FP7-ICT-611140 CARRE

Project co-funded by the European Commission
under the Information and Communication Technologies
(ICT) 7th Framework Programme

D.5.3. Advanced Visual Analytics Module

Y. Zhao, E. Liu, F. Parvinzamir

April 2016

CARRE Contacts

Project Coordinator: Eleni Kaldoudi kaldoudi@med.duth.gr

DUTH Democritus University of Thrace	Eleni Kaldoudi	kaldoudi@med.duth.gr
OU The Open University	John Domingue	john.domingue@open.ac.uk
BED: Bedfordshire University	Enjie Liu	Enjie.Liu@beds.ac.uk
VULSK: Vilnius University Hospital Santariškių Klinikos	Domantas Stundys	Domantas.Stundys@santa.lt
KTU Kaunas University of Technology	Arūnas Lukoševičius	arunas.lukosevicius@ktu.lt
PIAP Industrial Research Institute for Automation & Measurements	Roman Szewczyk	rszewczyk@piap.pl

Disclaimer

CARRE is a Specific Targeted Research Project partially funded by the European Union, under FP7-ICT-2013-10, Theme 5.1. “Personalized health, active ageing & independent living”.

Document Control Page

Project	
Contract No.:	611140
Acronym:	CARRE
Title:	Personalized Patient Empowerment and Shared Decision Support for Cardiorenal Disease and Comorbidities
Type:	STREP
Start:	1 November 2013
End:	31 October 2016
Programme:	FP7-ICT-2013.5.1
Website:	http://www.carre-project.eu/
Deliverable	
Deliverable No.:	D.5.3
Deliverable Title:	Advanced Visual Analytics Module
Responsible Partner:	BED
Authors:	Youbing Zhao, Enjie Liu, Farzad Parvinzamir
Input from:	All partners
Peer Reviewers:	E. Kaldoudi (DUTH), D. Stundys (VULSK)
Task:	T.5.3. Visual Analytics
Task duration:	10 months: 1 June 2015 to 30 April 2016
Work Package:	WP5: Data management & visual analytics for empowerment
Work Package Leader:	BED – Enjie Liu
Due Date:	30 April 2016
Actual Delivery Date:	18 May 2016
Dissemination Level:	PU
Nature:	R & D
Files and format:	Deliverable report: 1 pdf file Software source code available from project web site, see Annex 2: https://www.carre-project.eu/innovation/interactive-visual-interface/ and http://visual.carre-project.eu
Version:	03
Status:	☐ Draft ☒ Consortium reviewed ☒ WP leader accepted ☒ Coordinator accepted ☐ EC accepted

Document Revision History

Version	Date	Modifications	Contributors
v01.1	11.04.2016	Initial version	Youbing Zhao
v01.2	20.04.2016	First revision	Youbing Zhao
v01.3	05.05.2016	Duplicated parts removed	Enjie Liu
v01.4	09.05.2016	Design study added	Enjie Liu
v01.5	14.05.2016	Use case sequence diagram added	Youbing Zhao
v01.6	15.05 2016	Various minor changes	Enjie Liu
v02.0	16.05.2016	Reviewers comments and editing for uniformity	Eleni Kaldoudi
v02.1	17.05.2016	Reviewers comments	Domantas Stundys
v03	18.05.2016	Editing for reviewers comments & uniformity	Eleni Kaldoudi

Table of Contents

List of Figures

Executive Summary

CARRE personalised patient empowerment and decision support services require presentation and analysis of a large volume of heterogeneous data and metadata as harvested from a variety of data sources including sensors, risk factors, PHR, decision support, etc. Without proper tools it is impossible to achieve this goal.

Work Package 5 “Data Management & Visual Analytics for Empowerment” is proposed to meet this challenge. In Task 5.1 “Interactive Visual Interface”, an interactive visual analysis interface was designed to empower both the patients and the medical experts to view, utilise, analyse and understand the data.

This document is a deliverable report of D.5.3 “Advanced Visual Analytics Module” of WP5 in CARRE project. Some parts are based on previously submitted deliverable D.5.1 “Interactive Visual Interface”. In particular it covers the advanced visual analytics tasks designated in Task 5.1. This deliverable report focuses on the design and implementation of visual analysis components in order to provide personalised views and analysis of sensor data, PHR data, risk factors and decision support data. The report is organised to discuss visualization of data from different data sources, visual analytics interface and components, use cases and implementation details.

About CARRE

CARRE is an EU FP7-ICT funded project with the goal to provide innovative means for the management of comorbidities (multiple co-occurring medical conditions), especially in the case of chronic cardiac and renal disease patients or persons with increased risk of such conditions.

Sources of medical and other knowledge will be semantically linked with sensor outputs to provide clinical information personalised to the individual patient, to be able to track the progression and interactions of comorbid conditions. Visual analytics will be employed so that patients and clinicians will be able to visualise, understand and interact with this linked knowledge and take advantage of personalised empowerment services supported by a dedicated decision support system.

The ultimate goal is to provide the means for patients with comorbidities to take an active role in care processes, including self-care and shared decision-making, and to support medical professionals in understanding and treating comorbidities via an integrative approach.

Terms and Definitions

The following are definitions of terms, abbreviations and acronyms used in this document.

Term	Definition
API	Application programming interface (API) is a set of functions and procedures that allow the creation of applications that access the features or data of an operating system, application, or other service
DSS	Decision Support System
EC	European Commission
EHC	European Health Card
eHealth	Electronic Health
EU	European Union
HTML	Hypertext Markup Language, the standard markup language used to create web pages.
HTTP	Hypertext Transfer Protocol, a data communication for the World Wide Web.
HTTPS	Hypertext Transfer Protocol Secure
OWL	Web Ontology Language
PHR	Personal Health Record
REST	Representational State Transfer, is a software architecture style for building scalable web services.
RDF	Resource Description Framework - a standard model for data interchange on the Web.
SPARQL	RDF query language, that is, a query language for databases, able to retrieve and manipulate data stored in RDF
SVG	Scalar Vector Graphics, is an XML-based vector image format for two-dimensional graphics with support for interactivity and animation
UC	Use case
UMLS	Unified Medical Language System
URL	Uniform Resource Locator
VA	Visual analytics
Vis	Visualisation
WWW	World Wide Web

1. Introduction

CARRE personalised patient empowerment and decision support services are based on data from a variety of data sources including sensors, Personal Health Record (PHR) systems, manually health data entry system and public sources such as the CARRE risk factor database. Without proper tools it is almost impossible to present and analyse these large, heterogeneous, time-varying data.

Work Package 5 “Data Management & Visual Analytics for Empowerment” is proposed to address this challenge. In Task 5.1 “Interactive Visual Interface”, an interactive visual analysis approach is suggested be designed to empower both the patients and the medical experts to view, utilise, analyse and understand the data. Visual analytics is an integral approach which combines visualisation, human factors, and data analysis. This process incorporates automatic and visual analysis methods with a tight coupling through human interaction in order to view, analyse and understand the data.

As a science of analytical reasoning facilitated by interactive visual interfaces, visual analytics is critical in the applications development. Visual analytics interfaces allow the analysts to interact directly with the representation of the data or to modify visualisation parameters. In the visual analytics process, knowledge can be gained from visualisation, automatic analysis, as well as interactions between visualisations, models, and the human analysts.

In WP5 the visual analytics techniques deliver different perspectives on specific data types to provide role-specific (personalized) and goal-oriented representations of the data. The interactive visual interface directly supports complex decision making tasks. Standard interaction techniques will be implemented to facilitate data exploration and knowledge discovery.

The report of D5.1 interactive visual interface was the first deliverable report of task 5.1 “Interactive Visual Interface” with the aim to provide a design of visual analytics interfaces and components for CARRE patients and medical experts to facilitate understanding and analysis of sensor data, risk factor data, PHR data and decision support data, etc.

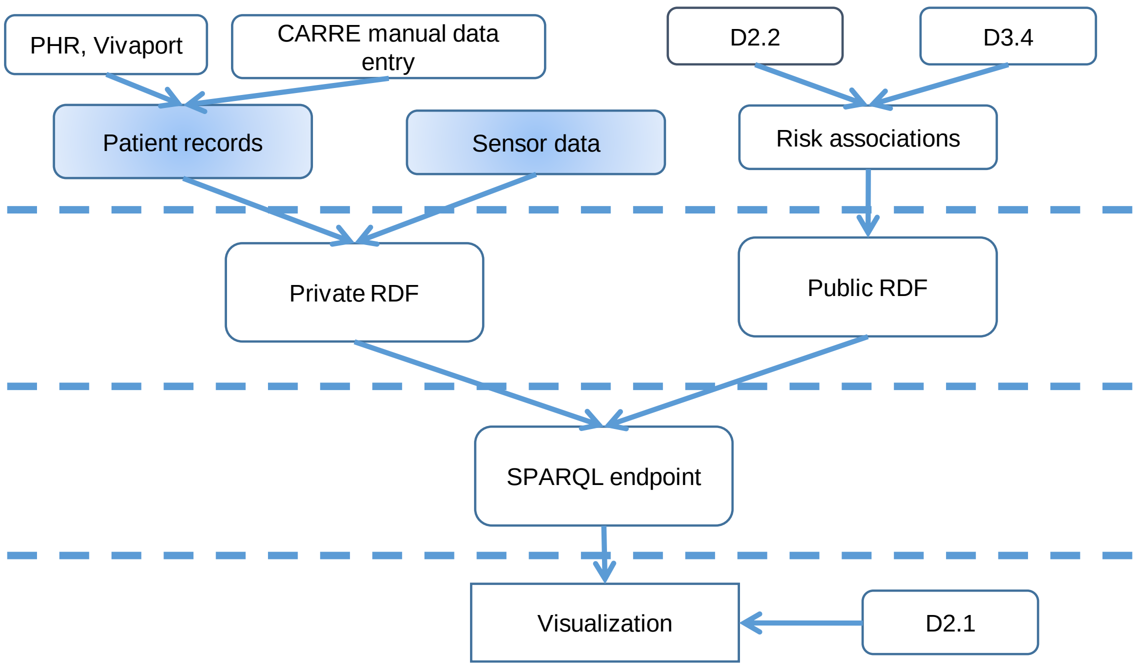


Figure 1. Relation of WP5 and other deliverables in CARRE

In this D.5.3 report, we describe development of the functions for the visual analytics interface, the use cases of the visual analytics, and also report on the improved graph visualization technologies. The deliverable also focuses on the following main challenges (see Section 1.2): (a) scalability to large graph data sets; (b) clarity, usability and aesthetical aspects; and (c) maximum portability using Web technologies. Finally, uncertainty is employed to support visual analytics.

The visual interface covers use cases that have been defined in D.2.1. The visual interface is intended as a common access point for all CARRE users, from which users can perform the typical actions as defined in the project. According to deliverable D.2.1, four high level visual analytics use cases have been identified: to allow patients to understand general disease progression; to show patients' disease progression based on personal monitored data; to show estimated disease progression if they change their lifestyle; to allow them to understand their disease by comparing their personal state with current medical evidence.

We define the visual analytics tasks provided by the CARRE visual interface as the follows:

- Visualise data from self-monitoring of own health-status including sensors, daily activities, symptoms and PHR data.
- Visualise individual risks and allow for methodical analysis of the impact of behaviour changes to the estimated personal risks.

Error! Reference source not found. shows the data collection and management in CARRE, and the data retrieving of WP5 from other work packages, and the functional interlink with other work packages.

2. CARRE visual analytics system

CARRE system is composed of different web-based functional modules that communicate with each other (see CARRE architecture, D.2.5). The visual analytics module is a web-based sub-system that provides supporting functions for data exchange, visualisation and analysis. The system is available online at http://visual.carre-project.eu. In this section, the visual analytics web interface and the advanced visual analytics components are introduced.

2.1. CARRE visual analytics system interface

The visual analytics system includes web hosting, access control and a web-based framework for hosting visual analysis components. The visual analytics system interface serves as a user interface for accessing CARRE services for patients. Figure 2 shows the login page of CARRE visual analytics web system. The user authentication will use private RDF repository authentication mechanism to achieve single-point access control. The web-based framework is composed of a dashboard and multiple visual analytics component containers.

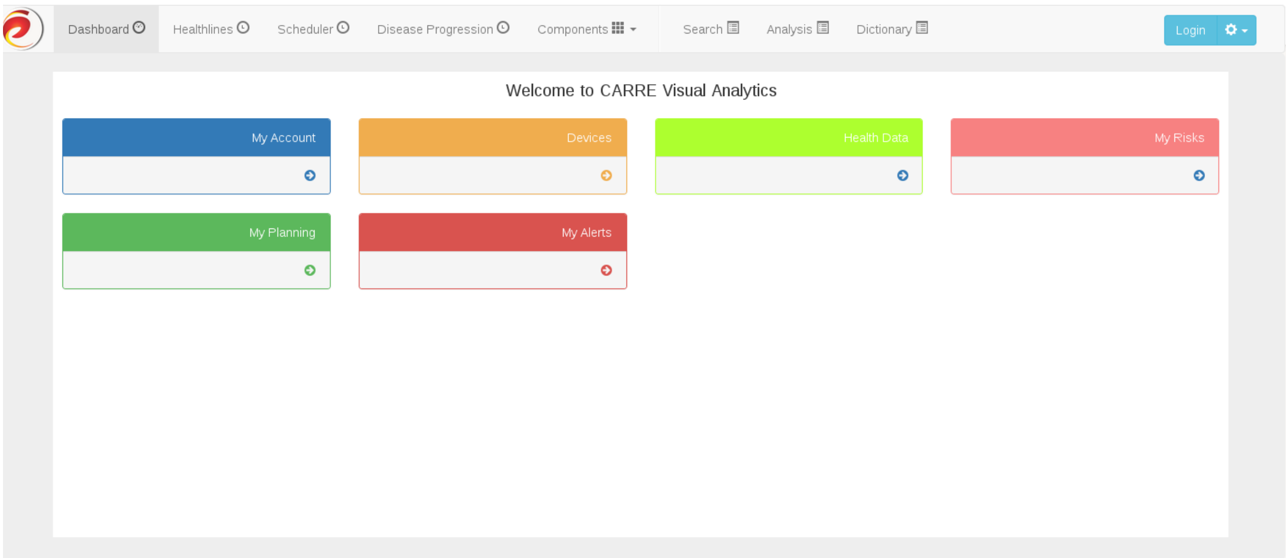


Figure 2 CARRE Visual Analytics Main User Interface

2.1.1. User Login

CARRE provides unified user access control. The user login module of the visual analytics system is based on that. When the user clicks on the “Login” button as shown in Figure 3**Error! Reference source not found.**, the web page will automatically redirects to the CARRE shared login site, as shown in Figure 4, which is designed and implemented as part of the CARRE middleware services (RESTful API and access control, see CARRE architecture, D.2.5). After the user inputs the correct user name and password, the user name will be displayed on the page, as shown in Figure 5.

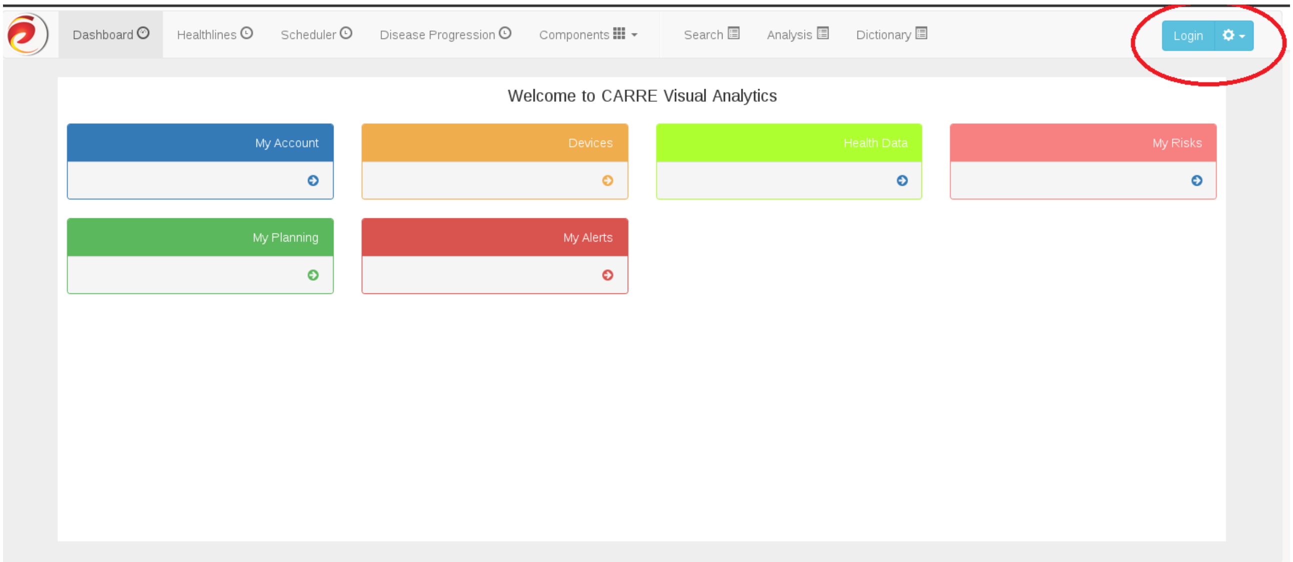


Figure 3. Login button on the landing page.

