2 036 763	21.7
Uppsala	166 025	1.8	167 865	1.8	333 890	3.6
Södermanland	134 120	1.4	135 776	1.4	269 896	2.9
Östergötland	214 935	2.3	213 440	2.3	428 374	4.6
Jönköping	167 966	1.8	168 490	1.8	336 455	3.6
Kronoberg	92 523	1	91 029	1	183 551	2
Kalmar	116 582	1.2	117 006	1.2	233 588	2.5
Gotland	28 355	0.3	28 891	0.3	57 245	0.6
Blekinge	77 613	0.8	75 296	0.8	152 909	1.6
Skåne	612 755	6.5	624 441	6.7	1 237 196	13.2
Halland	148 350	1.6	149 805	1.6	298 155	3.2
Västra Götaland	785 422	8.4	789 456	8.4	1 574 878	16.8
Värmland	136 370	1.5	136 891	1.5	273 261	2.9
Örebro	138 560	1.5	140 997	1.5	279 556	3

Show data possibility



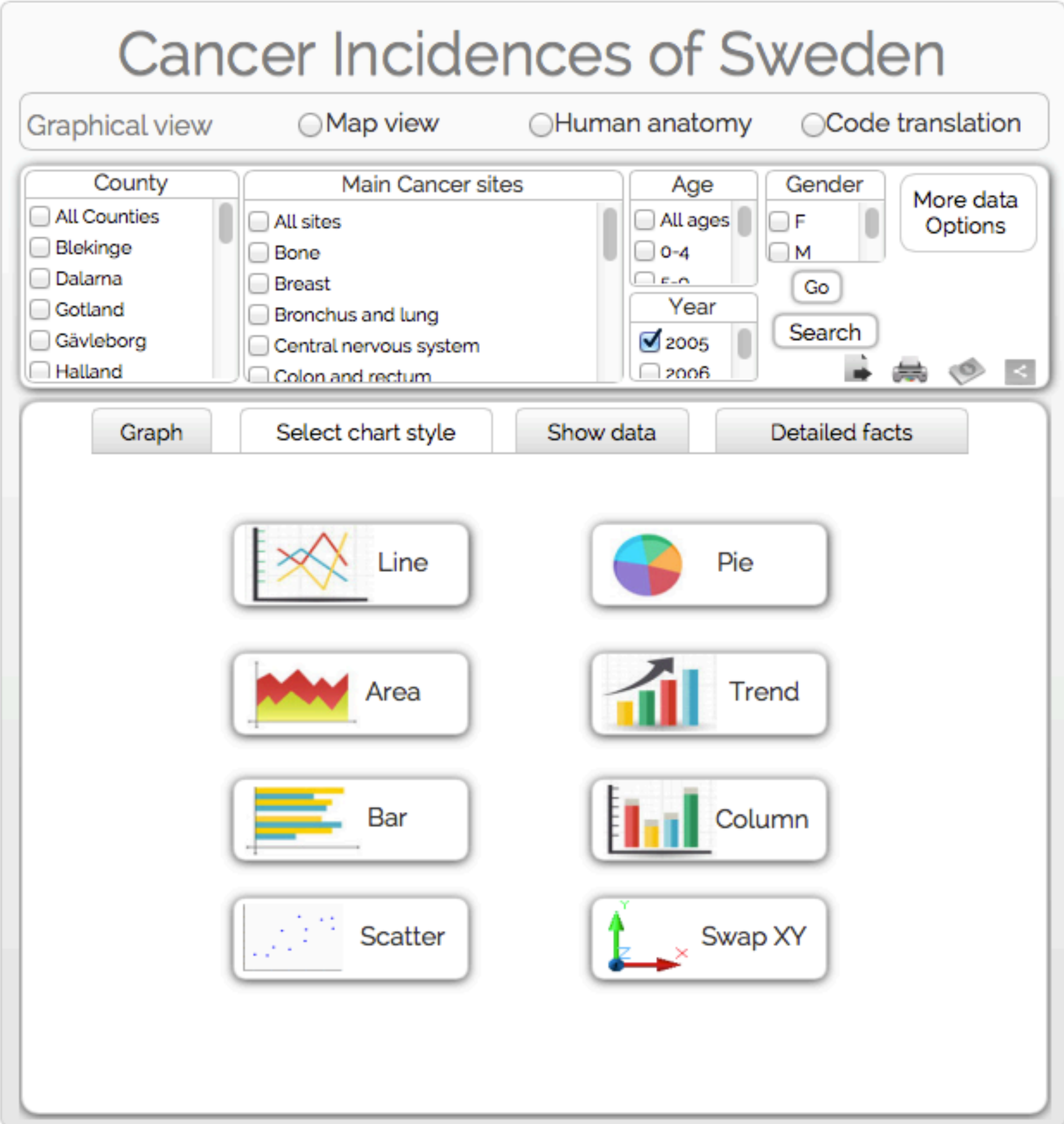
Detailed facts

3. Graphical View:

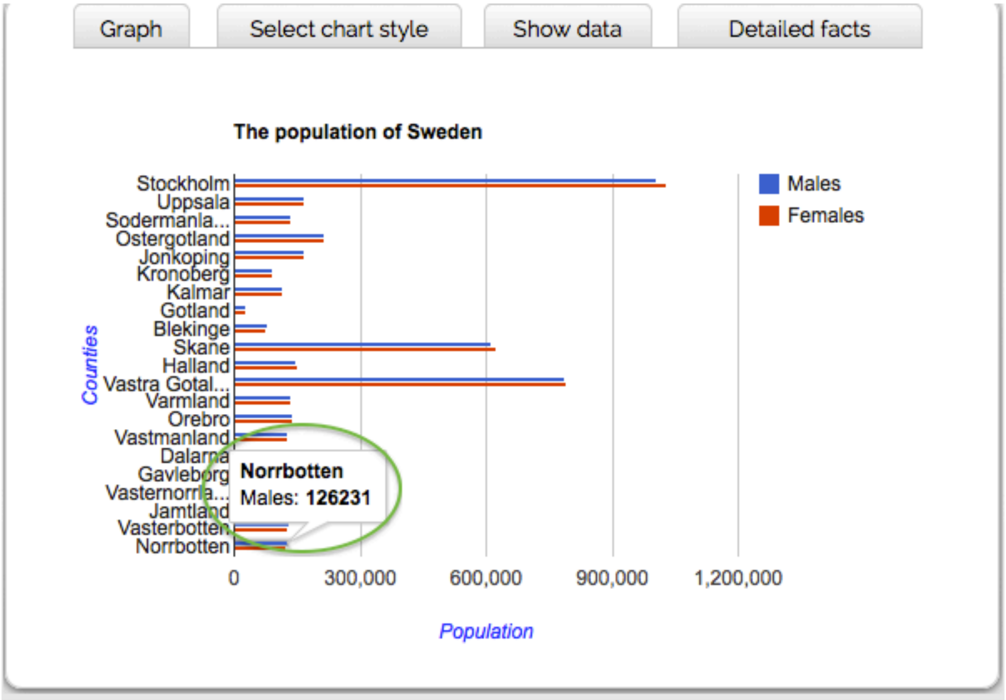


Graphical View

The sub-pages found under the '*Graphical View* 'are Select chart style, Show data and detailed facts. The *detailed facts* and *show data* follow the same concept as shown on the '*Map view*'.