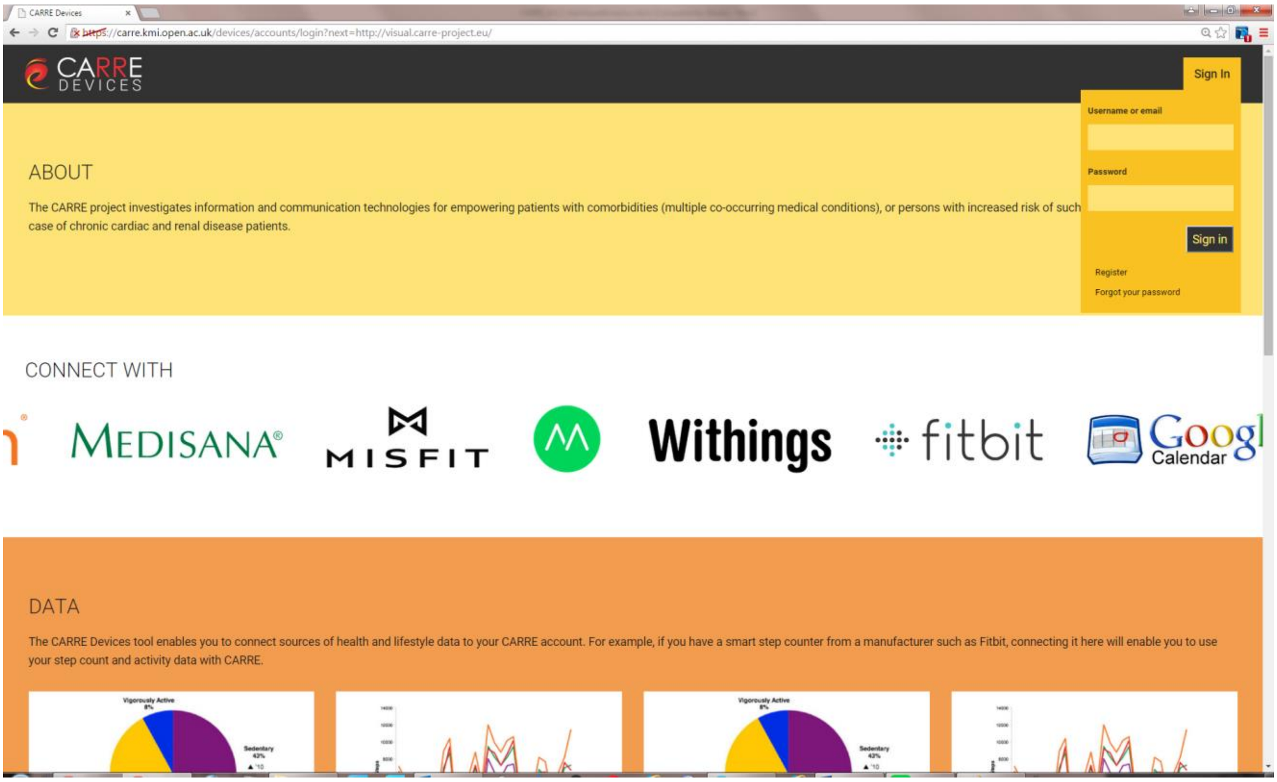


Figure 4. CARRE central user registration and sign-in control

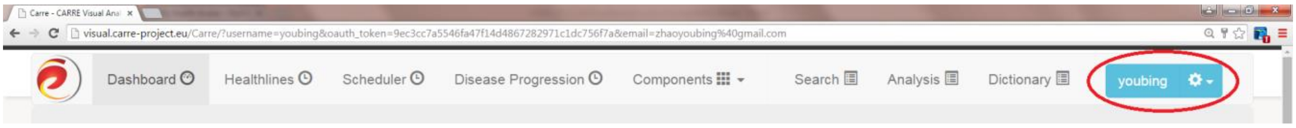


Figure 5. The login button shows the user name after the user has successfully signed in.

2.1.2. Data retrieval

The visual analytics data are retrieved from both private and public CARRE RDF repositories. The private RDF stores personal user data, such as the measurement data collected by sensors; while the public RDF stores disease risk associations as harvested from scientific literature.

Visual analytics provides an interactive interface for users to visualize the data stored via the visual analytics components that give different visual approaches to represent available data. The data retrieval is carried out via APIs (application programming interfaces) developed in the project (D.4.1.). For example, if we need to access the sensor measurement data for each patient, we can use the measurement API:

https://carre.kmi.open.ac.uk:443/ws/measurement?name=NAME&token=TOKEN

where: NAME is the RDF variable name and token is the user token which is acquired from the CARRE user management service.

The following URL is an example of retrieving the steps measurement name:

http://carre.kmi.open.ac.uk/ontology/sensors.owl#has_steps

The list of the measurements can be acquired from the following API:

https://carre.kmi.open.ac.uk:443/ws/measurementsList?token=USER_TOKEN

2.1.3. Dashboard

There are multiple visual analytics components that can be accessed by the user from the web-based CARRE visual analytics interface. However, as there is a variety of data sources and data types, it is difficult for a user to grasp an overview with important features from the scattered health status visualisation. To present the user a quick overview of their health status, CARRE visual analytics web interface provides a dashboard as a front page.

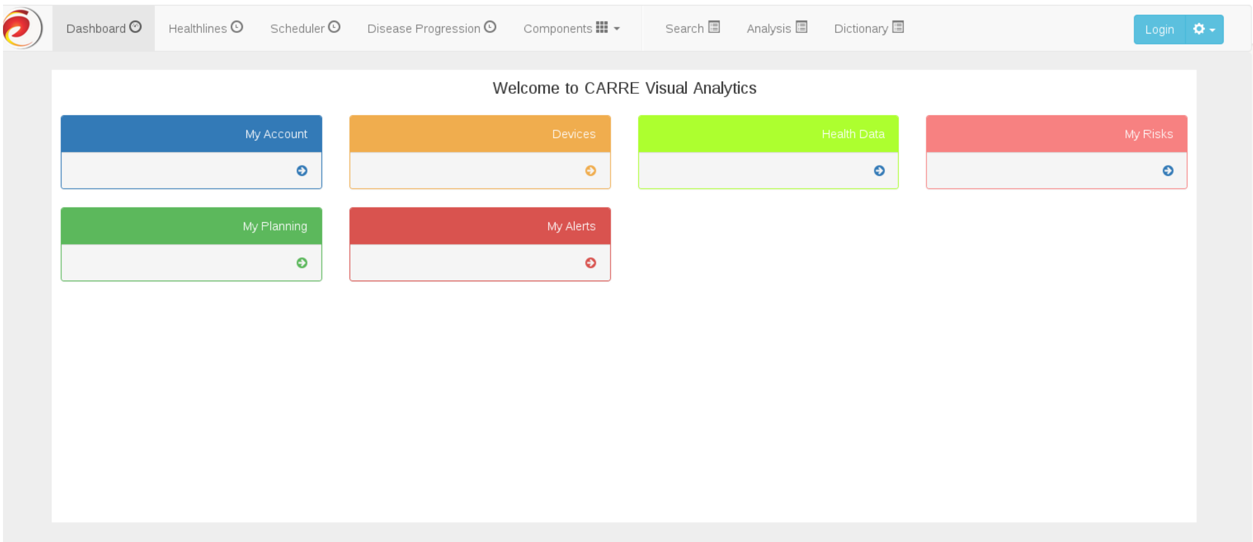


Figure 6. CARRE Visual analytics interface dashboard

The dashboard provides a summary of the user's latest health status and may present important notifications. It may include several visualisation components to present data in a relatively recent period. Figure 6 shows the example dashboard with data tiles, map and a timeline. The user can interact with the map and the timeline to obtain more detailed information.

2.1.4. Visual analytics component container

Visual analytics components are designed for visualisation and analysis of particular tasks. Rarely a single visual analysis component can meet all the requirements of a task. More commonly multiple visual analytics components need to be coordinated in synergy for a task. This implies that the related visual analytics components need to be organised in a container. The visual analytics component container is provided as a single page container for those visual analytics components. It supports automatic and manual layout setting of components. It further requires multiple visual analytics component containers which are organised as tabs in the CARRE visual interface. These tabs provide multiple visualisation containers in one place and are capable of performing coordinated tasks in terms of user interaction.

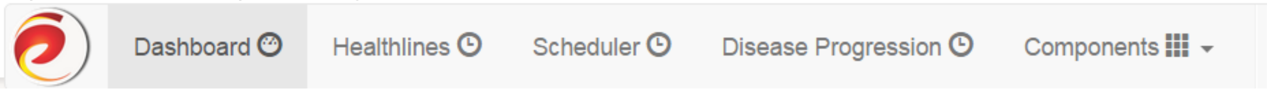


Figure 7 shows the current tab organisation in current CARRE visual interface implementation.

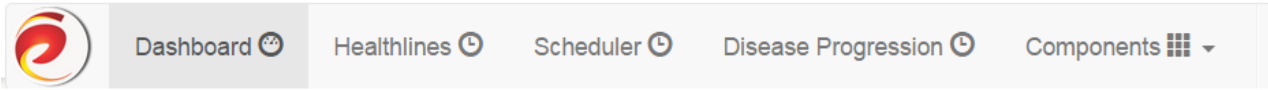


Figure 7. Tab-based component container organizer.

2.2. Sensor and PHR Data Visualisation

2.2.1. Healthlines

Time-dependant data are ubiquitous in many application domains as, for example, in business, medicine, history, planning, or project management. In CARRE, most of data, such as PHR and especially sensor data are time dependant. Providing appropriate methods to facilitate the visualisation and analysis of time-varying data is a key issue in CARRE. A timeline is a traditional method to visualise time-varying data and events in a linear layout and it is suitable for continuous variables which may cover a relatively long period, such as health indicators and medical measurements.

Figure 8. CARRE Healthlines

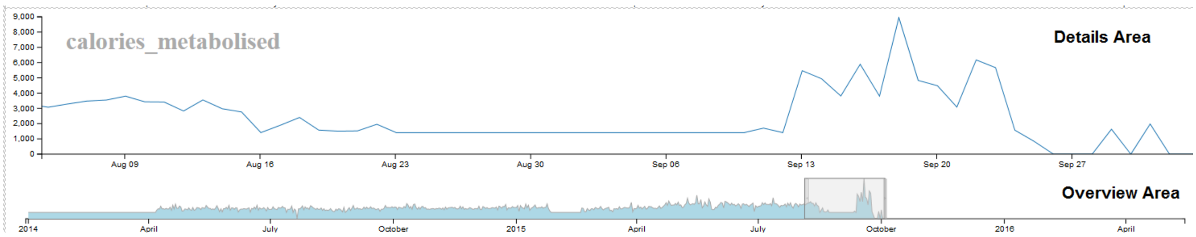


Figure 9. Overview area and detail area of each healthline.

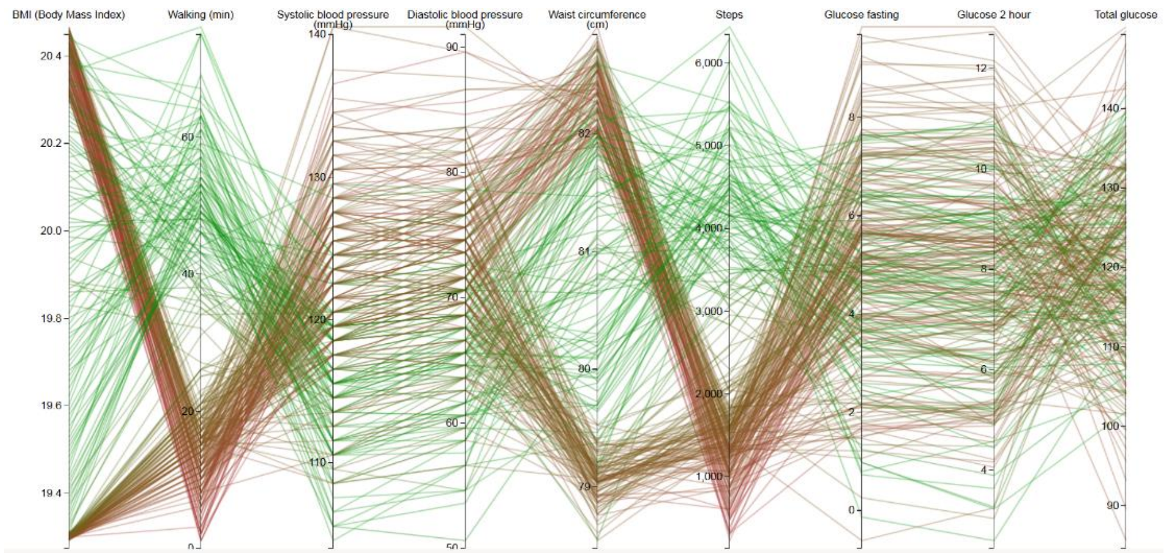
In CARRE the application of the timeline is mostly used for visualisation and analysis of the fitness statistics and the medical indicators. Data trends can be observed from the variable curves and data correlations may be discovered by comparison of the data curves of the multi-variables. As the data records may cover a long time range, interactive techniques such as zooming and “overview+details” are integrated with the visualisation. Bar charts and lines are available for variable visualisation. As there may be a number of variables to select, a convenient drag-and-drop approach is used for variable selection. Figure 8 shows multiple health indicators visualised in an interactive timeline in CARRE, while Figure 9 shows the additional functionality of selecting a window in the timeline to view it in a more detailed format.

The supported user interactions for the Healthlines are:

- Dragging variable names to the timeline to show the corresponding variable
- Overview+Details visualization: Select or change the time range of the data in the overview area to show the detailed data in the detail area
- Zooming and panning in the detail area

2.2.2. Parallel Coordinates

The technique of parallel coordinates is an approach for visualising multiple quantitative variables using multiple axis which are placed parallel to each other in the most common case[1]. The advantage of parallel coordinates is that it supports visualisation of multiple variables and correlation between attributes can be discovered by certain visualisation patterns. It is a common technique of visualising high-dimensional data and analysing multivariate data.



1 Inselberg, A., Dimsdale, B. Parallel coordinates: A tool for visualizing multi-dimensional geometry. Proc. 1st IEEE Symp. on Visualization, Oct. 1990, pp. 361-378

Figure 10. Health indicator correlation analysis based on Parallel Coordinates in CARRE

The parallel coordinates visualization is employed for multi-variable correlation analysis of the biomarkers. An example view of the parallel coordinate visualization is shown in Figure 10 where negative correlations can be found between walking minutes and blood pressures as well as BMI (Body Mass Index).

The supported user interactions for the Parallel Coordinates are:

- Axis reordering: the user can dragging an axis to a new position. In this way the user can reorder the axis to put closely related axis in a close neighborhood to better observe the data correlations.
- Brushing: the user can select a subrange on an axis, as shown in Figure 11. The data lines will be filtered based on the brush selection. The user can also move the brush selection on the axis vertically.

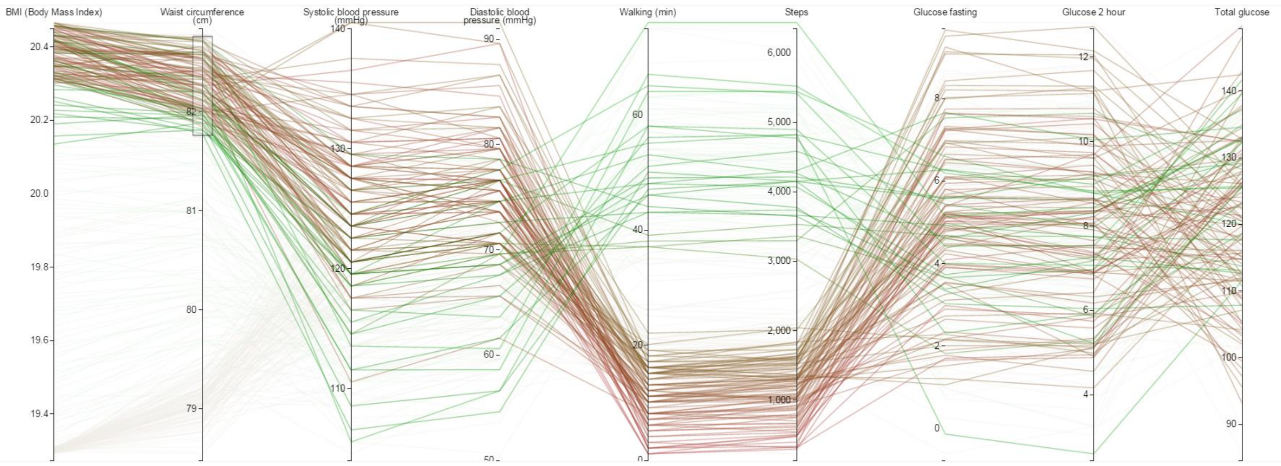


Figure 11. Brushing functionality

2.3. Risk Factor Data Visual Analytics Components

This section introduces the visual analytics components implemented.

2.3.1. Node-link Diagram

Node-link diagrams are usually used to visualise a tree or network graph data structure. In node-link visualisations of a network, entities are represented by nodes, the links or edges among those nodes represent relationships among entities. A node-link diagrams is the intuitive and natural way to represent relations between objects.

The basic graph layout is straightforward. Given a set of nodes with a set of relations (edges), it only needs to calculate the positions of the nodes and draw each edge as a curve. However, with as the number of nodes and edges increases, it becomes more and more difficult to make graphical layouts understandable and useful to ender users. Dynamic layout techniques can be used for node-link diagram to reduce difficulties in visualisation, such as force-directed layout[2] and Multi-Dimensional Scaling (MDS)[3].

Figure 12 is an example of risk factor visualisation in CARRE which uses additional channels such as colour, line width to visualise data attributes.

The supported user interactions are:

- Drag the nodes and links of the node-link diagram

2 Eades P. A heuristic for graph drawing. Congressus Numerantium, 42:149–160, 1984

3 Kruskal JB, Wish M, Multidimensional Scaling, Sage University Paper series on Quantitative Application in the Social Sciences, 07-011,Sage Publications, 1978

– Click on the nodes to show the risk observables of the corresponding risk element, as shown in Figure 13.
– Click on a link to show risk ratios associated with risk factor
– Filter the diagram based on patient-based diseases, as shown in Figure 14.

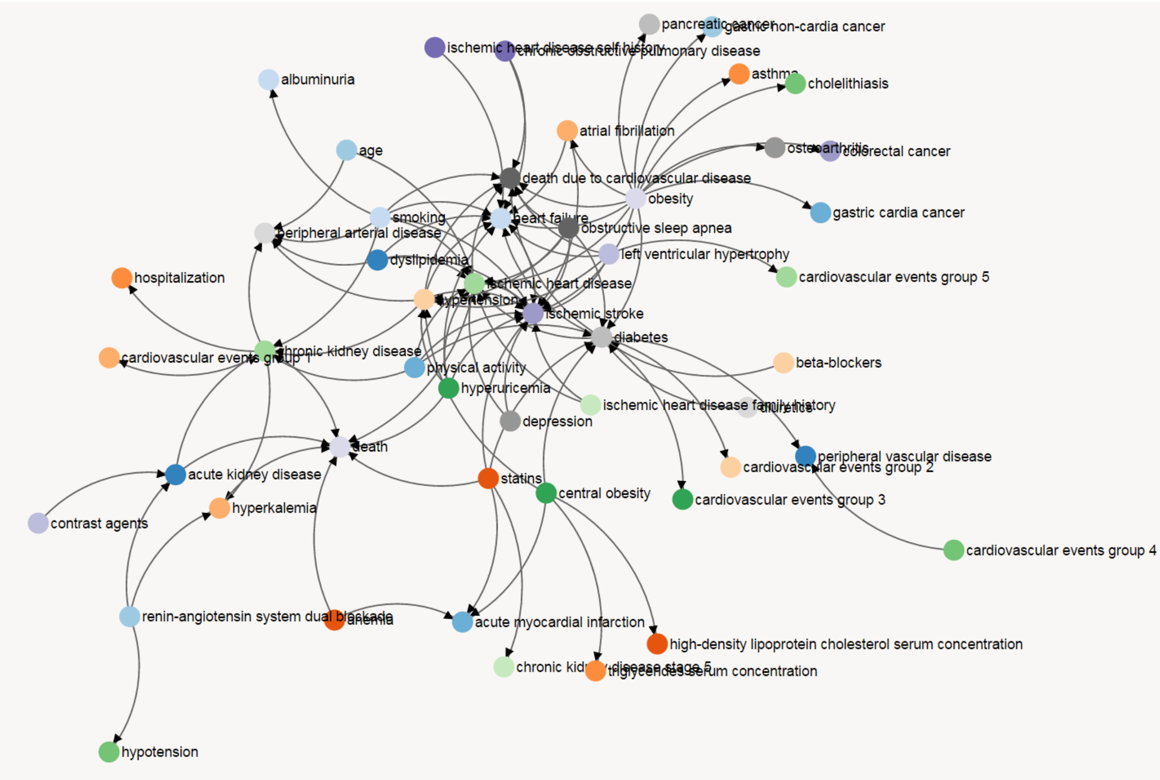


Figure 12. Node-link diagram example used in CARRE visual analytics to visualise risk factors

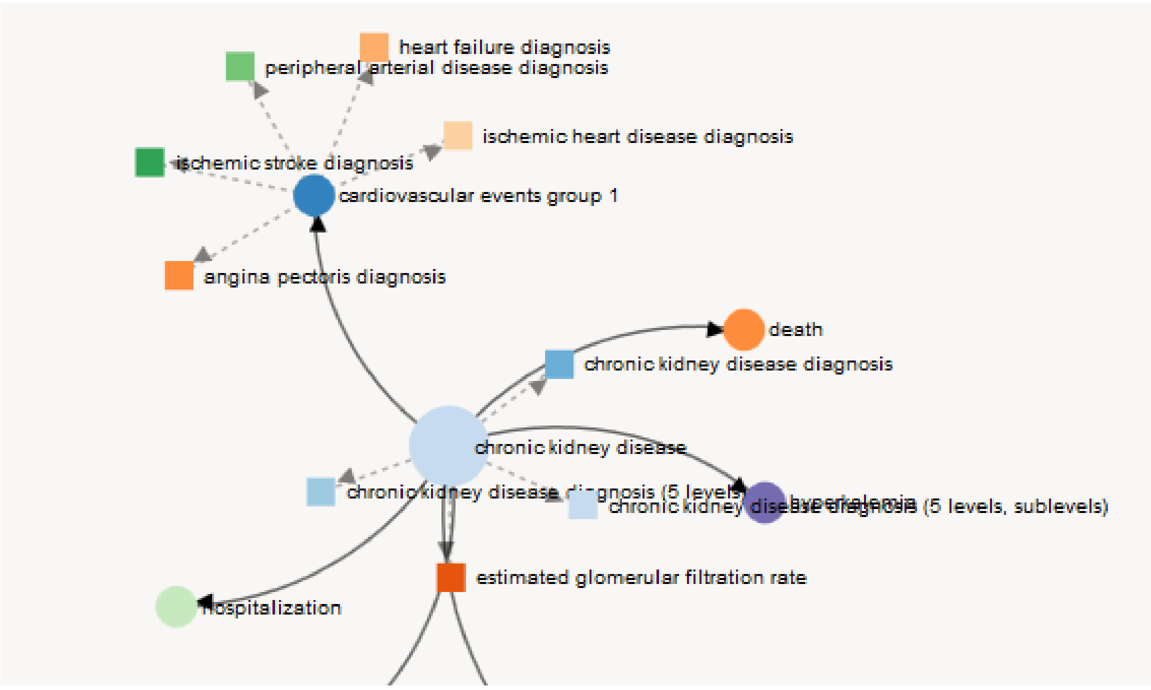


Figure 13. Clicking on risk elements shows risk observables.

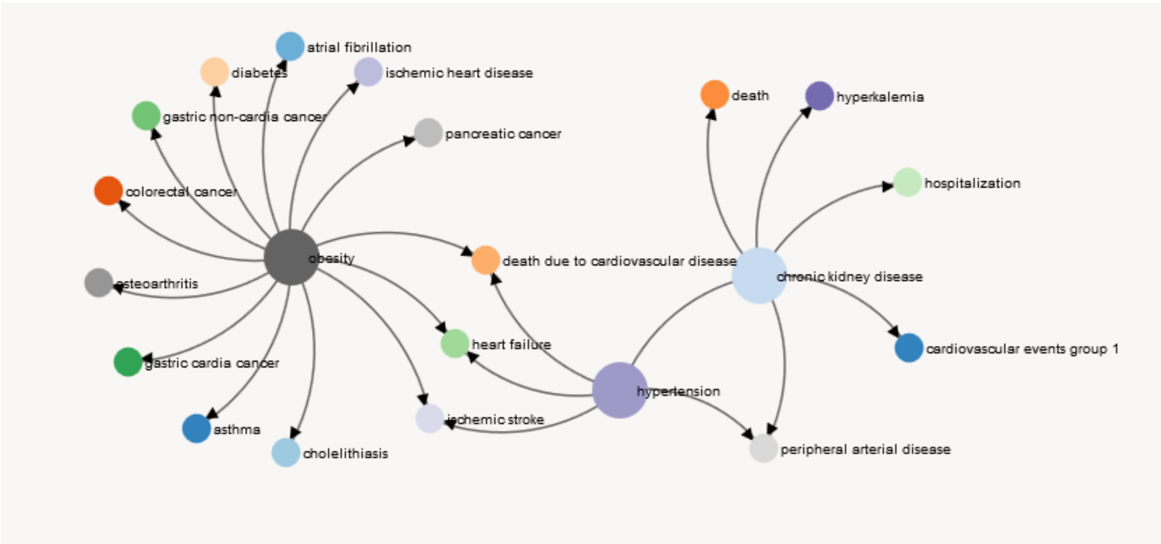


Figure 14. Patient specific risk factor node-link diagram.

2.3.2. Matrix

Although node-link diagrams are capable of presenting the overall structure of the connections, density has a strong impact on readability in node-link diagrams. Alternatively a network can be presented by an adjacency matrix, where rows and columns refer to nodes in node-link diagrams and cells refers to relationships. Compared to node-link diagram, when there is are large number of connections in a network, the advantage of matrix is that it can present all the relationships in the visualisation while node-link diagrams will inevitably result in excessive edge crossings and hairballs. However, the effectiveness of a matrix diagram is heavily dependent on the order of rows and columns: if related nodes are placed close to each other, it is easier to identify clusters and bridges.

Figure 15 is an example view in CARRE visual analytics to visualise causal relationships of risk elements. The filled cell means that there is a causal relationship from the column risk element to the row risk element. The colour represents the disease category and the darkness of the colour represents the number of occurrence of the relationship among all risk factors.

Supported user interactions are:

- Ordering by name, frequency or cluster
- Mouse-hover on the matrix elements to show the information of the corresponding of the risk factor.
- Filtering the diagram based on patient-based diseases, as shown in Figure 16.

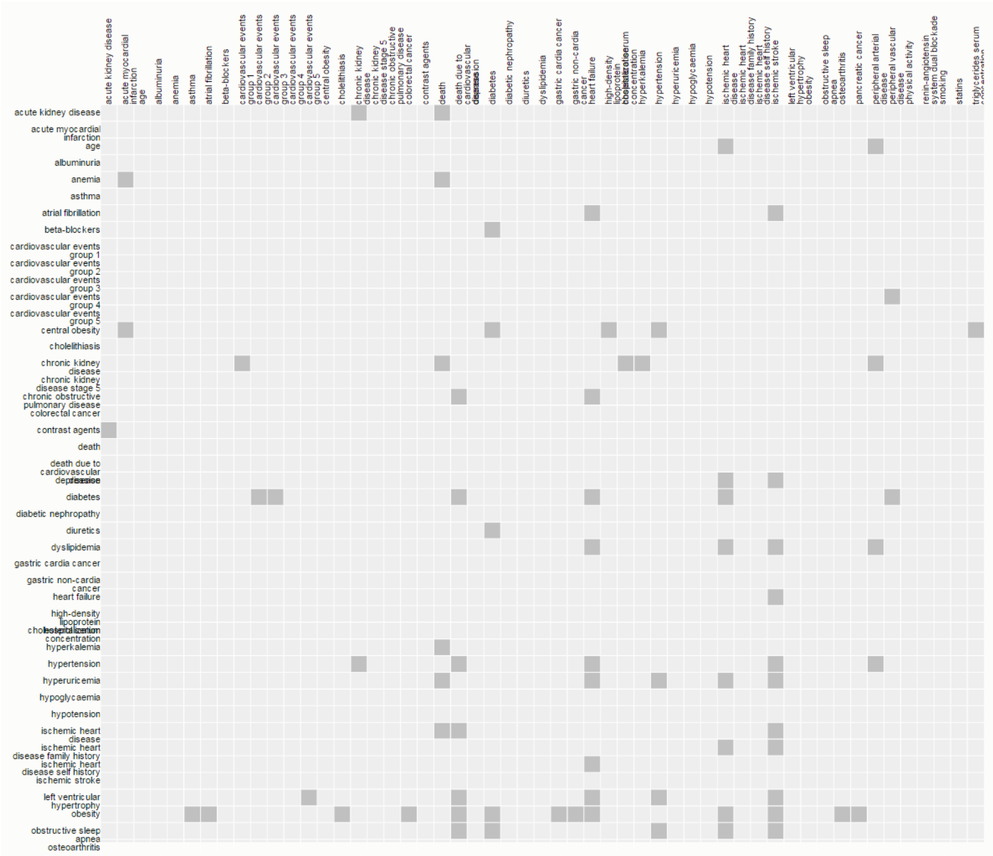


Figure 15. An example matrix view of all risk factors in CARRE repository.

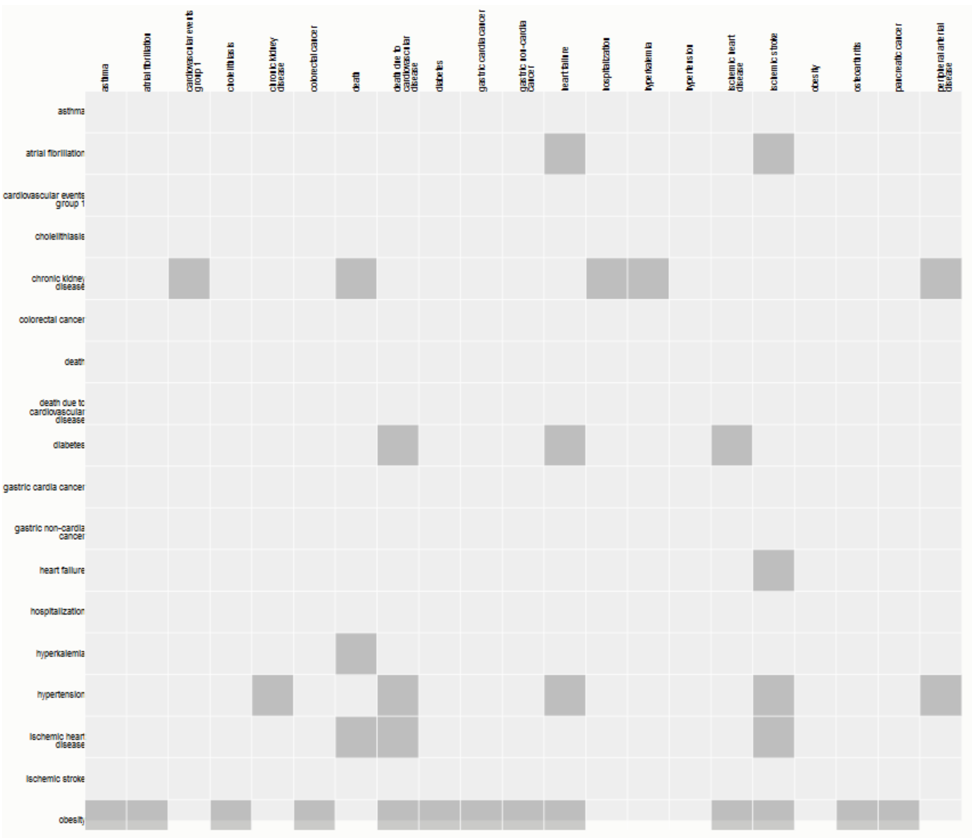


Figure 16. Patient-specific risk factor matrix diagram.

2.3.3. Chord Diagram

The disadvantage of the node-link diagram is that without proper handling, when the number of nodes and links increase, the visualisation will become increasingly messy for effective recognition by human beings, as obvious in Figure 12. Though filtering may be applied based on the conditions of a particular patient, it does not help for the visual analysis of the entire risk factor database. Fortunately there are some network visualisation techniques to alleviate this problem, such as the chord diagram[4], as shown in Figure 17.

The benefits of the chord diagram are that all the nodes are arranged on a circle and the edges from one node are grouped and bundled, which reduces the hairball problems which occur in the node-link diagram. With proper mouse hover interactions all the edges from or to one node can be highlighted, thus making the observation of the connections from or to one node much easier.

The chord diagram clearly visualises the relationships of all risk elements in the repository and is particularly useful when professionals check and insert new risks.

Supported user interactions are:

- Mousehovering on a certain risk elements is to show all risk factors starting or ending at the specific risk element, as shown in Figure 18.
- Filtering the diagram based on patient-based diseases, as shown in Figure 19.