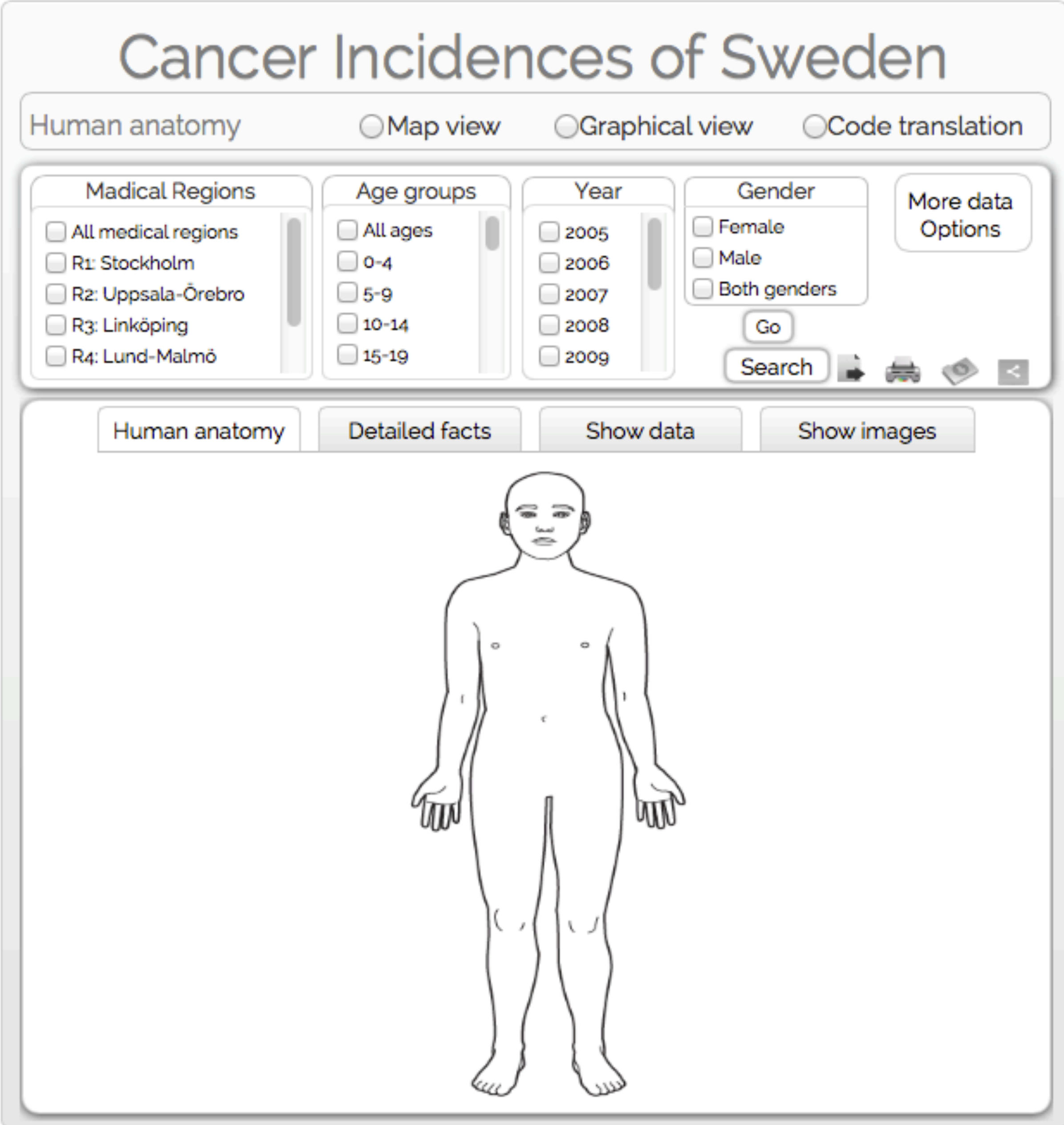
Select chart style



Graph showing the male population of Norrbotten

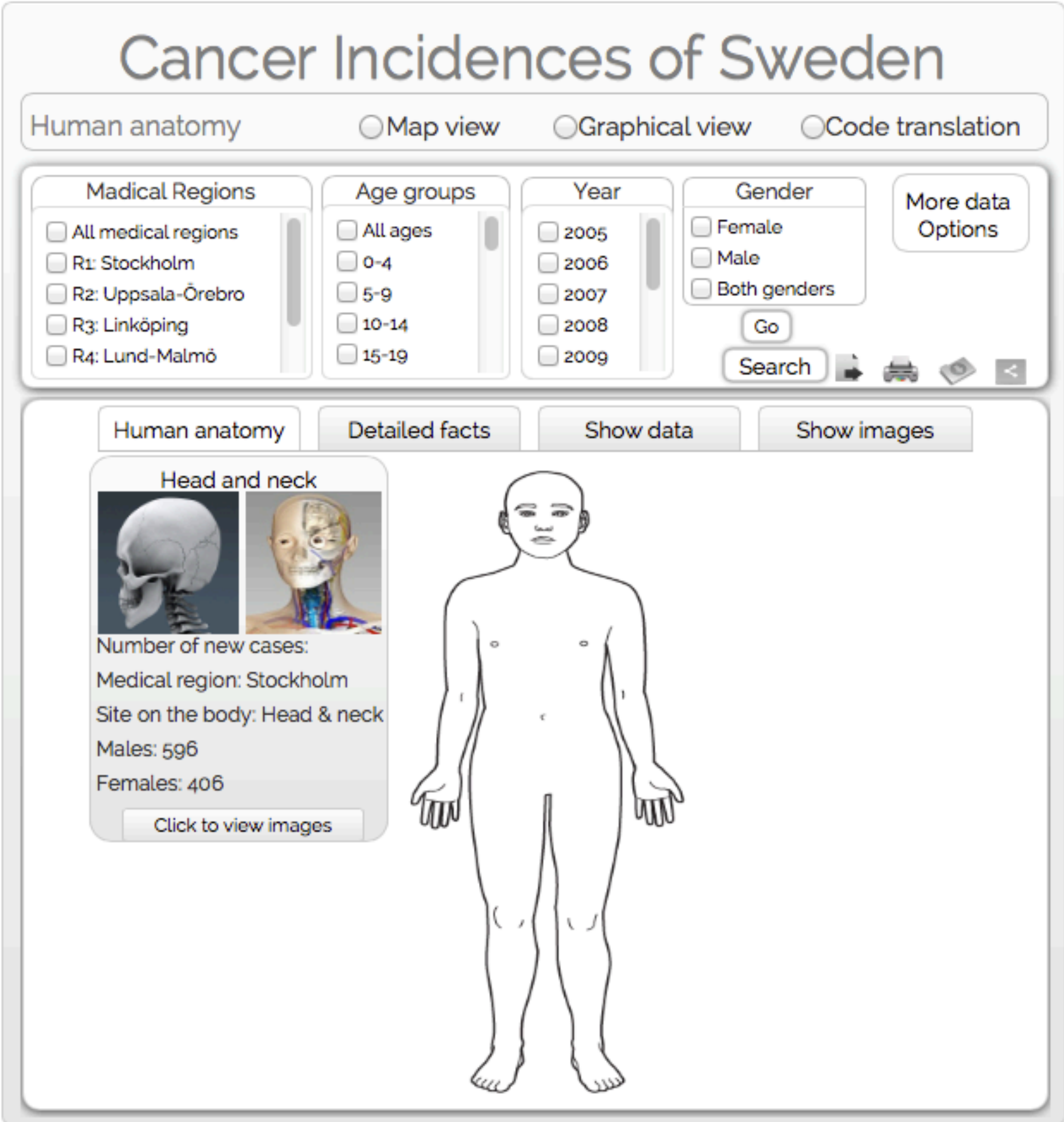
Users can interact with the graphs so as to have a closer look at the details shown. The example below shows the male population of Norrbotten when *population* was the chosen variable.

4. Human Anatomy View:

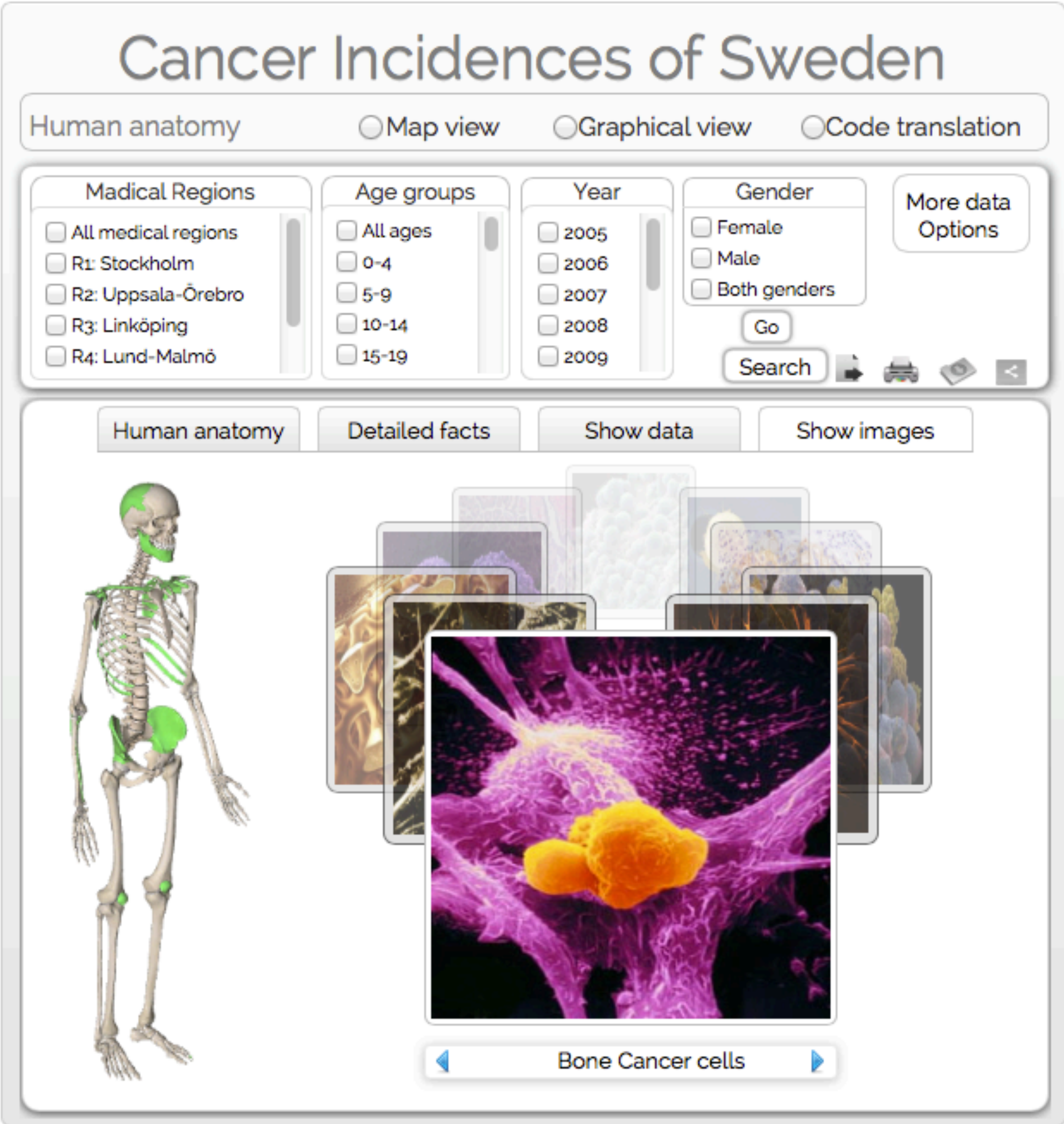


Human anatomy View

The sub-pages found under the '*Human Anatomy View*' are Detailed facts, Show data and Show images. The *detailed facts* and *show data* follow the same concept as shown on the '*Map view*'.



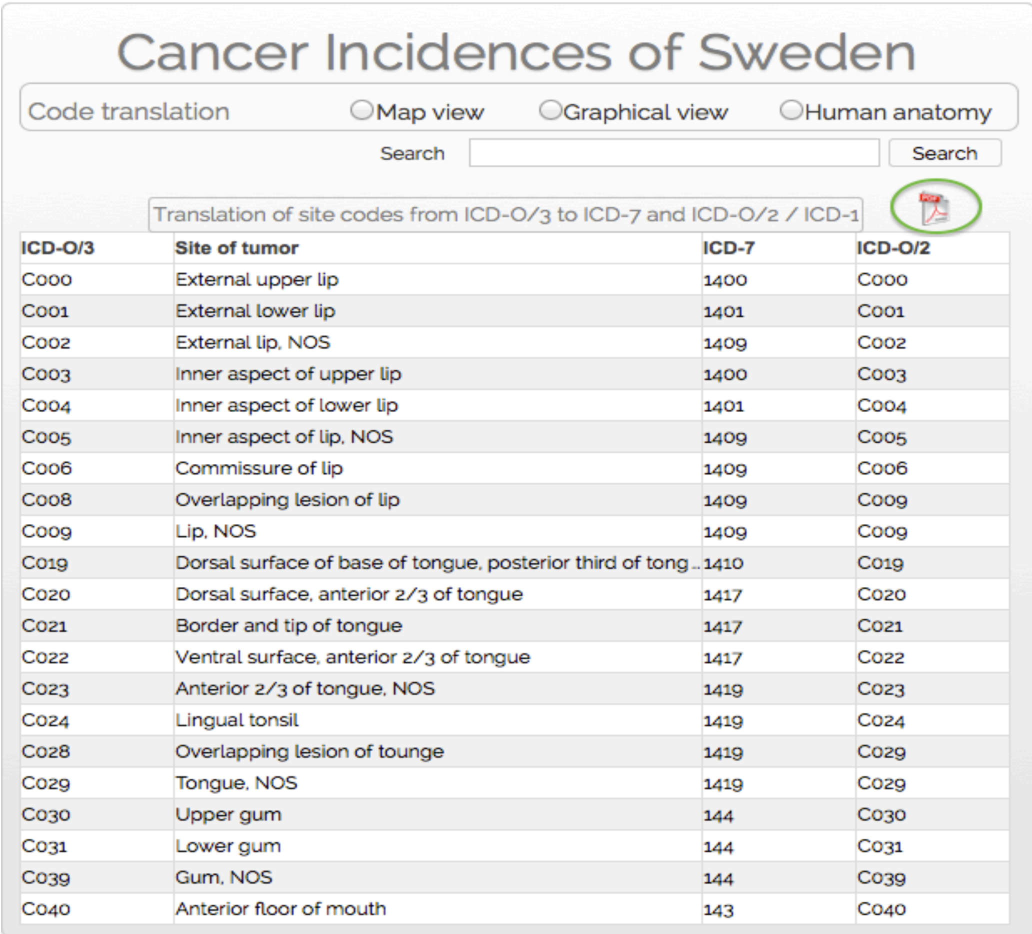
On mouse hover over the head and chest area, the user can see brief information according to the specified variables.