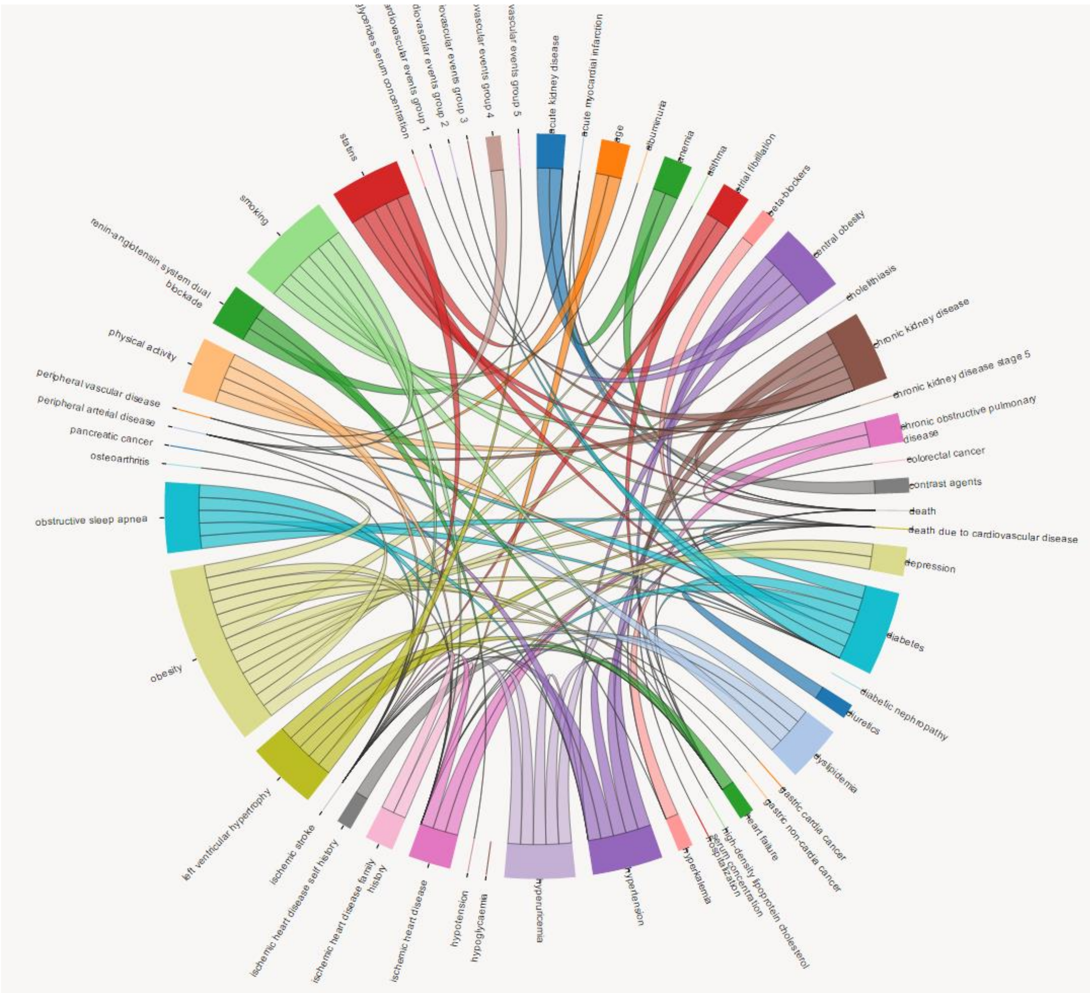


Figure 17. Chord Diagram visualisation of risk factors

[4] Holten D. Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data, IEEE Trans on Visualization and Computer, 2006, 12, (5).

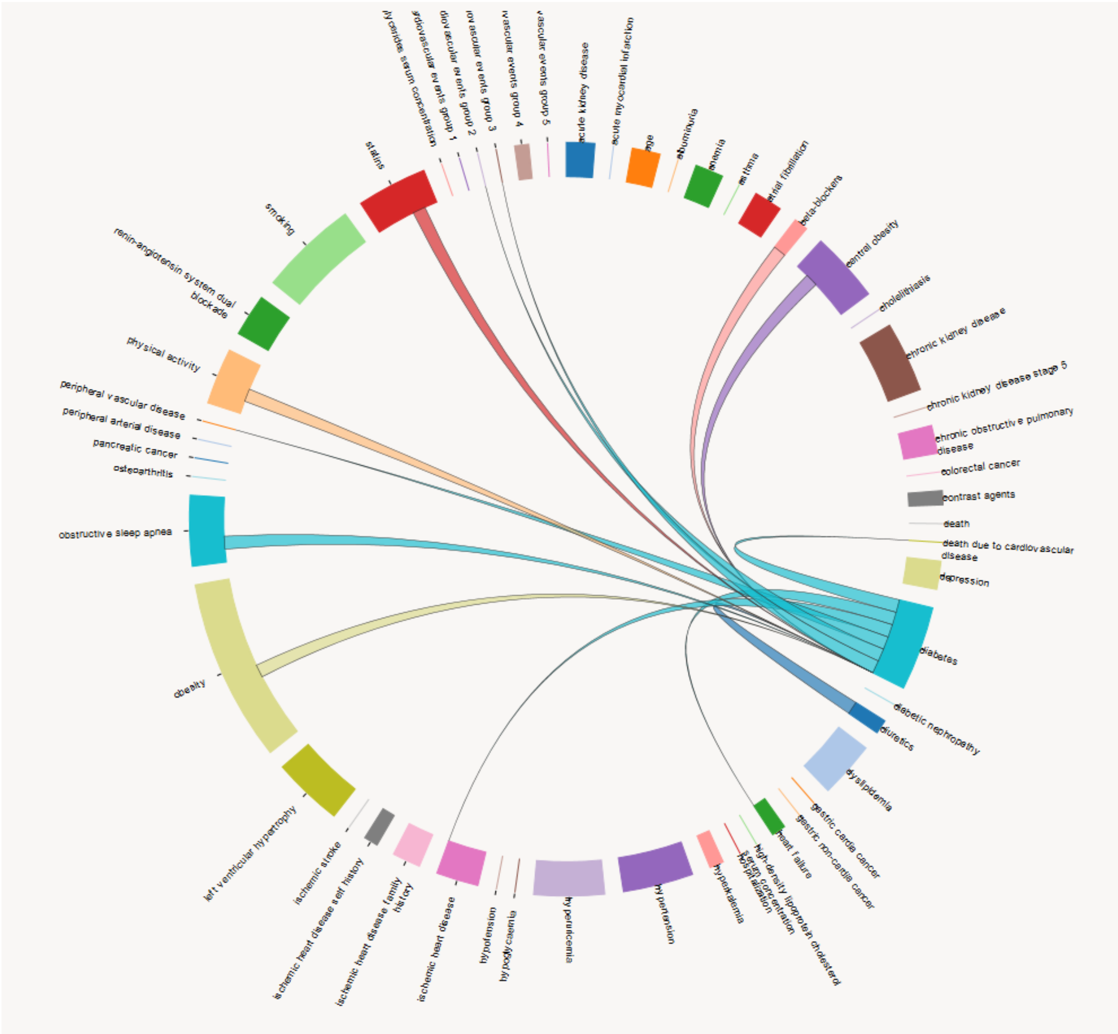


Figure 18. Mouse-hover on a certain risk elements showing all risk factors starting or ending at the specific risk element.

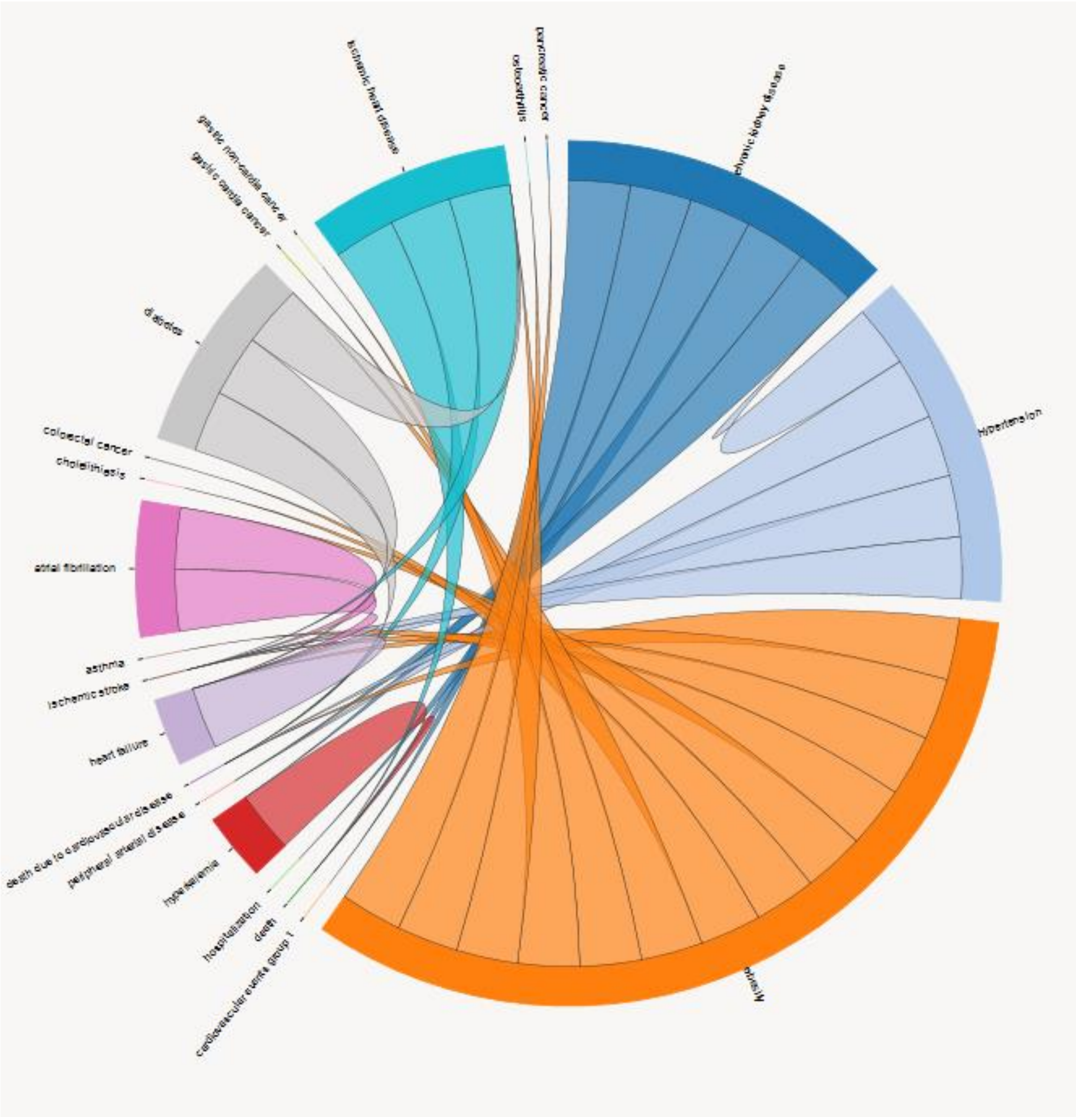


Figure 19. Patient-specific chord diagram.

2.3.4. Sankey diagram

The disadvantage of the chord diagram is that it can only show the associations between the risk elements but cannot show the time order of risk propagations in a static way. A Sankey diagram[5] is a better choice for visualising causal relationships. OutFlow[6] and DecisionFlow[7] use Sankey diagram style visualisation to visually analyse the causal relationships of events. The advantage of the Sankey diagram is that it shows the multi-layer causal relationships of the elements in a much more clear and understandable way than the node-link diagram, though it is not suitable for visualisation of general graph data. Figure 20 shows the Sankey diagram for risk visualisation and analysis in CARRE. The leftmost nodes represent the risk observables and other nodes are risk elements. It is fairly easy to identify the major risk observables relating to a disease (risk element) and to recognise the routes of risk propagation from the visualisation. Supported user interactions are:

[5] P. Riehmann, M. Hanfler, and B. Froehlich. Interactive sankey diagrams. In Proc. IEEE Symposium on Infomation Visualization, InfoVis 2005, pages 233–240. IEEE, 2005

[6] Wongsuphasawat, K., Gotz, D.: 'Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization', IEEE Trans. on Visualization and Comp. Graphics, Dec 2012, 18, (12), pp. 2659–2668

[7] Gotz, D.: Stavropoulos H.: Decisionflow: 'Visual analytics for high-dimensional temporal event sequence data', IEEE Trans. on Visualization and Comp. Graphics, 2014, 20 (12)

– Dragging a risk element to move it and all related risk factor connections to a user specified position
– Filtering the diagram based on patient-based diseases, as shown in Figure 21.

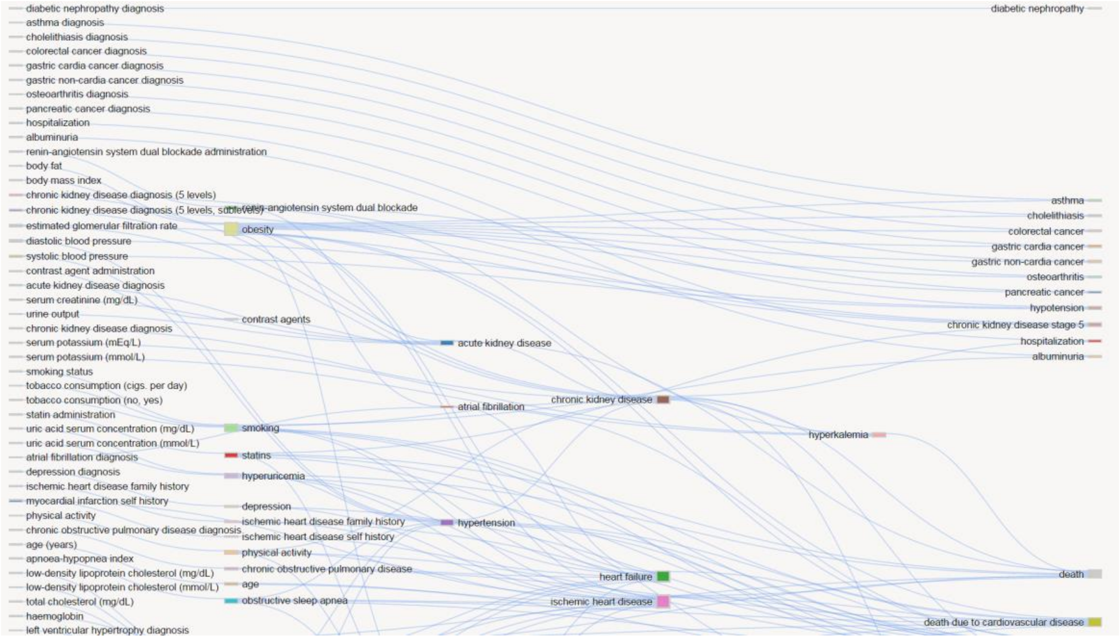


Figure 20. Part of the risk factor Sankey diagram.

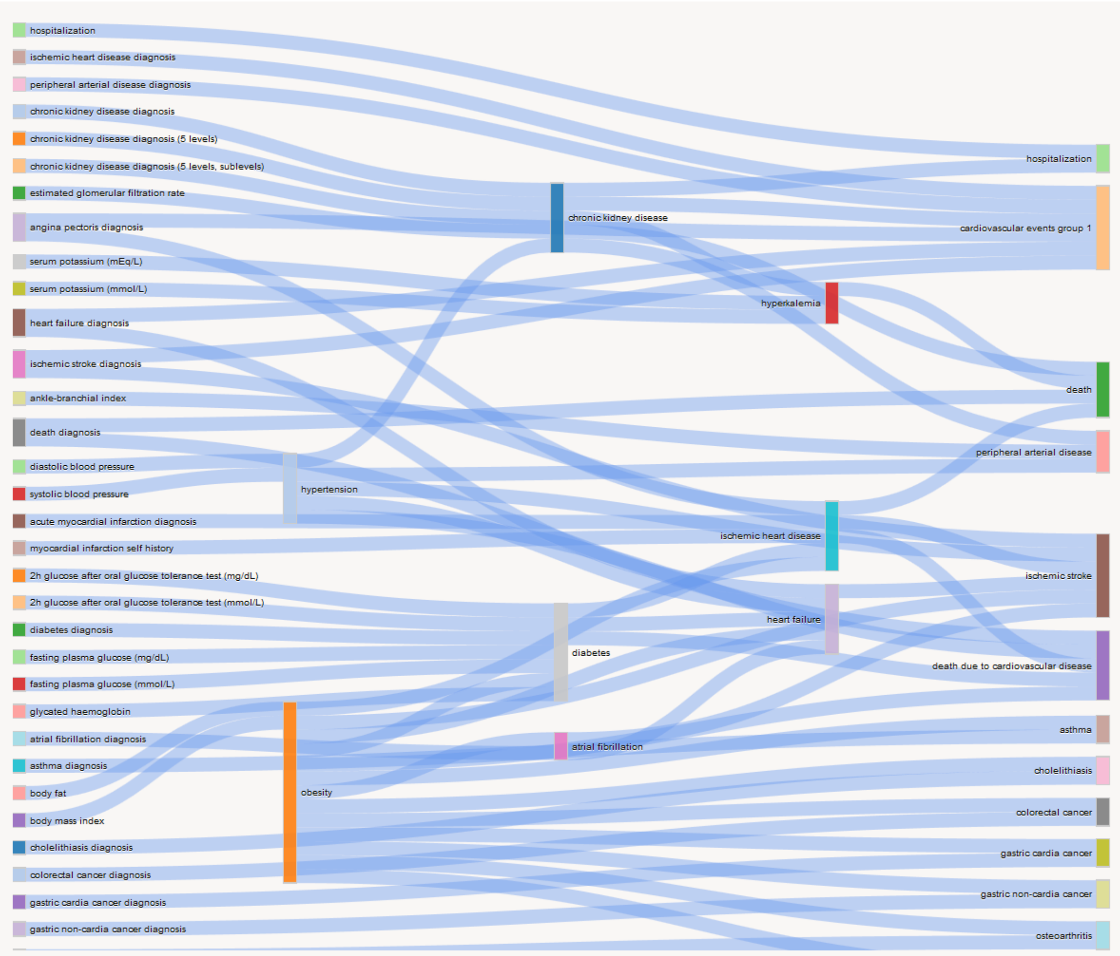


Figure 21. Patient-specific risk factor Sankey diagram.