Possibility to view images of the different kinds of cancer.

5. Code Translation:

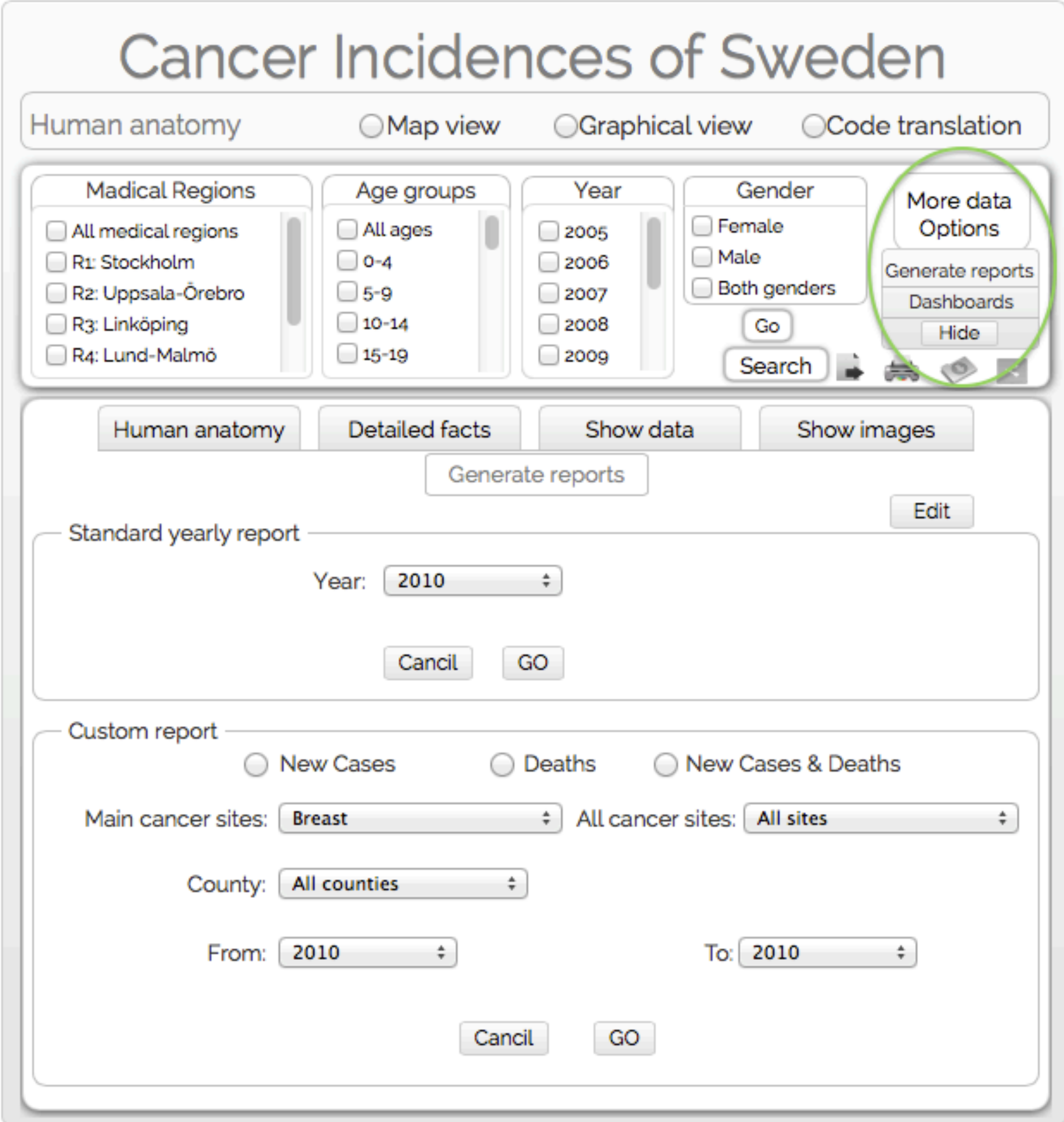
With a search function and PDF view possibility, the *'Code translation'* page contains a list of the codes used in the reports.



ICD-O/3	Site of tumor	ICD-7	ICD-O/2
C000	External upper lip	1400	C000
C001	External lower lip	1401	C001
C002	External lip, NOS	1409	C002
C003	Inner aspect of upper lip	1400	C003
C004	Inner aspect of lower lip	1401	C004
C005	Inner aspect of lip, NOS	1409	C005
C006	Commissure of lip	1409	C006
C008	Overlapping lesion of lip	1409	C009
C009	Lip, NOS	1409	C009
C019	Dorsal surface of base of tongue, posterior third of tong...	1410	C019
C020	Dorsal surface, anterior 2/3 of tongue	1417	C020
C021	Border and tip of tongue	1417	C021
C022	Ventral surface, anterior 2/3 of tongue	1417	C022
C023	Anterior 2/3 of tongue, NOS	1419	C023
C024	Lingual tonsil	1419	C024
C028	Overlapping lesion of tounge	1419	C029
C029	Tongue, NOS	1419	C029
C030	Upper gum	144	C030
C031	Lower gum	144	C031
C039	Gum, NOS	144	C039
C040	Anterior floor of mouth	143	C040

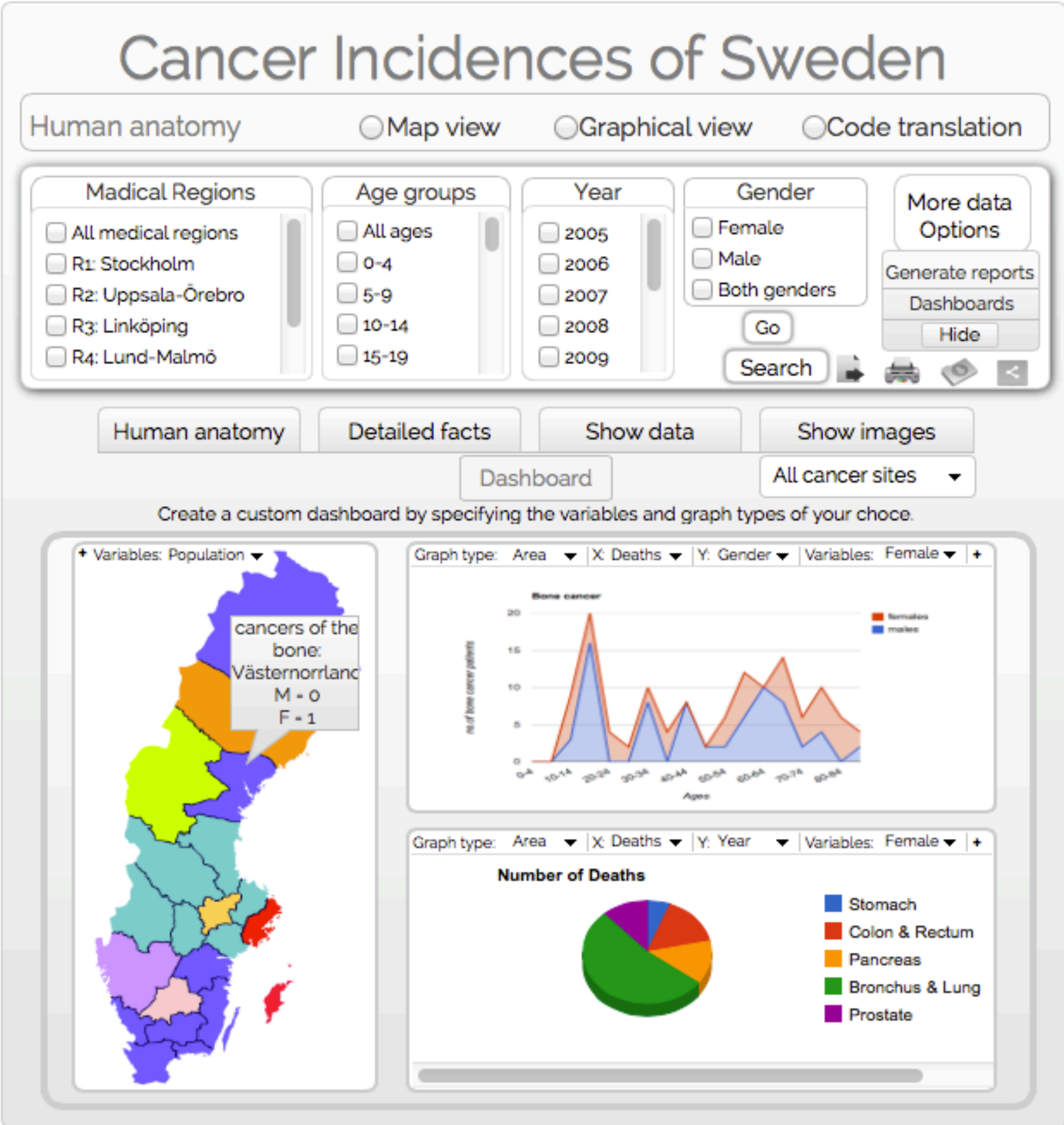
6. More data Options:
This consists of the '*Generate reports*' and '*Dashboard*' features.

***6a) Generate reports**:* Provides the possibility to generate custom reports and the normal yearly reports.



Possibility to generate reports

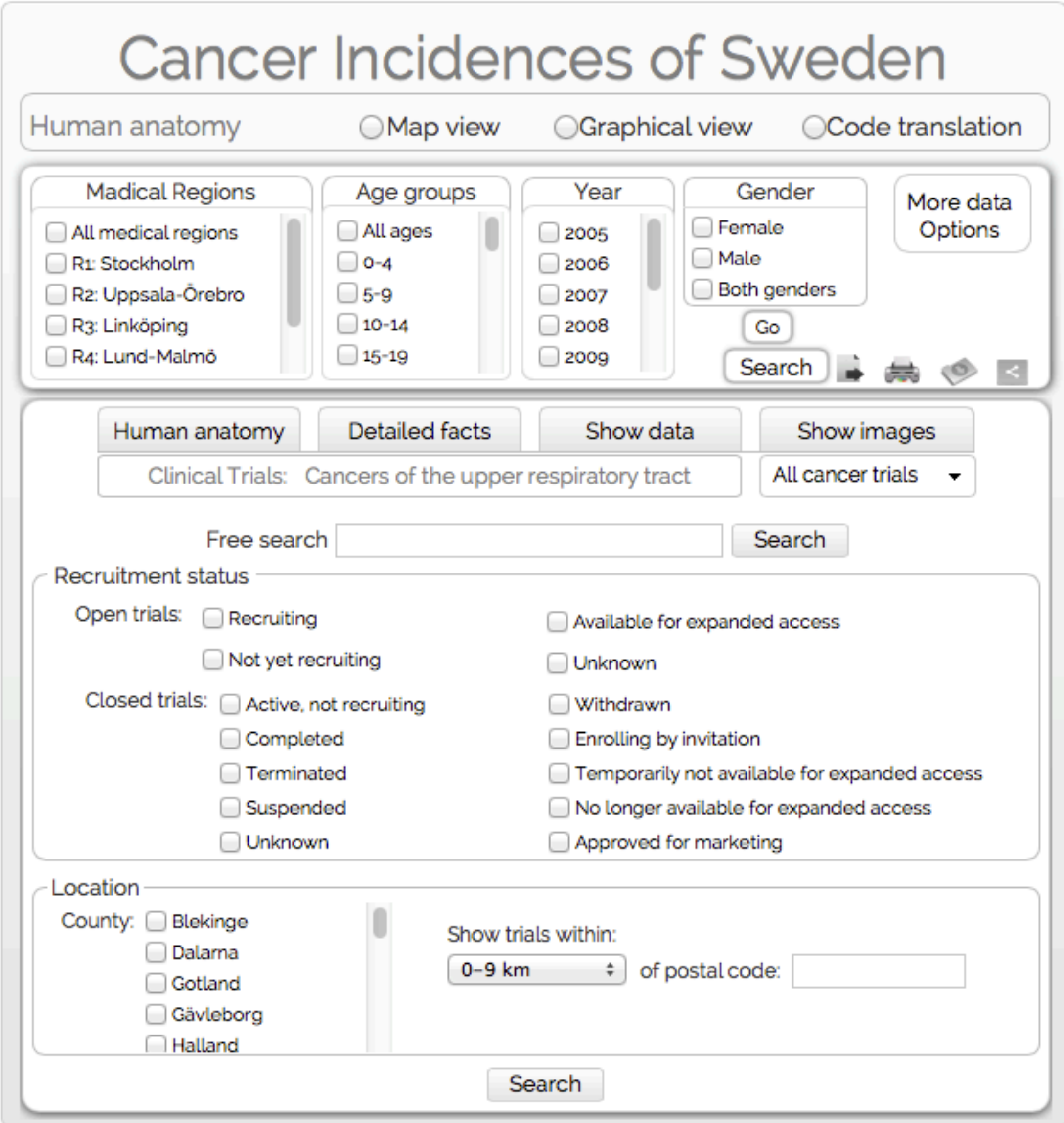
6b) Dashboard: This feature allows users to compare several variables. In addition, users can interact with the map, graph and chart displayed. On hover over a region on the map, the specified variable displays. For example, in the figure below, the variables were: cancers of the bone, Våsternorrland and both genders were specified for the map section of the dashboard.



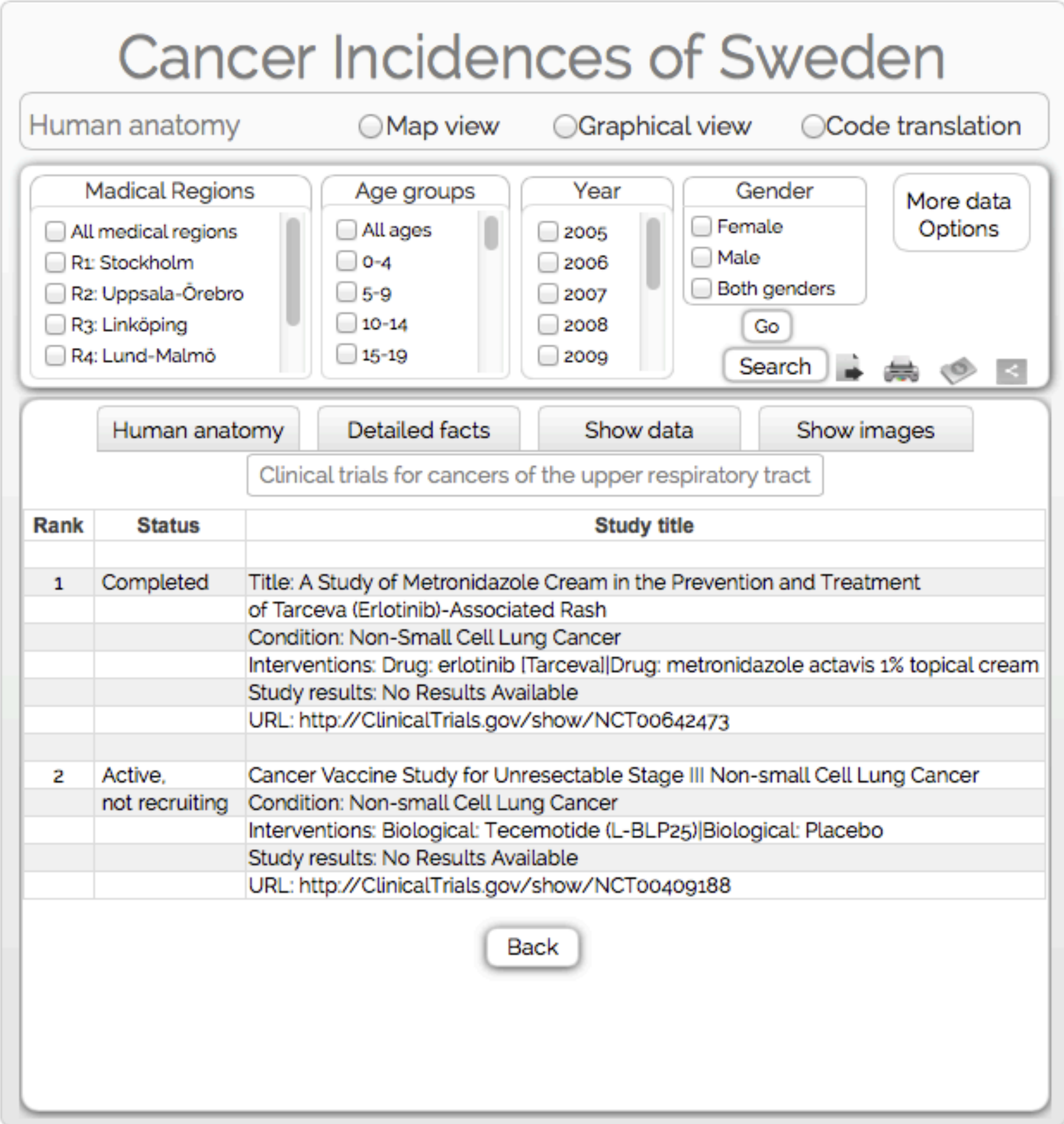
Dashboard

7. Clinical Trials:

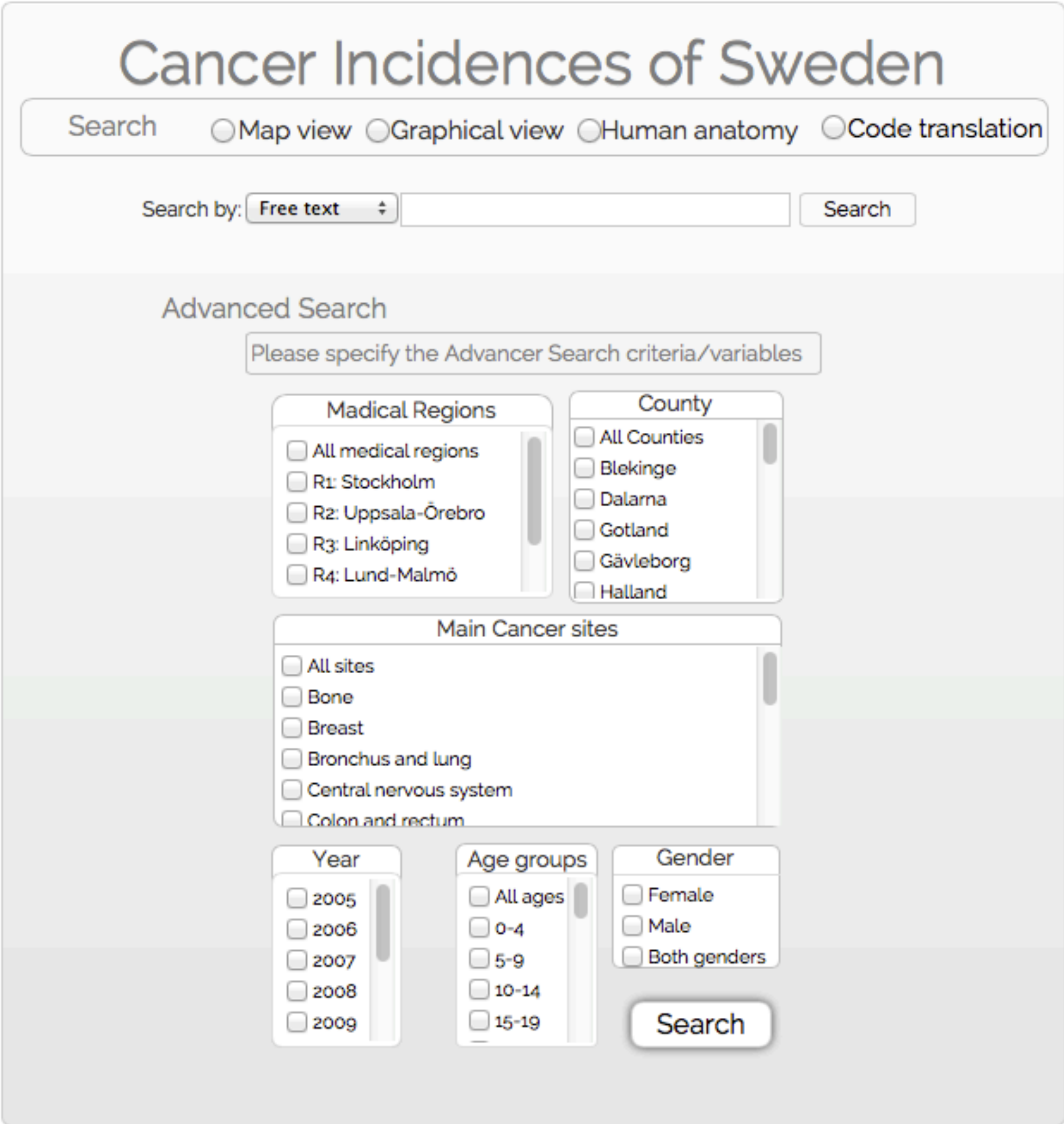
This provides users with a possibility to find out the clinical trials nearest to them. The example on the figures below shows clinical trials relater to cancers of the upper respiratory tract. The search results are obtained from clinicaltrials.gov, a public web based data resource for clinical trials.



Possibility to search for clinical trials



An example of clinical trials search results



The main search function

Appendix 6: Expert Evaluation

Appendix 6a: Expert evaluation form

Heuristic Evaluation Form.

I am a master student in Information Science at the University of Bergen writing a master's thesis about "Human Computer Interaction Design for Data Mining In Cancer Registries". I have designed a prototype for visualizing data mined from cancer registries and will appreciate your help in evaluating this prototype.
The evaluation consists of two parts; Performing some simple tasks by following the current practises and the designed prototype; and Evaluating the prototype following Nielsen's Heuristics where you will be asked to rate the prototype using a scale of 1-10 where 1=Very poor and 10 = Very good. Comments can be added under each heuristic.