3. Case Studies

After the design and initial implementation, we conducted survey amongst normal users, which include people with knowledge of visual analytics and people who are researchers but not familiar with the visual analytics. The purpose of this survey is to evaluate the function, usability, clarities of the visual diagrams, and efficiency of the diagrams. The survey is documented in the Appendix A. The survey helped us in designing the suitable diagram to visualise the data collected and used for CARRE scenario. The project will conduct a formal evaluation in a later stage (T.7.4), which will be documented in another deliverable (D.7.4).

In this section, four use cases for visual analysis are designed and described. We will explain the visualisation functions by mapping them onto the four high-level conceptual use cases defined in D.2.1. The purpose of use cases is to select representative scenarios in CARRE visual analytics and demonstrate how visual analysis can help patients and medical experts to achieve their goals.

We use virtual patient data to explain the visualisation interfaces and functions that will be provided by CARRE visual analytics to support general patients and medical experts. The use case diagram is show in Figure 22.

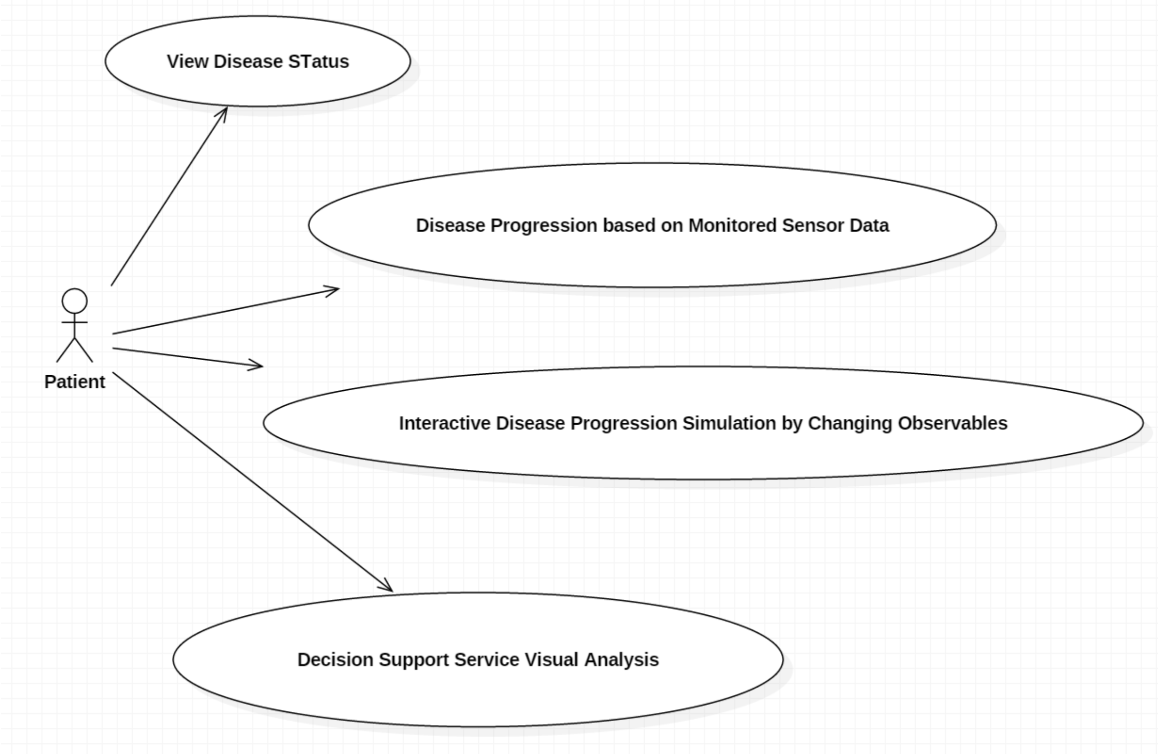


Figure 22. Use case diagram.

3.1. View Disease Status

3.1.1. Introduction

The goal of this use case is to demonstrate how patients can use visual analytics to visualise and understand their disease status interactively. Patients should be able to view those risk associations in an interactive and explorative way.

3.1.2. Sequence diagram

The sequence diagram of View Disease Status use case is shown in Figure 23.

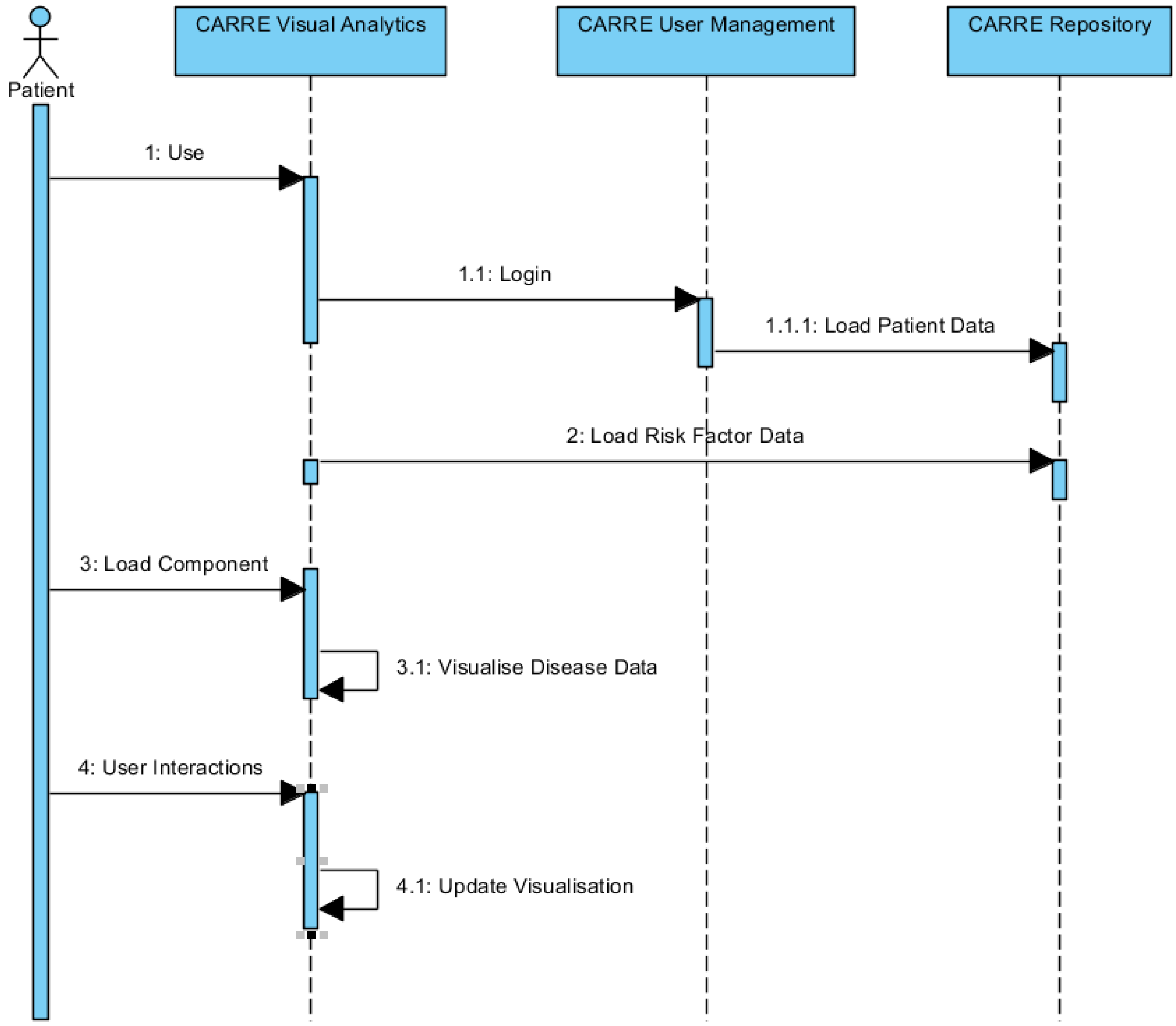


Figure 23. Sequence diagram of View Disease Status use case.

3.1.3. Components

When the user logs in, they can use the healthlines, node-link diagram, chord diagram or the Sankey diagram to view the related risk associations that are specific to their own health status. Figure 24 shows the related risks associated to virtual patient 2 in a node-link diagram. Figure 25 is a Sankey diagram of related risks factors and risk observables of virtual patient 2 where the risk flow can be observed easily.

3.1.4. Visual Analysis

The basic analysis operations include:

1. Select health indicators to visualise
2. Hide unwanted risk factors
3. Highlighting all risk factors related to a certain risk element
4. Clustering of the disease status if there are longitudinal data available
5. Find similar patterns of the disease status if there are longitudinal data available

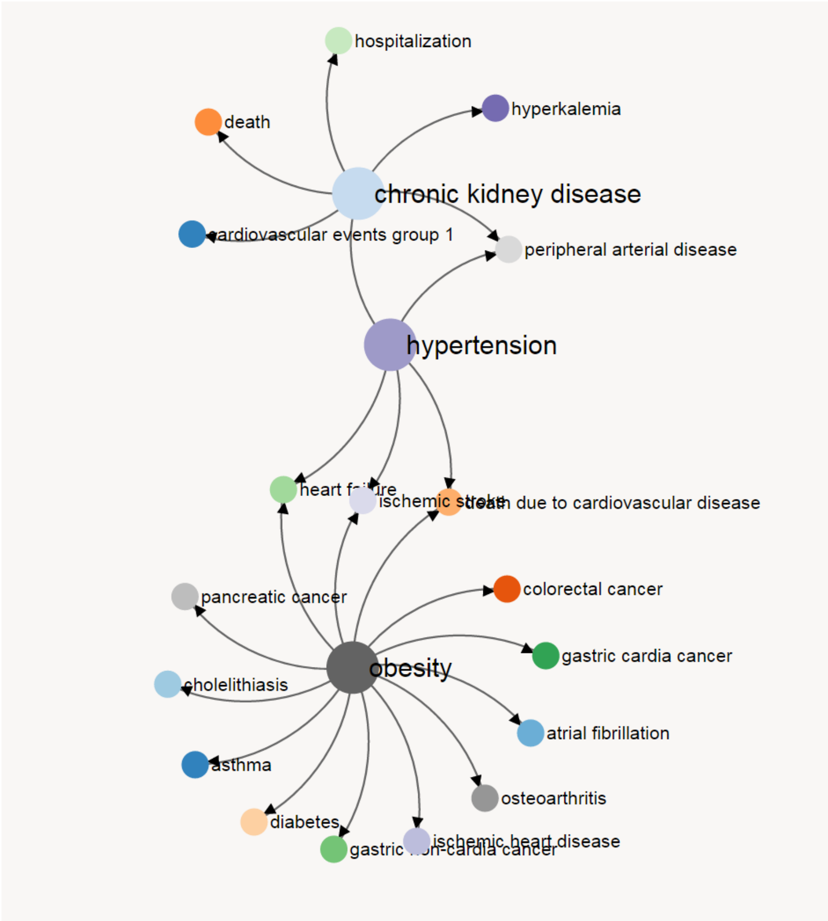


Figure 24. Related risk associations of virtual patient 2

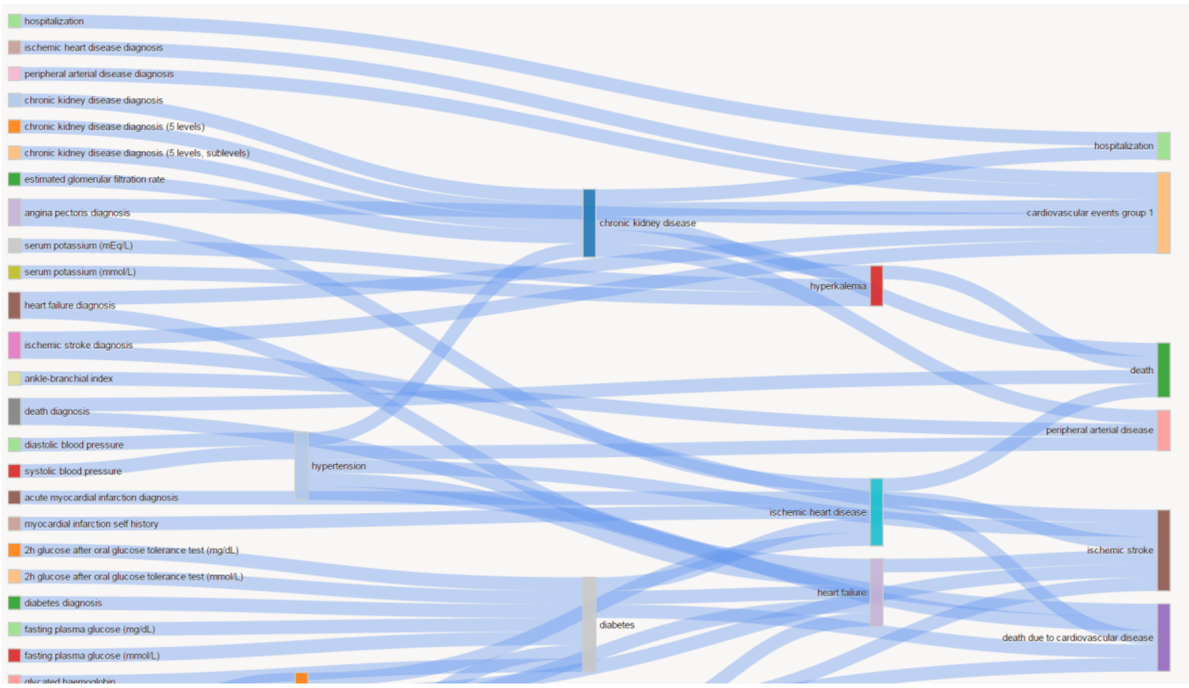


Figure 25. Related risks and risk observables of virtual patient 2

3.2. Disease progression based on monitored sensor data

3.2.1. Introduction

The goal of this use case is to demonstrate how patients can use visual analytics to understand disease progression based on personal monitored sensor data. Some activity tracking monitors, such as Fitbit, also may provide some additional life style information about this patient. In many cases, certain levels of activities are recommended by doctors as part of the disease self-management.

3.2.2. Sequence diagram

The sequence diagram of disease progression based on monitored sensor data use case is shown in Figure 26.

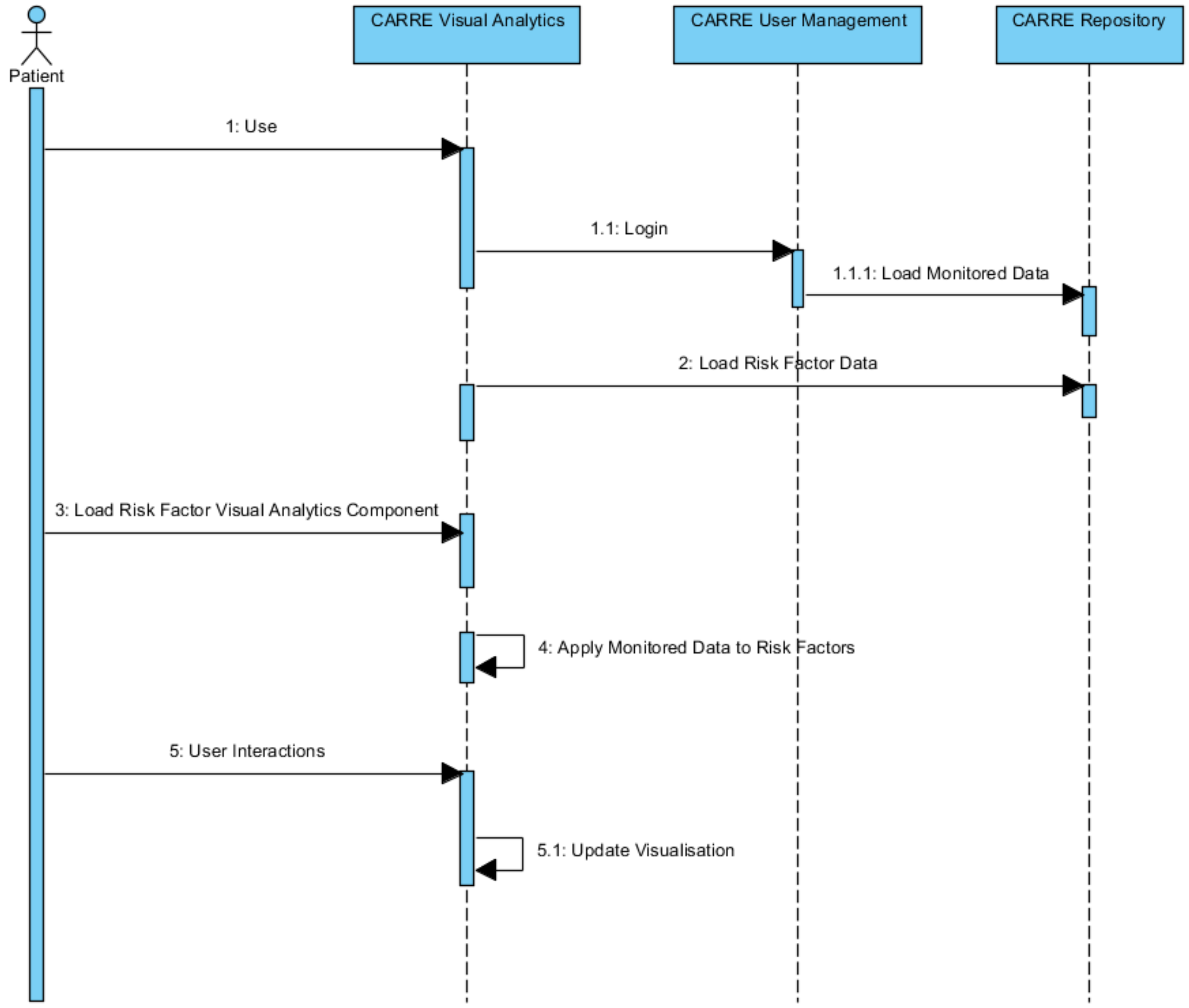


Figure 26. Sequence diagram of use case “Disease Progression based on Monitored Sensor Data”

3.2.3. Components

The purpose of disease progression visual analytics is to visualise related risk factors according to the patient’s health and lifestyle status. The node-link diagram introduced in section 6.1 shows the current disease status, it can also be used to show future disease progressions based on the monitored sensor data and the risk observable data. The analytics process starts from examining the existing sensor data. It then finds the related risk observables and risk factors. By checking the thresholds, it finds potential risks and expands the node-link diagram. Figure 27 shows a disease progression example for virtual patient #2. One can notice that there are new links and nodes which represent the disease progression.

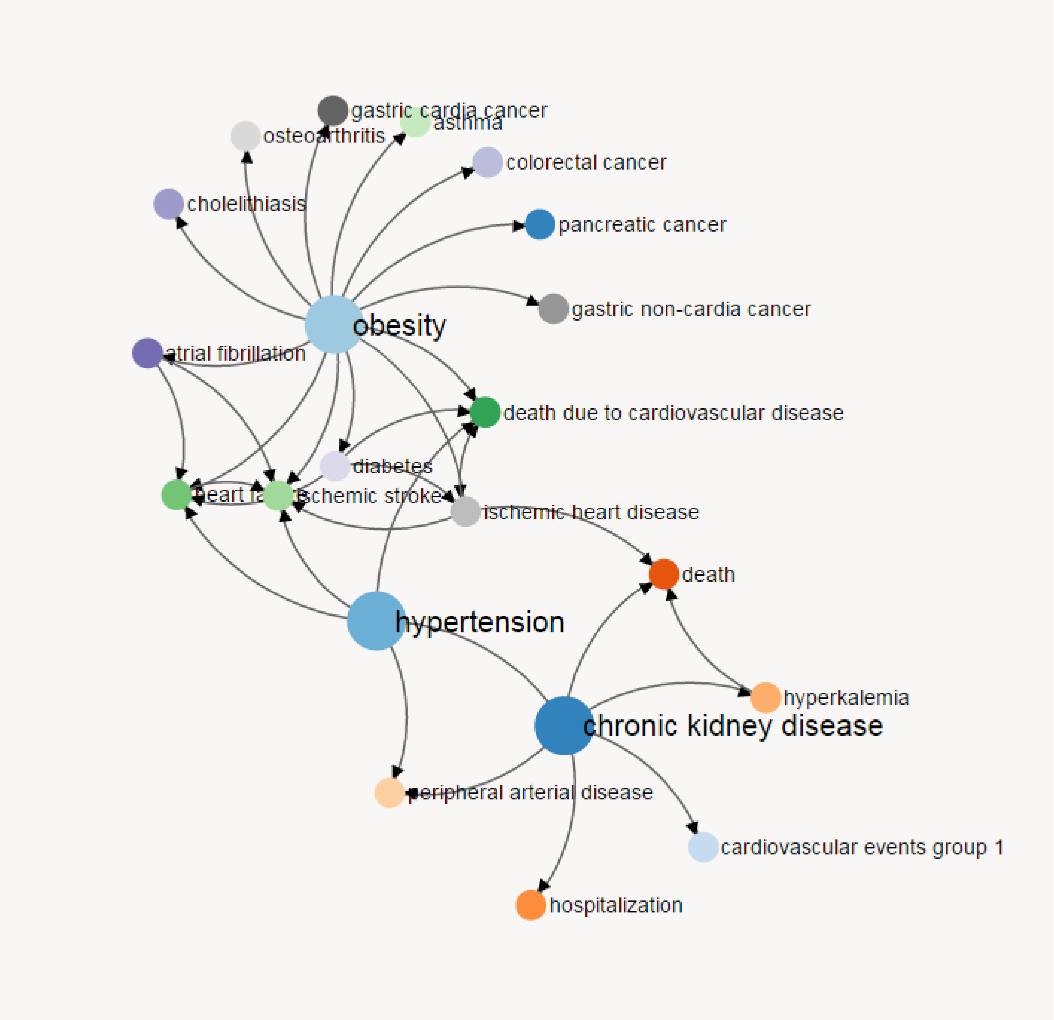


Figure 27. Disease progression based on monitored sensor data for a particular virtual patient.

The basic analysis operations include:

- Select sensor data
- Choose the time range of the sensor data if longitudinal data is available
- Hide unwanted risk factors
- Highlighting all risk factors related to a certain risk element

3.3. Interactive Disease Progression Simulation by Changing Observables

3.3.1. Introduction

The goal of this use case is to demonstrate how patients and medical experts can view and understand the changes in disease prospects interactively if they make changes to the current observables. The purpose of disease progression simulation visual analytics is to visualise related risk factors according to the patient's health and lifestyle status and to visualise the changes that may happen if the user changes his/her lifestyle and/or medical indicators.

The changes may be related to life style, in particular, multiple factors of life style changes, including, activities, food intake, and medicine intake can all be monitored using sensors. This kind of data are often too large and can often only make impact in a long run, such as the consequence of food intake. The functions provided by the visual interface allow users to aggregate the large dataset of the observations in many ways.

3.3.2. Sequence diagram

The sequence diagram of disease progression simulation by changing observables is shown in Figure 28.

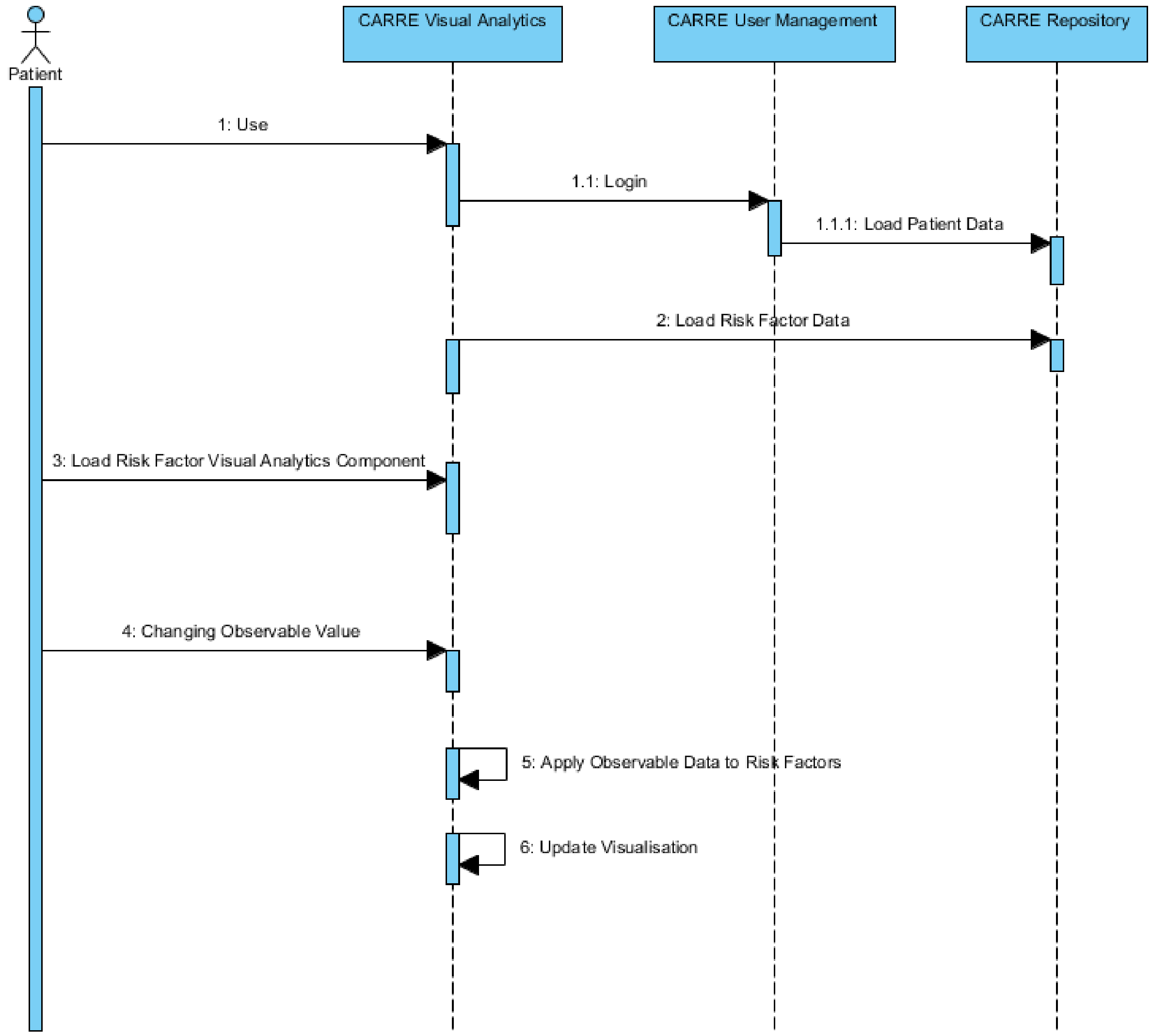


Figure 28. Sequence diagram of disease progression simulation.

3.3.3. Components

Figure 29 shows the user interface of disease progression visual analysis. The central part is a force-directed interactive node-link diagram visualisation of related risk elements and evidences. The right panel is for interactive adjusting of the related observables. Figure 29 shows the updated risk factor node-link diagram when the user increases the step number, which leads to the disappearance of the obesity and related diseases.

The basic analysis operations include:

1. Select risk observables
2. Hide unwanted risk factors
3. Highlighting all risk factors related to a certain risk element
4. Changing the values of the risk observables to view the change in risk factors

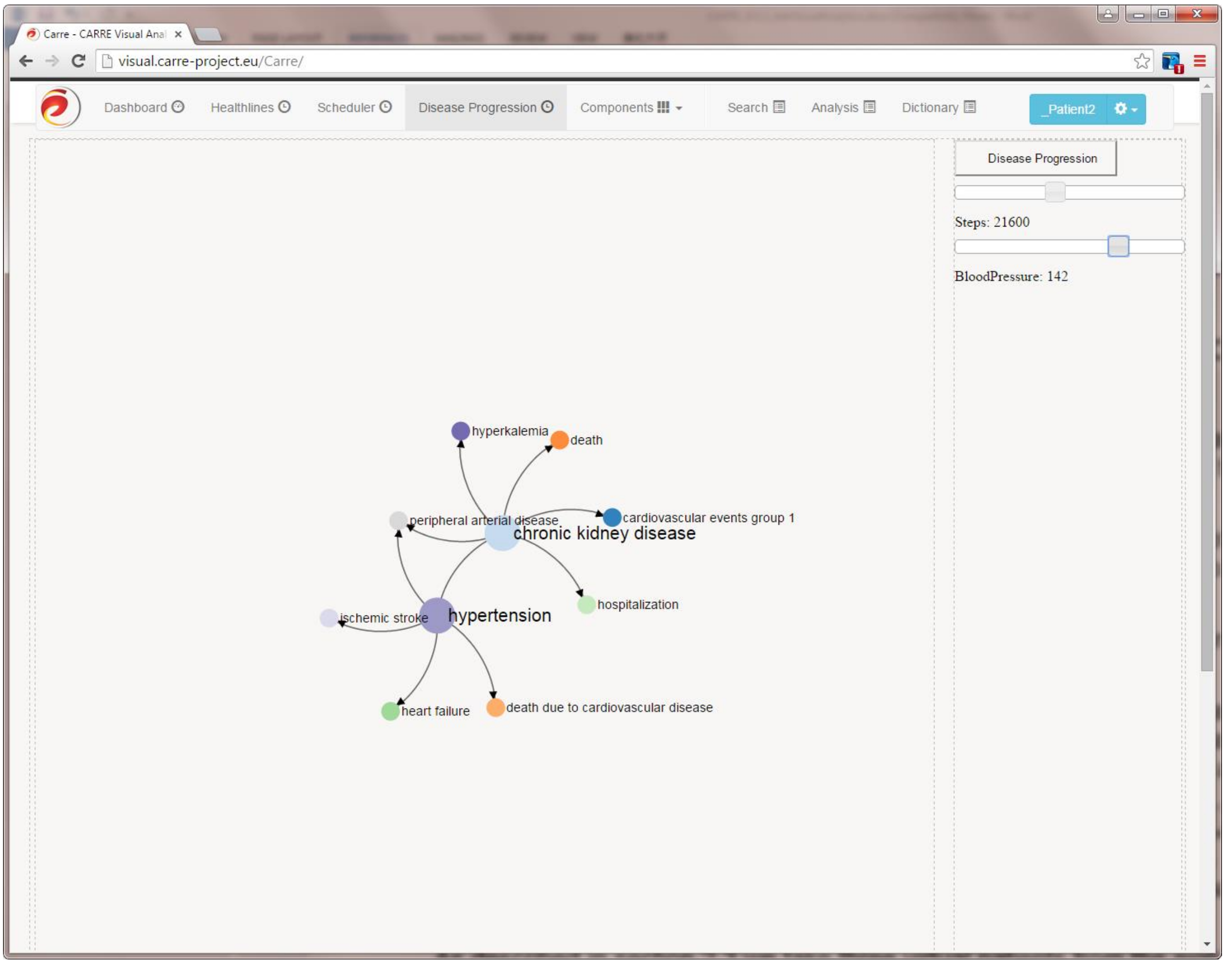


Figure 29. Disease progression visual analysis.

3.4. Decision support service visual analytics

3.4.1. Introduction

Decision support systems (DSS) can assist patients and medical experts by providing them advices, recommendations and diagnosis in cardiorenal domain, where the optimal solutions for a given sort of data about the possible consequences are determined similar as human experts in the field.

In CARRE system decision support service will determine the optimal solution, by mining RDF repositories data to predict future trends and patterns as well as information data analytics and formal reasoning from ontologies, which are the main techniques supported by RDF Linked Data and ontologies.

This method searches particular patient’s observables and assigns risk factors and evidences as well as showing the probability occurrence of a given risk factor and the most suitable risk ratio value.

DSS supports Patient Application and will be the main source of decision recommendations for CARRE. This includes the analysis of the generic and personalised risk model so as to allow the CARRE stakeholders to identify and assess critical medical conditions. Its aim will be to produce meaningful information that will be passed to the end-user interface with the synergy of the visualisation component.

3.4.2. Sequence Diagram

The sequence diagram of DSS visual analysis is shown in Figure 30.