H1: Visibility of system status.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

H2: Match between system and the real world.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

H3: User control and freedom.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

H4: Consistency and standards.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

H5: Error prevention.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

H6: Recognition rather than recall.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

H7: Flexibility and efficiency of use.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

H8: Aesthetic and minimalist design.

Very poor	1	2	3	4	5	6	7	8	9	10	Very good

Comments:

H9: Help users recognize, diagnose, and recover from errors.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

H10: Help and documentation.

	1	2	3	4	5	6	7	8	9	10	
Very poor											Very good

Comments:

Appendix 6b:Expert evaluation tasks

HEURISTIC EVALUATION OF THE PROTOTYPE

I am a master student in Information Science at the University of Bergen writing a master's thesis about "Human Computer Interaction Design for Data Mining In Cancer Registries". I have designed a prototype for visualizing data mined from cancer registries and will appreciate your help in evaluating this prototype.
The evaluation consists of two parts; Performing some simple tasks and Evaluating the prototype following Nielsen's Heuristics You will be asked to rate the prototype using a scale of 1-10 where 1=Very poor and 10 = Very good then add a comment where you think something could have been done differently.

TASK 1:
From the map view, find the statistics of breast cancer occurrences in Skåne for both genders and all ages in 2010.

TASK 2:
From the dashboard on the map view, find the number of deaths resulting from cancers of the stomach and pancreas.

TASK 3:
From the graphical view, find the Female population in Stockholm and the Male population in Norrbotten

TASK 4:
From the Human anatomy view, find out how many males and females from Medical region 1 (Stockholm) have been diagnosed with cancers of the head and neck in 2010.

TASK 5:
From the code translation, find the list of 'Lip' cancers.

TASK 6:
From the graphical view, show data represented on the graph.

TASK 7:
From the graphical view, select pie chart style.

TASK 8:
From the Human anatomy view, find at least one image of Lung cancer cells.

TASK 9:
From the map view, generate a custom report showing new cases of cancer of the Lip diagnosed in Stockholm medical region from the year 2006 to 2010.

TASK 10:
From the Human anatomy view, find the detailed facts about the cancers of the upper respiratory tract THEN
- Find the related (of the upper respiratory) clinical trials with closed recruitment *status*, which are completed and Active, not recruiting. The given county is Uppsala and the trials should be within 20km from postal code is 5005.
-Print the retrieved results.

Appendix 6c: List of heuristics briefly described

The TEN HEURISTICS according to Nielsen et.al.(2011).

H1: Visibility of system status:
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

H2: Match between system and the real world:
The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

H3: User control and freedom:
Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

H4: Consistency and standards:
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

H5: Error prevention:
Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

H6: Recognition rather than recall:
Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

H7: Flexibility and efficiency of use:
Accelerators—unseen by the novice user—may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

H8: Aesthetic and minimalist design:
Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

H9: Help users recognize, diagnose, and recover from errors:
Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

H10: Help and documentation:
Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Appendix 6d: Time tracking – Expert evaluation

Below are tables with the times, in minutes, taken by experts to perform the first five tasks when following both, the current ways and the suggested ways through the designed prototype. Each table represents one expert and it shows how long each task took to complete.

EXPERT 1

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:01:19	0:22:00
Task 2	0:03:59	0:42:00
Task 3	0:00:45	0:22:00
Task 4	0:04:48	0:39:00
Task 5	0:01:02	0:34:00
Total	**0:11:53**	**2:39:00**
Average	**0:02:23**	**0:31:48**

EXPERT 2

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:01:56	0:00:25
Task 2	0:02:01	0:00:32
Task 3	0:01:01	0:00:23
Task 4	0:03:59	0:00:23
Task 5	0:03:32	0:00:26
Toral	**0:12:29**	**0:02:09**
Average	**0:02:30**	**0:00:26**

EXPERT 3

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:02:04	0:01:34
Task 2	0:01:21	0:00:48
Task 3	0:00:53	0:01:29
Task 4	0:02:29	0:00:50
Task 5	0:01:23	0:00:15
Toral	**0:08:10**	**0:04:56**
Average	**0:01:38**	**0:00:59**

EXPERT 4

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:03:54	0:01:25
Task 2	0:03:27	0:02:08
Task 3	0:01:13	0:00:53
Task 4	0:05:11	0:00:40
Task 5	0:01:54	0:00:29
Total	**0:15:39**	**0:05:35**
Average	**0:03:08**	**0:01:07**

EXPERT 5

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:01:49	0:00:51
Task 2	0:01:15	0:00:41
Task 3	0:00:29	0:00:39
Task 4	0:03:53	0:00:23
Task 5	0:05:49	0:00:27
Toral	**0:13:15**	**0:03:01**
Average	**0:02:39**	**0:00:36**

EXPERT 6

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:00:35	0:00:59
Task 2	0:01:52	0:01:12
Task 3	0:02:16	0:00:55
Task 4	0:03:52	0:00:53
Task 5	0:01:11	0:00:26
Toral	0:09:46	0:04:25
Average	**0:01:57**	**0:00:53**

EXPERT 7

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:04:21	0:00:42
Task 2	0:02:59	0:01:02
Task 3	0:01:32	0:00:31
Task 4	0:01:09	0:00:21
Task 5	0:00:49	0:00:19
Total	**0:10:50**	**0:02:55**
Average	**0:02:10**	**0:00:35**

EXPERT 8

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:05:23	0:01:38
Task 2	0:03:27	0:02:31
Task 3	0:01:20	0:00:12
Task 4	0:01:14	0:00:15
Task 5	0:00:22	0:00:11
Toral	**0:11:46**	**0:04:47**
Average	**0:02:21**	**0:00:57**

EXPERT 9

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:03:21	0:00:26
Task 2	0:01:49	0:00:39
Task 3	0:00:59	0:00:25
Task 4	0:01:32	0:00:32
Task 5	0:00:24	0:00:21
Toral	**0:08:05**	**0:02:23**
Average	**0:01:37**	**0:00:29**

EXPERT 10

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:01:04	0:00:21
Task 2	0:01:21	0:00:54
Task 3	0:00:59	0:00:58
Task 4	0:00:39	0:00:29
Task 5	0:03:32	0:00:21
Toral	**0:07:35**	**0:03:03**
Average	**0:01:31**	**0:00:37**

EXPERT 11

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:04:19	0:01:28
Task 2	0:02:21	0:02:19
Task 3	0:01:03	0:00:48:
Task 4	0:03:29	0:00:23
Task 5	0:02:30	0:00:21
Toral	**0:13:42**	**0:04:31**
Average	**0:02:44**	**0:01:08**

EXPERT 12

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:03:02	0:00:49
Task 2	0:02:21	0:01:01
Task 3	0:00:59	0:00:21
Task 4	0:02:09	0:00:19
Task 5	0:00:55	0:00:23
Toral	**0:09:26**	**0:02:53**
Average	**0:01:53**	**0:00:35**

EXPERT 13

Tasks	Current ways (Time in min.)	Suggested ways (Time in min.)
Task 1	0:01:04	0:01:01
Task 2	0:01:39	0:01:12
Task 3	0:02:01	0:00:57
Task 4	0:03:49	0:00:59
Task 5	0:01:21	0:00:47
Total	**0:09:54**	**0:04:56**
Average	**0:01:59**	**0:00:59**

AVERAGE TIME IN MINUTES

Experts	Current ways (Time in min.)	Suggested ways (Time in min.)
Expert 1	0:08:13	0:02:49
Expert 2	0:02:44	0:01:08
Expert 3	0:08:13	0:02:49
Expert 4	0:05:29	0:01:59
Expert 5	0:06:24	0:02:16
Expert 6	0:06:10	0:02:11
Expert 7	0:06:13	0:02:12
Expert 8	0:05:48	0:02:04
Expert 9	0:06:30	0:02:17
Expert 10	0:06:01	0:02:08
Expert 11	0:06:13	0:02:12
Expert 12	0:06:08	0:02:11
Expert 13	0:06:09	0:02:11
Total	**1:20:14**	**0:28:28**
Average	**0:06:10**	**0:02:11**

Appendix 6e: Results from expert evaluation

(Follow-up answers)

	SCORES	COMMENTS
HEURISTIC 1	10	-Dashboard hidden
	10	--Sometimes you don't know if the system is responding eg. after clicking- go-
	10	it is not easy to know when to hover the mouse, it the selection has been
	9	-solution: make a small hint by text or highlight on the map
	10	or that human body when the data is updated and ready.
	8	-The system had good information aout what pages you were on.
	8	-Some operations took a long time and no progress indicator was shown.
	9	This left me unsure as to whether or not anything was happening
	9	with the system or if I should try to click the element again.
	9	
	9	
	6	
	7	
HEURISTIC 2	10	-It is a good idea that users can choose to view the data on the map or graph or that human body.
	10	It becomes simple to find the different kinds of cansers.
	10	-Not easy to know all the counties.
	10	-Skeleton ok for bone cancers, maybe 3D body for the rest?
	10	--Good use of icons and images - The map,body and graph are easy to use.
	8	-After looking at the pdf-document this system seems to fix all the problems related to searching and stuff.
	9	Easy to understand. Of course some of the words were hard to understand for me,
	10	who don't have any knowledge about the cancer and cancer cases.

	10	But even though i did not know anything about the field of study,
	10	I felt i understod the most.
	10	-Ikke mye å utsette her, forsto stort sett hva de forskjellige menyene ledet til.
	10	
	8	
HEURISTIC 3	10	-Nice search under Code Translation
	9	-The way to return to the main page were a bit hard, and should maybe be a bit mote clear.
	9	Even though I tought it was easy to "escape" most of the situations.
	10	It was good that the different views are always available to switch between
	10	regardless of what information is showed in the lower part.
	7	
	9	
	10	
	10	
	10	
	7	
	9	
	5	
HEURISTIC 4	10	-Consistent positioning of the navigation bar was very good. -East to know where you are.
	10	Good!
	10	Good use of standard web widgets such as tabs, combo boxes and radio buttons
	10	made me able to navigate easily without any prerequisite knowledge of of the system.
	10	However, I miss an option to reset the search criteria before starting a new search.
	9	It could be cumbersome to remove all the criteria from the last search if there are many of them.

	10	
	10	
	10	
	10	
	9	
	8	
	7	
HEURISTIC 5	10	-No error occured
	10	--No errors at all
	10	-no errors
	9	-Good, not many errors that occured.
	10	Maybe there should be some red start next to forms that have to be filled in to get some information.
	10	-I didn't encounter any errors, so I can only assume that it is good.
	10	
	10	
	10	
	10	
	9	
	10	
	10	
HEURISTIC 6	10	- I think the clinical trials were not easy to find, also the dashboard.
	10	-Difficult to locate clinical trials, but it was a very interesting feature when i finally located it.
	10	-Generate report and Dashboard and clinical trials was difficult to locate. **Suggestion:-Make them tabs
	10	
	10	

	8	
	8	
	8	
	10	
	9	
	10	
	9	
HEURISTIC 7	10	-There was a search function hidden deep in a menu,
	9	that could be easier to find to make experts get to that function quicker.
	10	When I've done the tasks more then once I easily learned the way of doing it.
	10	If I had done different tasks one more time i think i would be going pretty much faster.
	10	-So good!
	7	
	10	
	10	
	10	
	8	
	10	
	8	
	5	
HEURISTIC 8		- Very clean design. - It made me find the data i wanted to see very fast.
	10	-Simple and beautiful design
	10	Nice design, but it could be larger in height and width.
	10	Very minimalist and nice design.