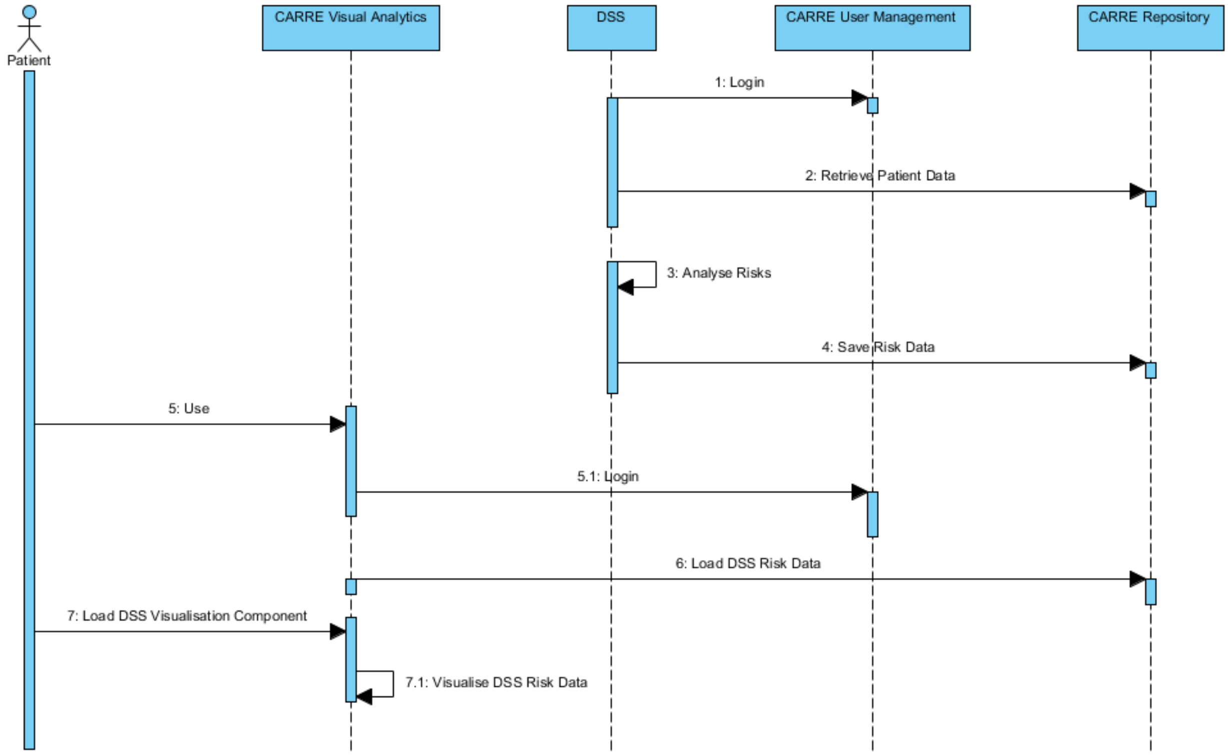


Figure 30. Sequence diagram of DSS visual analysis.

3.4.3. Components

DSS provides the patients and medical experts with tools to make decisions based on sensor data, PHR data and risk factor models. The visual analytics interface will be similar to disease progression visual analysis. The difference is that due to the nature of decision support service, the calculation cannot be completed in real time, which means less user interactivity can be achieved. The DSS result is fetched from the CARRE repository via the DSS API and displayed on the dashboard. The final DSS visual analysis will be completed during the system integration stage.

The basic analysis operations include:

- Set and hide undesired risks
- Risk clustering and comparison if longitudinal risk data is available

4. Implementation and performance

4.1. Implementation technologies

The web-based visual analysis framework is developed in Java, Javascript and HTML and deployed on a Linux environment. The backend is based on Java programming language and Spring Framework[8] technology stack. Java runs on all major platforms include Windows, Linux and Mac OS, which makes module based on Java is generally more portable between different OSs, also JVM's proven high performance is crucial for our potential large user base. Spring Framework is the de-facto standard in enterprise Java programming, the developer team's high experience in Spring Framework makes it our pick on implement the project.

The frontend web-based UI is mainly based on Twitter Bootstrap, HTML, jQuery[9] and AngularJS[10]. jQuery is used to facilitate javascript programming. The interactive visual analysis components are implemented in javascript and d3.js[11] which is a javascript based scalar vector graphics (SVG) library. Each component is placed in a DIV element in the component container and the container is managed automatically as tabs in the main user interface.

The sensor data, risk factor data and decision support data are all fetched from the CARRE public and private repositories. The sensor data can be directly accessed via jQuery APIs while risk factor data and decision support data need to be accessed by SPARQL queries.

The frontend is a relatively separate Grunt project which uses Bower to manage the frontend JavaScript dependencies. The frontend source code is unit tested and built on a continuous integration manner. Twitter Bootstrap 3 is used to support both mobile and desktop browser, mobile first responsive design is applied for the web UI. AngularJS and jQuery JavaScript libraries are used to help shape the frontend logic, and interact with the REST service provided by backend. The combination of cutting-edge frontend technology stack does give us an enjoyable developing experience and a great dynamic user-friendly UI.

The project is hosted on Ubuntu 12.04 LTS VPS and is available at http://carre.ccgv.org.uk:8080/Carre, Apache is used to server static content while Tomcat 7 is reverse proxies to server the dynamic contents. A HAProxy server is configured to work as load balancer and Radis server is tested to be the cache server for scale to larger user groups.

This visual analytics web system uses several open source software, including D3.js for chart drawing, angularJS for the main framework of the web site, jQuery for front end javascript programming. We build up our own front end algorithms in javascript with the help of these libraries. The current implementation utilises Apache Maven for software project management and comprehension.

Code quality analysis has been conducted on the Javascript file at the stage. We use JSLint for code metrics. JSLint is a static code analysis tool used in software development for checking if JavaScript source code complies with coding rules.

---

8 www.springsource.org

9 http://jQuery.com

10 http://www.angularjs.org

11 http://d3js.org/

4.1.1. Code metrics summary

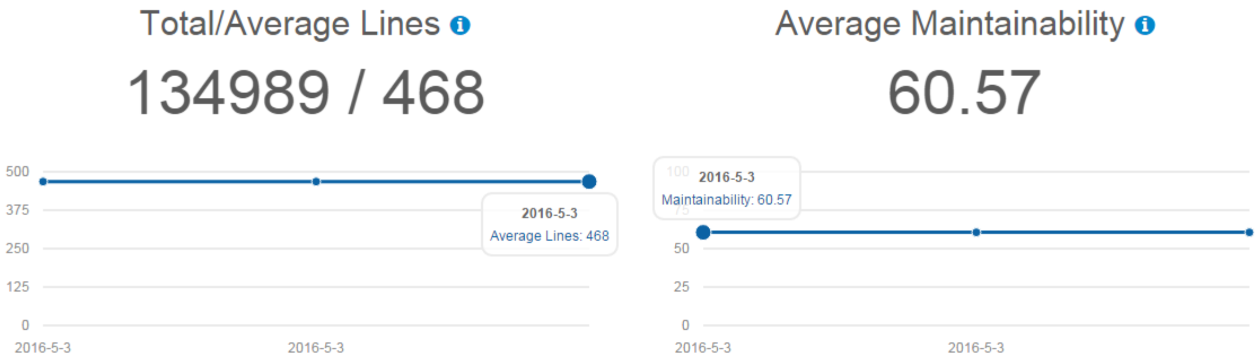


Figure 31. Summary of JSLint analysis result

4.1.2. Maintainability

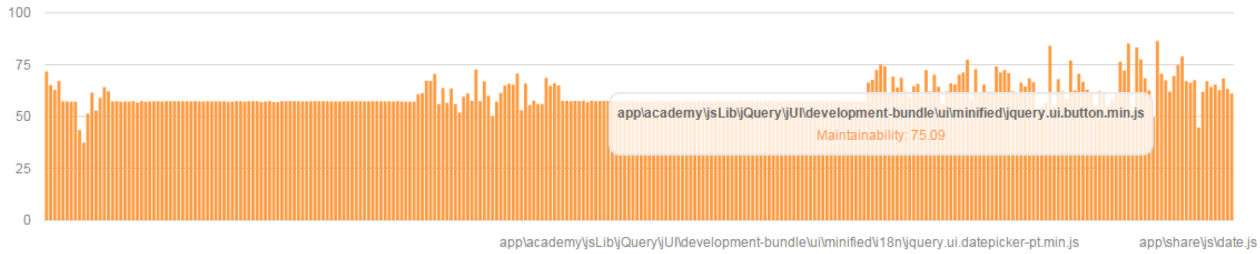


Figure 32. Maintainability of JSLint analysis result

4.1.3. Lines of code

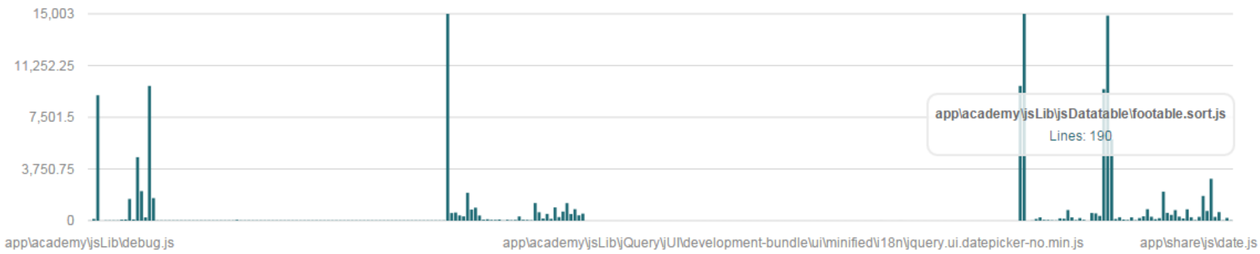


Figure 33. Lines of code of JSLint analysis result

4.1.4. Estimated Errors in Implementation

Figure 34. Estimated Errors in JSLint analysis results

4.2. Scalability to large graph data sets

Scalable visualisation is a general requirement for visual analysis of large scale data. In CARRE, the longitudinal data from a variety of data sources will produce a huge amount of data, without scalable visualisation techniques, CARRE cannot achieve its goal of effective visual analytics.

Scalable visual analytics techniques used in CARRE include:

- Data filtering. Data sources typically contain large amount of data which might be too big for visualisation processing. However, visualisation usually needs only a specific subset of data that meets certain criteria. Users can select specific data by using filters. For example, rather than retrieving information about the full time range, the user can create filters to select data in a given date range. This is almost indispensable for interactive scalable visualisation of large amount of data as it is usually not possible to load the whole dataset into the visualisation application. Data filtering is used by the risk factor visual analysis as well as other components.
- Overview+details. Overview+details technique places an overview of the graphic next to a zoomed "detail view." As the user drags a viewport around the overview, show that part of the graphic in the detail view. This technique is used in the healthlines to show both the overall data shape and the detailed data of the selected time range.
- Clustering. Clustering is a fundamental technique in data mining and visual analysis. Clustering uses unsupervised learning to find structures in a collection of unlabeled data. It organises objects into groups based on similarity of objects. In this way, the number of objects to be visualised is greatly reduced and part of the underlying knowledge in the objects is shown to the user. Hierarchies of objects can be built and used for interactive exploration and analysis. Common techniques of data clustering include k-means, hierarchical clustering, Fuzzy C-means, Gaussian Mixture Model (GMM), etc[12]. Clustering can be used by the risk factor analysis to group nodes with similar properties into the same group to reduce the complexity of the visualization.

4.3. Uncertainty

Uncertainty is common and crucial in a number of fields. The exploration and analysis of three-dimensional (3D) and large two-dimensional (2D) data with uncertainty information demand effective visualisation augmented with both user interaction and data analysis.

Uncertainty visualisation is one of the principal and pioneering research topics in visualisation. As uncertainty is very hard to be eliminated, visualisation techniques are required to present and to explore uncertainty information in an expressive way. Uncertainty visualisation aims to present data in a way that it is easy and intuitive for users to become aware of the locations and degree of uncertainties in the data to make better analysis and decisions.

4.4. Maximum portability using Web technologies.

The web-based visual analysis framework is developed in Javascript and HTML. The web-based UI is mainly based on Twitter Bootstrap, HTML, jQuery. The interactive visual analysis components are implemented in javascript and d3.js13 which is a javascript based scalar vector graphics (SVG) library.

The sensor data, risk factor data and decision support data are all fetched from the CARRE public and private repositories. The sensor data can be directly accessed via jQuery APIs or SPARQL queries.

The cross platform nature of HTML and Javascript implies the high portability of the visual analytics module. It runs on a variety of platforms (including Windows, Linux, Android, iOS, etc.) and browsers (including Google Chrome, Firefox, Safari, etc.).

[12] Lior Rokach, Chapter 14 A survey of Clustering Algorithms, Data Mining and Knowledge Discovery Handbook (2nd ed), 2010

[13] http://d3js.org/

4.5. Clarity, usability and aesthetical aspects

In the design of CARRE user interface, each component is placed in a DIV element in the component container and the container is managed automatically as tabs in the main user interface. It is easy to add, remove or reposition a certain component or group components under a new tab to improve the clarity of the system. Visualisation components are designed according to the general principles recognized in the visualization design community. The use of colours also follows the colour design rules to achieve better usability and aesthetical standard.

5. Conclusion

Visual analytics is an integral approach combining visualisation, human factors, and data analysis. This process combines automatic and visual analysis methods with a tight coupling through human interaction in order to gain knowledge from data.

The target of Work Package 5 in CARRE project is to provide effective data management and visual analytics tools to empower patients and medical experts to access, view, understand and analyse patients' health status and possible disease progressions.

In this deliverable report of D5.3, the design and implementation of task 5.1 "Advanced Visual Analytics" is presented. It is based on D5.1: "Interactive Visual Interface". Three virtual patients are selected for visual analytics use case demonstration. Visualisation components such as charts, timeline are used for visual analysis of sensor data and PHR data. Node-link diagram and matrix are used for visual analysis of risk factor data and decision support data.

A web-based CARRE visual analytics web system is provided as the hosting system of the visual analysis components. A dashboard is designed for an overview of the user's health status and functional visual analytics components are organised in visual analysis containers for different visual analysis tasks.

The design and implementation meets the demands of visual analytics of sensor data, PHR data, risk analysis, disease progression analysis and decision support analysis for patients and medical experts.

Annex 1

Design Study Evaluation

1. Introduction

In D5.1, we described initial design of the visual analytics interfaces. Later, based on the discussions with the users and developers, we came up with a set of graphic descriptions for the data that need to be visualised in CARRE. The following section will explain the functions of the graphic descriptions. Dashboard is the start point for everyone, and it also includes the summary data for the logged in user. Users of the interface are advised to use one or more graphics to visualise the data collected from sensors (including the ones from third party and the CARRE products), some graphics are particularly designed for visualising risk factors in various formats that suit different use cases.

After the design and initial implementation, we conducted a survey amongst normal users, which include people with knowledge of visual analytics and people who are researchers but not familiar with the visual analytics. For the latter group, we gave some briefing on CARRE project and the visual interfaces. The purpose of this survey is to evaluate the function, usability, clarities of the visual diagrams, and efficiency of the diagrams. The purpose of the evaluation is to make sure that we provide the suitable diagram to visualise the data collected and used for CARRE scenario. In the later section, we will explain how to use the visual functions for patients to see their risks, risk progressions based on several popular use cases; and then followed by a discussion on using visual analytic techniques to make further analysis, such as clustering.

1.1. Questionnaires – General Information

Gender?		
Male	Female	Prefer not to say

Age?					
< 20	20 -35	36 - 45	46 - 55	56 - 65	> 65

Highest level of education?			
No Academic qualifications	Elementary school	Vocational qualification (e.g. technical college)	Higher degree

Have you ever had a job in healthcare?	
No	Yes
Other:	Yes as a Provider (e.g. nurse, doctor, etc.)
	Yes as a Researcher (e.g. university, public health, etc.)
	Yes as Other (e.g. hospital manager, health charity, pharmacist, etc.)

Do you have any long term health conditions?			
No	Yes		
	If yes please choose all that apply:		
	Prefer not to say	Heart disease	Chronic bronchitis or emphysema
	Diabetes	High blood pressure	Back problems and/or arthritis
	Asthma	Anxiety or depression	Alzheimer's disease or other dementia

	Fibromyalgia	Migraine headaches	Problems related to alcohol or drugs
	Effects of stroke	Cancer	Bowel disorder
	Epilepsy	Thyroid condition	Urinary incontinence
	HIV/AIDS	Cataracts	Sexually transmitted disease
	Glaucoma	Other:	

How would you rate your computer skills?						1 = Novice / 10 = Expert			
1	2	3	4	5	6	7	8	9	10

Are you a member of any of the following social networking services?				
Facebook	Twitter	LinkedIn	XING	Google+
Other:				

Have you ever participated in other CARRE project surveys?	
Yes	No

Have you ever heard anything about Visual Analytics?		
No	Yes	
	If yes where did you hear it from?	
	The media (e.g. newspaper, radio, television)	The Internet (e.g. news article, Google)
	The hospital (e.g. GP)	Word of mouth (e.g. friend or relative)
	Other:	

Have you used Visual analytics tools to analyse data?	
Yes	No

1.2. Questionnaires – Visual analytics functionalities

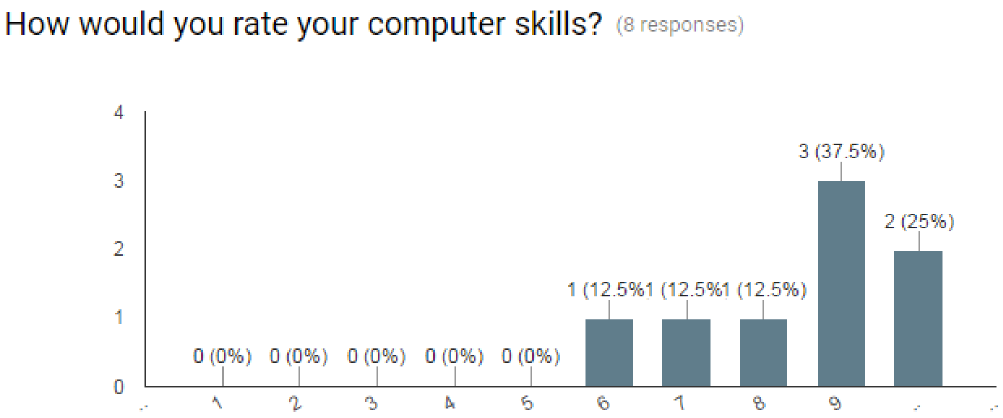
Functions						
Healthlines		Rating (1 low, 5 high)				
	(software quality characteristics)	1	2	3	4	5
Functionality	Can this function tell you the measurements collected over a time period?					
	Are graphs displayed sufficient information?					
	Are these diagrams easy to understand?					
Efficiency	How quickly do you see the graphs?					