	10	Very much information on some pages, but probarbly needed for those tasks.
	7	I miss clearer visibility of which options are currently selected.
	10	For instance, the selected tab is white and the deselected tabs are light grey. I
	10	t is not that easy to see the difference quickly.
	10	
	9	
	8	
	7	
	7	
HEURISTIC 9	10	-did not experience any errors
	10	-Didn't encounter any errors
	10	
	9	
	9	
	10	
	10	
	10	
	10	
	8	
	5	
HEURISTIC 10	9	-Could not find help and documentation
	10	-It was clear how to use it, so did not need documentation
	10	-5 because i pressed the scale,

	9	I don't think extra documentation is needed if the user gets a proper introduction to
	9	the system and knows the specific terms for medical stuff.
		-The documentation and way to undestand different codes
	10	is much easier in this system than in the old PDF document.
	9	-I miss having tooltips pop up describing the element when i hover over a button or another control.
	10	
	5	
	10	
	7	
	5	

Related Academic Publications

Data Mining in Cancer Registries: A Case for Design Studies

G. Kanza, BA. 1, A. Babic, Ph.D.1,2

[1] University of Bergen/Department of Information Science and Media Studies, Bergen, Norway
[2] Department of Biomedical Engineering, Linkoping University, Sweden

Abstract— **Cancer registries are created, managed and data mined to gain knowledge about long term patient outcomes, effects of medication, clinical factors influencing patients' well-being. Equally important is the insight into the cost effectiveness of cancer treatments, and securing data input from different medical centers and enable competent data analysis and meaningful results. Interest among different user groups (physicians, researchers, health care administrators, policy makers) cerates expectations regarding the results and active role in the development and in interactive use of the information. This paper discusses several design cases in which data mining could be implemented to enable efficient and user friendly knowledge extraction. Three important design cases have been identified following the pathways that the users typically make: 1. ensemble data mining from long term national registries; 2. ensemble data mining form the dedicated clinical web-databases; 3. ensemble distributed data mining and analysis.**

Keywords— **HCI design, big data, data visualization, data mining, cancer registries**

I. Introduction

Long term studies with a high number of subjects are trusted to give statistically meaningful results and a reliable insight into the patient outcomes. Clinical trials have established very strict study conditions and collect and process a wide range of clinical variables to provide answers to the study hypotheses. Registries have been developed using fewer data items, limiting the number of primary and secondary end points, but they gather data from several centers, both nationally and internationally. They allow several periodic analyses and some flexibility in generating reports. However, for many years researchers and physicians had to deal with more than one tool to obtain, analyze and visualize data from the registry. Reports have been published off line: once the gathered data was verified, data sets are sent to bio-statisticians. Reports resulting from those analyses typically come back in a form of a file which could be tedious to follow: having to look at never ending rows and columns of numbers is cognitively a demanding task that could be eased by human computer interaction methodology. A hypothesis could be made to explore how a smart design for visualizing could benefit mining the data from what is usually called big data [1]. The notion of big data stands for the combination of various clinical data, images, and textual information, all of which is the reality of the patient management, clinical research and medical education. We will look at the example of Swedish Cancer Registry and the report based on it (Design Case 1).

In addition, we will look at specialized web-sites offering information regarding registries covering a great number of clinical indications and outcomes, study duration, data types, clinical centers involved, study status. At best those sites offer some information retrieval function and no possibility for data mining. A good example of a very well established web-site is *ClinicalTriasl.gov* that is visited by clinicians, researches, industries, and patients interested in the clinical trials (Design Case 2).

Cancer registries, and web-sites containing information about cancer related studies, are subjects of change as new data brings new information and holds insights into the patient population, disease treatment, and changing patterns. It is one of the challenges of data mining to design a distributed system in which users would contribute their data and track those changes (Design Case 3).

II. Research Framework and Methods

There is a large amount of clinical and administrative data that healthcare information systems produce for each patient in different care processes. However sufficient knowledge discovery methods can be developed to retrieve information and implicit knowledge about evidence-based care processes and patient treatments. When statistical methods are combined with data mining techniques, reuse of previous experiences can be utilized in order to improve the clinical guidelines for treatments and services [2].

Table 1 presents two main approaches to the data mining offered by the statistics and artificial intelligence

Table 1 Overview of the data methodology.

MULTIVARIATE STATISTICS	Descriptive Statistics, Cluster Analysis Discriminant Analysis, Regression Analysis
ARTIFICIAL INTLLIGENCE	Decision Trees, Neural Networks, Bayesian Rules, Fuzzy Logics, Rough Sets

An excellent example of the data mining systems implemented in Java, the WEKA system [3] is a collection of machine learning algorithms useful for extracting information from large databases. It is an open source system executable on different platforms. It could provide data in two main ways. The first is by allowing the possibility of loading data from databases, files and Universal Resource Locators (URLs) with the help of supported formats. The second is the possibility to generate data from artificial data sources where the generated data can be edited manually with a dataset editor.

Applications of data mining in a specific domain such as cancer have special requirements, users expect a certain kind of result such as optimal treatment or a survival rate. Methods that are capable of extracting knowledge in forms closer to human perception are those that induce decision trees, classification rules. However, often and more used are statistical methods due to the robustness and validity which are not transparent. Clinical areas with successful applications are numerous and growing [4,5], however more research should be done to aid users by designing the interactions that would provide transparency and clarity of data mining steps. Therefore, the research need to consider work flow and in particular the tasks, procedural steps, organisations and people involved in managing and analysing cancer registries. Resulting patterns such as resource, data, and interaction patterns will be assessed for their potentials to influence the design process.

An overview of the research described in this paper is illustrated in the Figure 1. The main goal was to explore the clinical data and indication for analysis, as well as to understand clinical and all other related work practices. By further looking at how data mining is conducted currently a set of design features could be identified (Figure 1).

The research has been focused on the system design (objective, elements, services, workflow) and interactions (functions, user behaviours, workflow) [6]. The resulting, new automatic, user-friendly system for data mining from the clinical registries is to be tested and evaluated by various expected user groups such physicians, health care administrators, and potentially medical students. The design science knowledge has been instrumental in this study, and In particular the seven design guidelines presented by Hevner et al. (2004) [7]: *design as an artifact*, *problem relevance, design evaluation, research contributions, research rigor, design as a search process, and communication of research*.

From the various evaluation methods proposed by Rogers, Sharp, and Preece (2011) [8], this research employs the *analytical evaluation* where experts will be involved in the evaluation process.

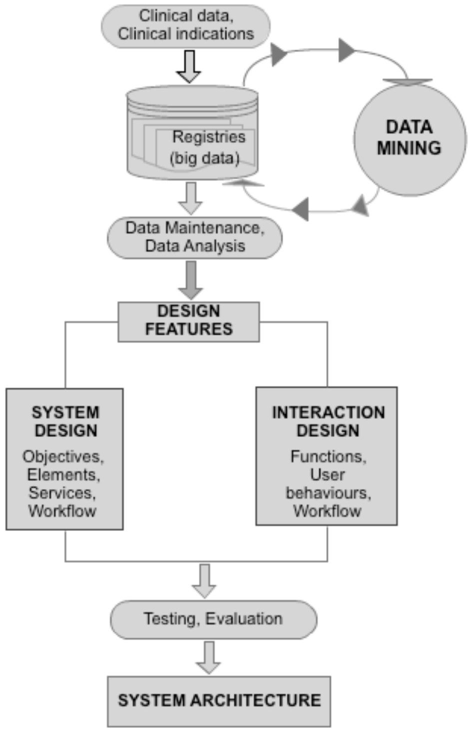


Fig.1 Design steps in data mining.

The design should also look at the number of quality issues and how they could resonate with the users, both medical and administrative. Table 2 lists the quality aspects such as appropriateness of the data, way the results and knowledge could be presented and validation issues.

Table 2 Aspects of quality assurance of importance for data mining.

DATA	Is the data representative, reproducible, of good quality?
PATTERNS	Are the patterns presentable, understandable, easy to interpret and reason about?
KNOWLEDGE	What is feasible/ which is the best way to extract and present the knowledge: Regression and/or classifications formulae, decision tree, neural networks, statistical tables?
VALIDATION	Can we validate methods? Are results comparable to the published ones?

Design Studies

There are three major cases of design to be considered. They are resulting from a pre-study in which we looked at three most typical ways of working with the cancer registries: 1. obtaining in depth analysis, 2. conducting meta analysis of web-based resources, 3. working with the distributed databases in different locations.

Design Case 1: Data mining from long term national registries

The Swedish Cancer Registry [9] has been a data source for epidemiologists and clinicians for over fifty years now. This data is used both at national and international levels to provide an overview of the country's cancer incidences. Researchers, physicians and other medical staff continuously use this data to study the disease in order to improve the treatments that already exist while aiming at finding its cure [9, 10]. Currently, extracting data from the registry is done by sending a request to the cancer registry (statisticians) who retrieves the data, analyses it and sends back tables, graphs back to researchers and physicians who analyze, draw conclusions, and suggest additional analyses. This process is time consuming as it involves several stages starting with often repeated requests and corresponding, resulting sets of results. Each iteration could take up to few weeks. In spite of already well established routines, there are times when doctors may want to get an original solution for a problem at hand. *For example: A research and administrative group wants to do a series of investigations focused on the intersection of aging and cancer, a study should discover information about treatment efficacy, non-compliances, tolerance and effects of co-morbidities in the treatments of elderly colorectal cancer patients.* In this case, data mining should consider different age groups, but that may cause problems since the age is measured purely chronological in most contexts. Another challenge might be to conduct a long term follow up where other populations registries needs to be followed- up and their data combined.

Design Case 2: Data mining form the dedicated web-databases to make their resources available to the users

ClinicalTrials.org is an international registry with 185 countries participating with their own data in the studies carried out for a number of disease indications. In this on-line registry, there is currently a total of 716 cancer studies registered from Sweden, which involve international collaborators from different institutions. Of these, 230 studies are still open and 486 closed.

Institutional commitment to so many national and international studies shows a strong intention to do research and assistance in making bigger databases in which population representative and meaningful patterns could be extracted. The sense of community, collaborative work forms are important design factors to explore. The Table 3 shows just a few registries with a long duration and a high enrollment.

Table 3 Examples of currently running colorectal cancer registries .

STUDY TITLE	TO ENROLL	COLLABORATORS	TIME
Epidural Versus Patient-controlled Analgesia for Reduction in Long-term Mortality Following Colorectal Cancer Surgery	300	-Örebro University, Sweden - University Hospital, Linköping	Mar 2011 to Dec2018
The Northern-European Initiative on Colorectal Cancer	66000	Norwegian Department of Health and Social Affairs - Maria Sklodowska-Curie Memorial Cancer Center, Institute of Oncology - Erasmus Medical Centre, Dep. Of Gastroenterology, Rotterdam, The Netherlands - Landspitali University Hospital - Uppsala University Hospital, Dep. Of Surgery, Uppsala, Sweden - Karolinska Institutet - Riga Eastern Clinical University Hospital, Riga, Latvia - Memorial Sloan-Kettering Cancer Center - Harvard School of Public Health	May 2009 to Jun 2026
Rectal Cancer And Pre-operative Induction Therapy Followed by Dedicated Operation. The RAPIDO Trial	885	- University Medical Centre Groningen -Karolinska University Hospital - Leiden University Med. Center -Uppsala University Hospital -Dutch Cancer Society	Jun 2011 to Jun 2016

Design Case 3: Distributed data mining data.

Clinically based information systems, patient records and registers are designed for recording data more than reporting data. In normal clinical settings, there are several sources of information that could be eligible and useful for exploring clinical hypotheses (Figure 3).

Data mining technique and machine learning methods can bring the facility to discover patterns and connections (hidden knowledge) within the medical databases. Therefore, it is important will utilize the data by:

- integrating information coming from various database covering clinical, research, economic, and other aspects of patient management;
- outlining recommendations for users on how to move across platforms and systems.

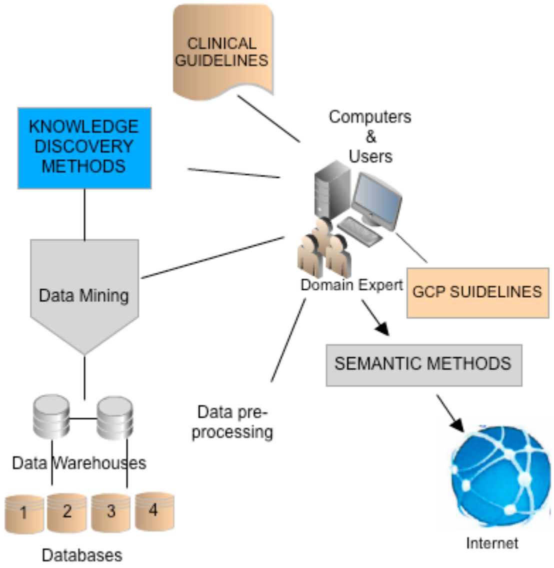


Fig.3 Design points in distributed data mining.

III. DISCUSSION AND CONCLUSION

Here presented design research aims at transforming the current state of interacting with cancer registries to a better and user friendlier stage. User experiences in working with the data sources are different, but demanding in several ways. The most detailed clinical knowledge is being extracted *off line* and with a help of a statistician. A quality checks performed throughout, do guarantee a reliable facts representative of the patient population. However, there is little flexibility in extracting knowledge, for example, it may not be simple to learn about *just* the mortality in a certain age group given a selected treatment. Automated systems would allow to get several interesting comparisons among selected patient groups. Communication in real time would allow testing hypotheses, identifying special cases which are invisible within the whole population. Ad-hoc searches could be performed to confirm findings reported in the literature in a cost efficient manner.

Retrieving information from the dedicated web-sites results in the meta-knowledge of the studies or conditions that physicians are interested in; some sites offer final and periodic reports, but this could be furthered by semantic methodology that would look deeper into the content. In the cases when there are several database covering different clinical areas (*cancer and palliative care*, for example), user is in even bigger need of a good data mining system that would overcome geographical and electronic barriers. We plan on developing several design solutions for the design cases identified and reported in this paper.conclusion

Human computer interaction design methodology provides meaningful means to make data mining easier for users [11]. It needs to be seen what factors will be key to the acceptance of new means of working with the clinical data. Clinical experts will need to continuously evaluate the results of knowledge extraction and to provide suggestions for improvements of the whole process [12].

REFERENCES

1. Eaton, Chris, Deroos, Dirk, Deutsch, Tom, Lapis, George, & Zikopoulos, Paul. (2012). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data* Retrieved from http://public.dhe.ibm.com/common/ssi/ecm/en/iml14296usen/IML14296USEN.PDF
2. Bifet, Albert, Holmes, Geoff, Pfahringer, Bernhard, Kranen, Philipp, Kremer, Hardy, Jansen, Timm, & Seidl, Thomas. (2010). MOA: Massive Online Analysis, a Framework for Stream Classification and Clustering (pp. (3-16)): University of Waikato, Hamilton, New Zealand, Aachen University, Germany.
3. Hall, Mark, Frank, Eibe, Holmes, Geoffrey, Pfahringer, Bernhard, Reutemann, Peter, & Witten, Ian H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations, 11*(1), (10-18).
4. Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med, 34*(2), 113-127. doi: 10.1016/j.artmed.2004.07.002
5. Wilson, Andrew M., Thabane, Lehana, & Holbrook, Anne. (2003). Application of data mining techniques in pharmacovigilance. *British Journal of Clinical Pharmacology, 57*(2), (127-134). doi: 10.1046/j.1365-2125.2003.01968
6. Shaw, Michael J., Subramaniam, Chandrasekar, Tan, Gek Woo, & Welge, Michael E. (2001). Knowledge management and data mining for marketing. *Decision support systems, 31*(1), (127-137).
7. Hevner, Alan R., March, Salvatore T., Park, Jinsoo, & Ram, Sudha. (2004). Design Science in Information Systems Research. *MIS Quarterly, 28*(1), 75-105
8. Rogers, Yvonne, Sharp, Helen, & Preece, Jenny. (2011). *Interaction design: beyond human-computer interaction* (3rd ed.). UK: John Wiley & Sons Ltd.
9. Å berg, Anders, Ericsson, Jan, Holmberg, Lars, Rozell, Barbro Lundh, Ayoubi, Shiva, Khan, Staffan, & Klint, Å sa. (2011). Cancer Incidence in Sweden 2010 *Statistics – Health and Medical Care* (pp. 1-100).
10. NSD, Norsk samfunnsvitenskapelig datatjeneste. (2012). Personvernombudet for forskning . Retrieved 30 March, 2013, from http://www.nsd.uib.no/personvern/forskningstemaer/registerstudier.html
11. MacLean, Allan, Young, Richard M., & Moran, Thomas P. (1989). Design Rationale: The argument behind the artifact. *SIGCHI Bull., 20*, 247-252. doi: 10.1145/67450.67497
12. Nielsen, Jacob, & Molich, Rolf. (1990). Heuristic evaluation of user interfaces. *CHI '90*, 249-256. doi: 10.1145/97243.9728

Agile Development for Smart User Interfaces to Cancer Registries

G. Kanza, MSc[1], and A. Babic, PhD[1,2]

[1] University of Bergen/Department of Information Science and Media Studies, Bergen, Norway
[2]Linkoping University/Department of Biomedical Engineering, Linkoping, Sweden

Abstract— **The paper studies different user interfaces developed for visualizing data mined from cancer registries. The motivation behind this research is a need to create more flexible and smart, easy to use interfaces that will assist users in utilizing and exploring clinical information. The design process combined Parallel and Iterative design process models together with Personal Kanban for managing the development process. The developed prototype provides users with a choice of several data visualization possibilities, depending on the user's tasks and goals. A preliminary user interface was evaluated resulting in recommendations for further development. Heuristic evaluation with potential experts had scores on all the high scores on all the design dimensions: visibility of system status; match between the system and the real world; user control and freedom; consistency and standards; error prevention; recognition rather than recall; flexibility and efficiency of use; aesthetic and minimalist design; help users recognize, diagnose, and recover from errors; and help and documentation**

Keywords— **Cancer registries, User interface, Personal Kanban, Parallel and Iterative design, Data visualization.**

I. INTRODUCTION

Clinical registries have been a good source of information and are used for research purposes in areas such as epidemiological research and health care planning and monitoring. They are used as a guide for evaluating and planning cancer control programs across the different regions. This research looks at reports based on the data mined from the Swedish Cancer Registry as a starting point in suggesting how this data can be presented to the audience in smart and user-friendly interfaces that enhance the utilization of information kept in such data resources.

These particular registries are a collection of data about the incidences of cancer in the different regions of the country [1]. They contain among others, individual cancer patients' information; information about the types of cancer and the tumor's location in the body; their development phase at diagnosis; and where and when it was diagnosed. This information is obtained from other clinical systems such as the Electronic Patient Records (EPRs).

Human Computer Interaction (HCI) methodologies in combination with agile development approaches are applied in providing a solution that can improve the current situation of presenting this kind of data. Due to the information needs in several settings, three different user groups were identified: clinicians, medical students and the public; and two design case studies are presented as a result of the design process.

II. USER INTERFACE DESIGN AND USER PROFILES

The developed prototype utilizes the Information Technology (IT) that is close to the users' experience and understanding. This was expected to improve the data mining from the cancer registry, which is typically a demanding task. The artifact resulting from the design process presents several user interfaces that allow different user groups to choose the way in which they want to view, explore and use the data. This approach is a remedy for the current system in which users are presented with the data in a bulk format, which they manage on their own, best they can, often with the help of a statistician.

Designing user-friendly interfaces for data visualizing makes it easy to navigate through the data and enables retrieval of any information that would be hard to manage due to the presentation and amount of information. Smart interfaces make data more accessible since users can specify exactly what data they want to see using just a few clicks. Several cases will therefore be demonstrated and will show that it is easier to explore data through simple natural interactions. User will have easier time to go through a complex annual report that summarizes all kinds of relevant data typically in what seems to be endlessly many tables and graphs all detailing information on the cancer types, treatments and outcomes.

Three different user groups were identified; clinicians, medical students and the public.

Clinicians

This is the user group with the widest spectra since they continuously use data mined from the cancer registries for research purposes and improving clinical practices. They do not have direct interaction with the registries, but use statisticians to obtain the data. For instance, for reporting certain cancer incidences at regional level, they would obtain the data after going through a series of steps and only then, can they analyze this data and make the report of particular interests to them. The data is often delivered with some delay and comes in many rows and columns, which makes it difficult to easily extract information. Therefore, by having smart and user-friendly interfaces changes could be made to improve interactions with the registry system. Direct interactions between the clinicians and the data will reduce the steps they are now performing. This will eventually prevent loosing the site over the information and it will provide users with alternative interfaces to view the data.

A good example is when researchers explore the mined data in order to detect trends or make hypothesis, it is important for them to have an overview of all the variables, which they think could be of impact.

Medical students

For students, the resulting prototype opens up learning possibilities as it can be used as an e-learning tool by defining tasks in which the mined data could be used, for instance, connect types of cancers, their treatments and the study outcomes to do research even at the student level. User interface with a good visualization technique will allow a student to explore relationships in the data to combine different representation forms and to understand relationships which otherwise could be hard to find.

The Public

The current system allows the public to view the data from the cancer registries in form of the annual reports. The public gets a limited access due to the nature and sensitivity of the data stored in these clinical data resources. Even though a lot of this data is publicly available, it is still not easy to extract information due to the way the data is presented. The many rows and columns of the data are tedious and not at all encouraging to go through in search for specific information. Therefore, the need to ease the ways of presenting data is of crucial importance. The designed prototype provides possibilities to look at the published reports resulting in several easy to follow interfaces.

The inclusion of other related data from several data resources will assist users in obtaining more information about the particular type of cancer selected. For example, the prototype provides the possibility for users to search for ongoing clinical trials that are related to the selected type of cancer where they can see the purposes of the trials, outcomes, enrollment criteria and the distance to the nearest center performing clinical trials.

III. Methods

Innovative Design

In order to design an innovative solution to improve data visualization from the cancer registries, this research has opted for Genius design approach [2] in the initial iteration, which did not include users. Rogers et.al [3] suggest that even though users often can not tell designers what they want, they will know what they do not want after seeing and trying out a product. Therefore, presenting the users with the concept that will eventually solve the current problem opened up for the users views on what they really want. Consequently, we took this view to develop concepts, present them for users and made them choose. In addition, research based personas and scenarios [3] were constructed in order to establish the initial system requirements.

Innovative design was chosen specifically for the initial design in our research since we expected it to be easier for users to consider tentative solutions and to make them think about their own wishes. The speed of technology changes is another reason for choosing innovative design since it could quickly and efficiently presents solutions that would be otherwise difficult to envision.

Prototyping

This research project has portrayed the design ideas for visual presentation of the data mining from the cancer registries through the development of several prototypes in three iterations.

Due to schedule and time constraints, distributed resources, and the complex nature of the cancer registries and medical data, the final prototype will not be implemented fully in this paper. Nevertheless, emphasis was put on suggesting features that are technically feasible for the development team to implement and that can be utilized even beyond the current project.

A combination of Parallel [4] and Iterative [3] design models was applied in order to achieve high quality user interfaces. Given the limited time and resources, this combination was cost effective, fast and efficient to use for prototyping.

Pre-iteration tasks comprised of the following: gaining an overview of the current situations and the ways the data is retrieved, handled and presented from the cancer registries; and there after, specifying goals and requirements that led to improved ways of visualizing data from these rich data resources.

Three different design alternatives were created in the first iteration in the form of sketches, both hand-drawn and electronic. More than one set of sketches was made in order to present the design solution from different angles. Emphasis was put on the top features of the designed alternatives. The sketches at this stage only presented the positioning and labeling of the intended buttons and functions, but none of them worked. These kept on evolving into rich electronic interactive vertical prototypes [3] by the end of the last iteration.

During the second iteration, some interactivity was added to the initial sketches and usability testing was carried out using all three design alternatives. Users were able to interact with the artifact and provide feedback, which was utilized in the iteration that follow to adjust the design features of the final prototype. The Question, Option and Criteria (QOC) [9] method used to single out one designed interface to be further developed.

Alternations made to the prototype in this iteration were results from the user feedback on the previous iteration.

Personal Kanban

Personal Kanban [5] was used to manage the development of the design process. This agile approach emphasizes on work visualization and limiting work-in-progress (WIP) and it is suitable for projects carried out by individuals and small teams.

It was important to use this Agile development [6] approach in order to systematically keep track of all the design tasks; focus, visualize the amount of tasks involved when prototyping; and to prioritize the tasks execution.

The Personal Kanban board used was divided into three columns; *backlog*, *in progress* and *done*.

The backlog column consisted of all the tasks to be completed when prototyping. They were moved from left to right across the board following their order of priority.

A WIP limit of five tasks was set to the '*in progress'* column in order to control the number of tasks to be handled at the same time and to minimize task switching. In this way, the tasks were executed systematically, which in turn decreased the number of unfinished tasks and helped keeping control of the workflow.

For instance, a task was first to be moved to the 'in progress' column when it was ready to be processed, but only if there was room for it. Upon its completion, it was moved to the 'done' column.

The Pomodoro Technique [5] was useful not only for tracking time used on each task, but also was a reminder of taking in-between breaks in order to enhance productivity while working on the tasks. A series of short 5-minute breaks were taken after devoting full concentration on a task for 25 minutes. In this context, these short breaks are known as the Pomodoros. A 20-minute break followed after working 5 pomodoros. This approach was an encouragement to breaking down the work into simple and focused tasks.

IV. Design case studies

Design Case Study 1:

Problem: Finding the number of males and females who were diagnosed with breast cancer in 2010 in the county of Skåne.

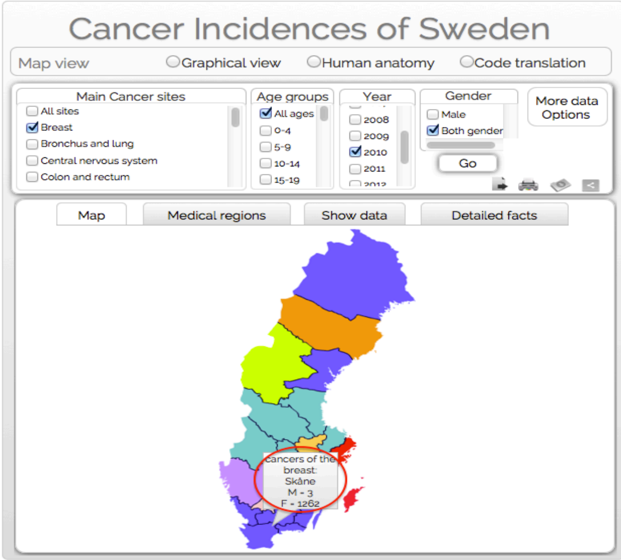


Fig 2: Design Case Study 1, Viewing data from a map

Solution1: The current solution for doing such a straightforward task involves the following steps; looking through several report pages in search of the categorized statistics then when the appropriate page is found; searching through dense lists with many cancer types in order to find the intended type; and at last browsing through the rows to find the right region. The search result is the number found at the intersection between breast cancer and the column representing the county of Skåne. Handling data this way is a tedious and demanding task for any user, which discourages data exploration.

Solution2: The interface in this prototype simplifies the task by allowing users to select variables and click on the '*GO'* button so that they can view the numbers.

Figure 2 shows the practical solution to this problem allowing the user to hover the mouse over Skåne after specifying the variables *breast cancer*; *all ages; year 2010* and *both genders* on the top section above the map.

Design Case Study 2:

Problem: Finding the number of occurrences of cancers of the bone for both genders living in Blekinge, and at the same time, see how the same type of cancer has been diagnosed in patients aged between 2 – 84 years. In addition, find how many autopsied cases in the course of death register of 2009 had cancers of the Stomach; Colon and rectum; Pancreas; Bronchus and lung; and Prostate.

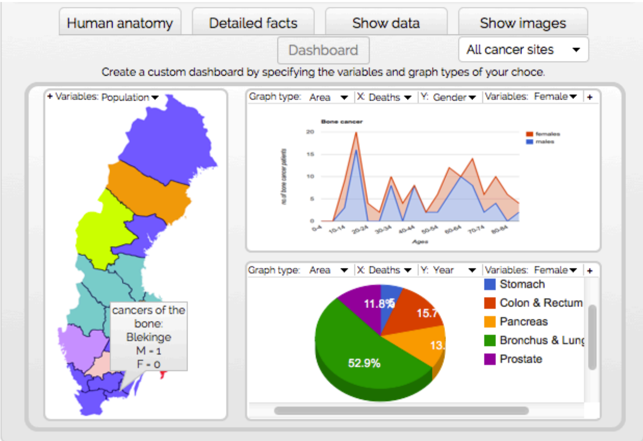


Fig. 3: Design Case Study 1, A Dashboard solution

Solution 1: For such a complicated task, the current solution executes each of these sub-tasks separately following the same steps as in design case study one, which ends up in using too much time going through the bulk data, something that is not very pleasant for the users. Extracting information from performing this task may lead to information loss, as the user is not provided with possibilities to look at this data from different angles.

Solution 2: Even though this task may seem complicated, the designed prototype provides users with an easy to use dashboard function with possibilities to view all the data from all the sub-tasks in one screen. *Figure 3* shows a solution to the stated problem on a dashboard with interactive features allowing users to directly interact with the system.

V. Evaluation and Conclusions

Nielsen's ten heuristics [4] were used as a guideline in order to evaluate the elements of the user interface during the evaluation process. Master students with a background in HCI and who have experience in practicing usability methods were used as experts to evaluate the designed artifact. The prototype scored highly on all ten dimensions: H1=Visibility of system status scored 9.8 out of 10; H2=Match between the system and the real world scored 10; H3=User control and freedom scored 9.8; H4=Consistency and standards scored 10; H5= Error prevention scored 9.8; H6=Recognition rather than recall scored 10: H7=Flexibility and efficiency of use scored 9.8; H8=Aesthetic and minimalist design scored 10; H9=Help users recognize, diagnose, and recover from errors scored 9.6; and H10= Help and documentation scored 9.4.

The results suggest that users were comfortable with the interaction and were able to extract meaningful information. In e-learning settings the interface could provide in-depth understanding of clinical data by enabling users to explore different dimensions and types of data. The designed prototype enables possibilities for the broader audience to interact with the system, visualize and acquire data for their personal needs such as type of cancer, complications, outcomes, clinical trials, and all in an efficient and user-friendly manner.

There is a huge increase of software applications in clinical areas [7][8], which provides opportunities to largely improve the ways we use clinical data. We have shown the contribution of smart, simple interfaces to explore data and utilize the information that is otherwise hidden in the reports and databases.

References

1. Ericsson, Jan, Åberg, Anders, Holmberg, Lars, Rozell, Barbro Lundh, Ayoubi, Shiva, Khan, Staffan, & Klint, Åsa. (2011). Cancer Incidence in Sweden 2010 *Statistics – Health and Medical Care* (pp. 1-100).
2. Saffer, Dan. (2010). Designing for interaction: creating innovative applications and devices: New Riders.Smith J, Jones M Jr, Houghton L et al (1999) Future of health insurance. N Engl J Med 965:325–329 DOI 10.10007/s002149800025
3. Rogers, Yvonne, Sharp, Helen, & Preece, Jenny. (2011). Interaction design: beyond human-computer interaction (3rd ed.). UK: John Wiley & Sons Ltd.
4. Nielsen, Jacob, & Molich, Rolf. (1990). Heuristic evaluation of user interfaces. CHI '90, 249-256. doi: 10.1145/97243.97281
5. Benson, Jim, & Barry, Tonianne Demaria. (2011). Personal Kanban: Mapping Work, Navigating Life: Modus Cooperandi Press.
6. Dingsøyr, Torgeir, Dybå, Tore, & Moe, Nils Brede. (2010). Agile software development: an introduction and overview Agile Software Development (pp. 1-13): Springer.
7. Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med, 34(2), 113-127. doi: 10.1016/j.artmed.2004.07.002
8. Wilson, Andrew M., Thabane, Lehana, & Holbrook, Anne. (2003). Application of data mining techniques in pharmacovigilance. British Journal of Clinical Pharmacology, 57(2), (127-134). doi: 10.1046/j.1365-2125.2003.01968.x
9. MacLean, Allan, Young, Richard M., & Moran, Thomas P. (1989). Design Rationale: The argument behind the artifact. *SIGCHI Bull., 20*, 247-252. doi: 10.1145/67450.67497