	How smooth are the healthline operations?					
	Disease progression	Rating (1 low, 5 high)				
	(software quality characteristics)	1	2	3	4	5
Functionality	Can this function show clear information of a disease progression?					
Efficiency	Do graphs display sufficient information?					
	Are the diagrams easy to understand?					
	Components (data correlations)					
	(software quality characteristics)	1	2	3	4	5
Functionality	Can this function tell you the likelihood of correlations between measurements?					
Efficiency	Do graphs display sufficient information?					
	Are the diagrams easy to understand?					
	Components (Matrix view)					
	(software quality characteristics)	1	2	3	4	5
Functionality	Can this function show clearly the relation between two risk factors?					
Efficiency	Do graphs display sufficient information?					
	Are the diagrams easy to understand?					
	Components (Chord view)					
	(software quality characteristics)	1	2	3	4	5
Functionality	Can this function show clearly the relations between two risk factors?					
Efficiency	Do graphs display sufficient information?					
	Are the diagrams easy to understand?					
	Components (Sankey view)					
	(software quality characteristics)	1	2	3	4	5
Functionality	Can this function show clearly the relations between two risk factors?					
Efficiency	Do graphs display sufficient information?					
	Are the diagrams easy to understand?					
The visual analytics interface						

Usability						
	How quickly do you see the graphs?					
	Can you use the functions without further explanations?					
	Does the system utilize CPU and memory efficiently (answer this if you can)					
	When checking the relationship between risk factors, which diagram is easy to understand: matrix view, chord view or sankey view?					
	When checking the relationship between risk factors, which diagram provides richer information: matrix view, chord view or sankey view?					
Reliability	Have most of the faults in the software been eliminated over time?					
	Is the software capable of handling errors?					
	Can the services resume working & restore lost data after failure?					
Maintainability	Can the software be tested easily?					
Portability	Can the software be moved to other environments?					
	Can the software easily replace other software?					
Quality in use	How accurate and complete is the software for the intended use?					
	Does the software improve the time or reduce resource for the intended goal? (say, it has the information you need with regards to personal health status)					
	Does the software satisfy the perceived achievement of pragmatic goals? (say, help you understand your personal health)					
	Can the software harm people in the intended context of use? (say, the disease progression)					

1.3. Results

Figure 35 shows the general background of the people that we surveyed.



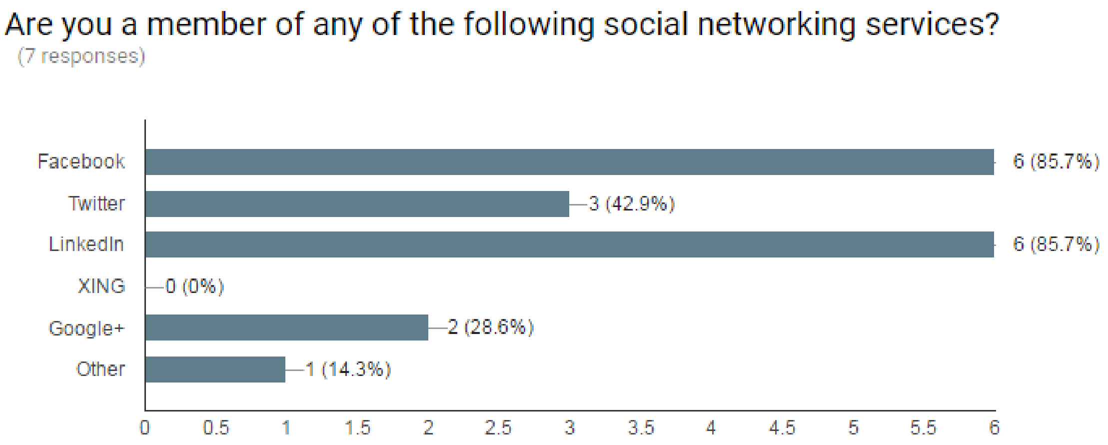


Figure 35. The general background of the people that we surveyed.

Figure 36 shows the survey results for healthlines, which displays the biomarker results over a time period. The graph view is straightforward to people. With regards to the insufficient information, we think the question itself is a bit misleading. After some further discussion, we realised that the problem is about the usability of the function. It is not that straightforward for users to realise what to do. We plan to add relevant 'help' manuals.

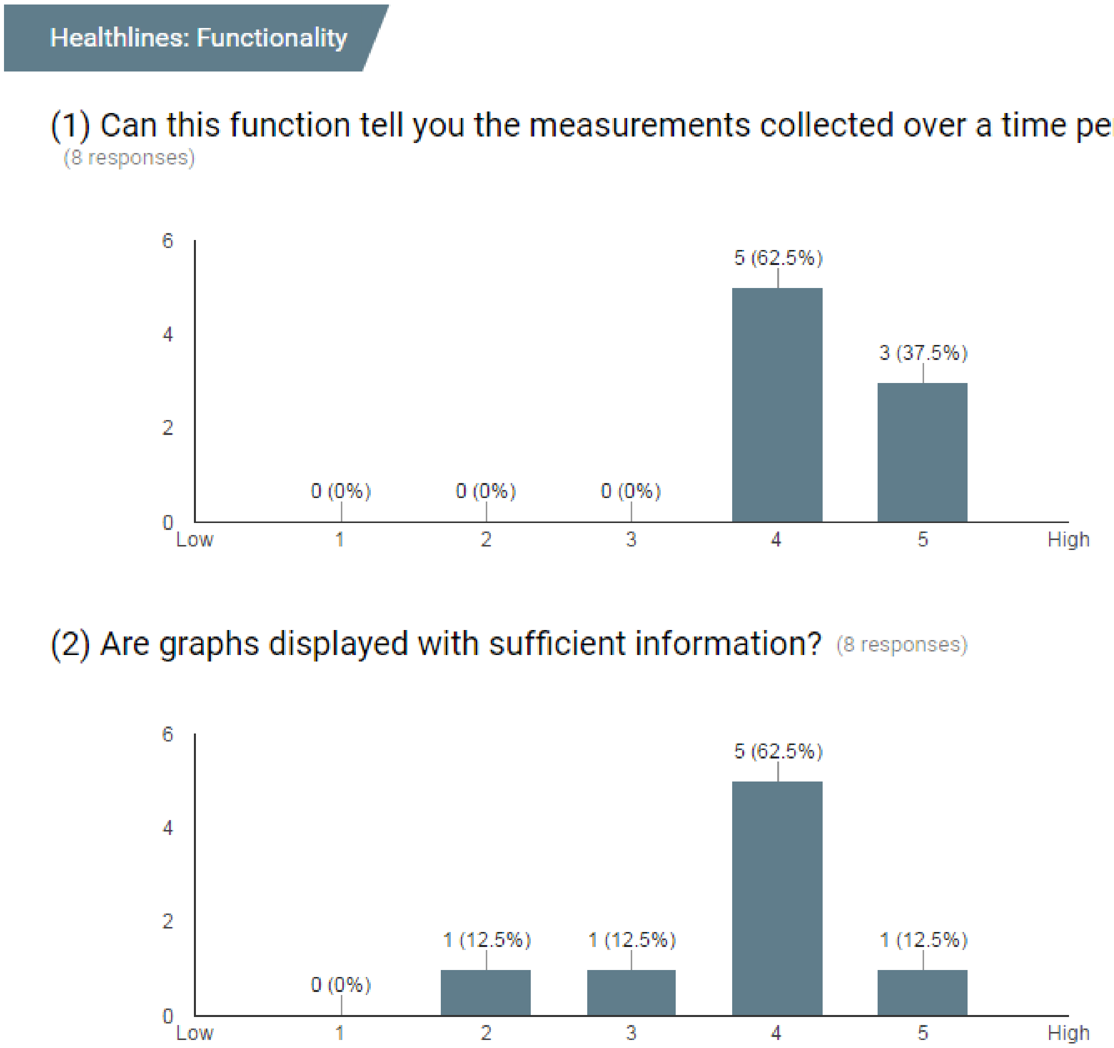


Figure 36. Survey results for the Healthlines module

Figure 37 shows the disease progression, which from the survey, it is well accepted.

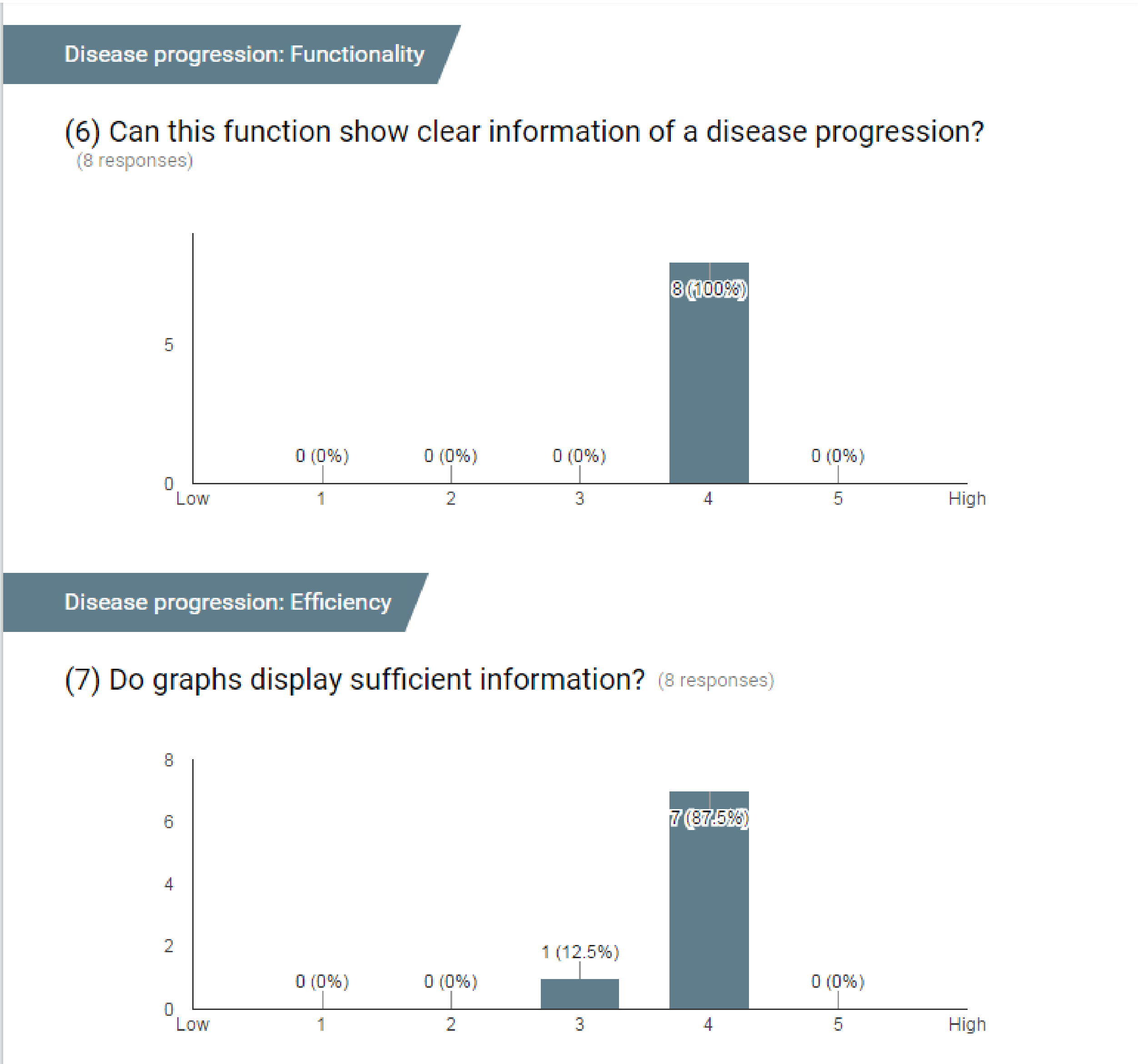


Figure 37. Survey results for the Disease Progression module

Figure 38 deals with the function under the component button. Data correlation is generally well accepted within this survey group, which is as we anticipated. The members of the group are all well trained as researchers who are used to explain and analyse graphs as such. On one hand, we decide not to include this graph for use cases, and on the other hand, we will keep this graph for medical professional.

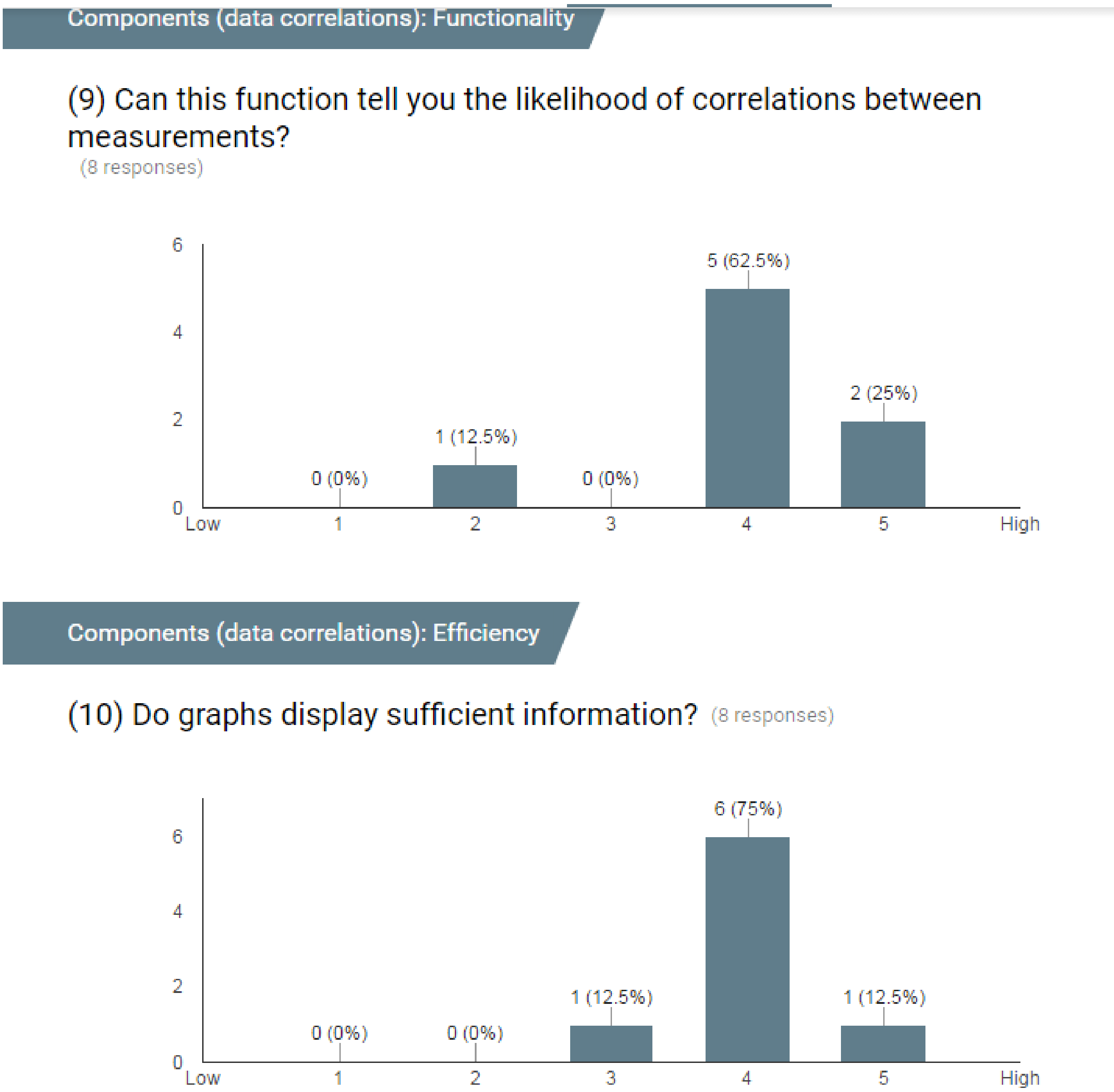


Figure 38. Survey results for the data correlation analysis module

Figure 39 displays easy to spot relations between risk element, and it is generally well accepted.

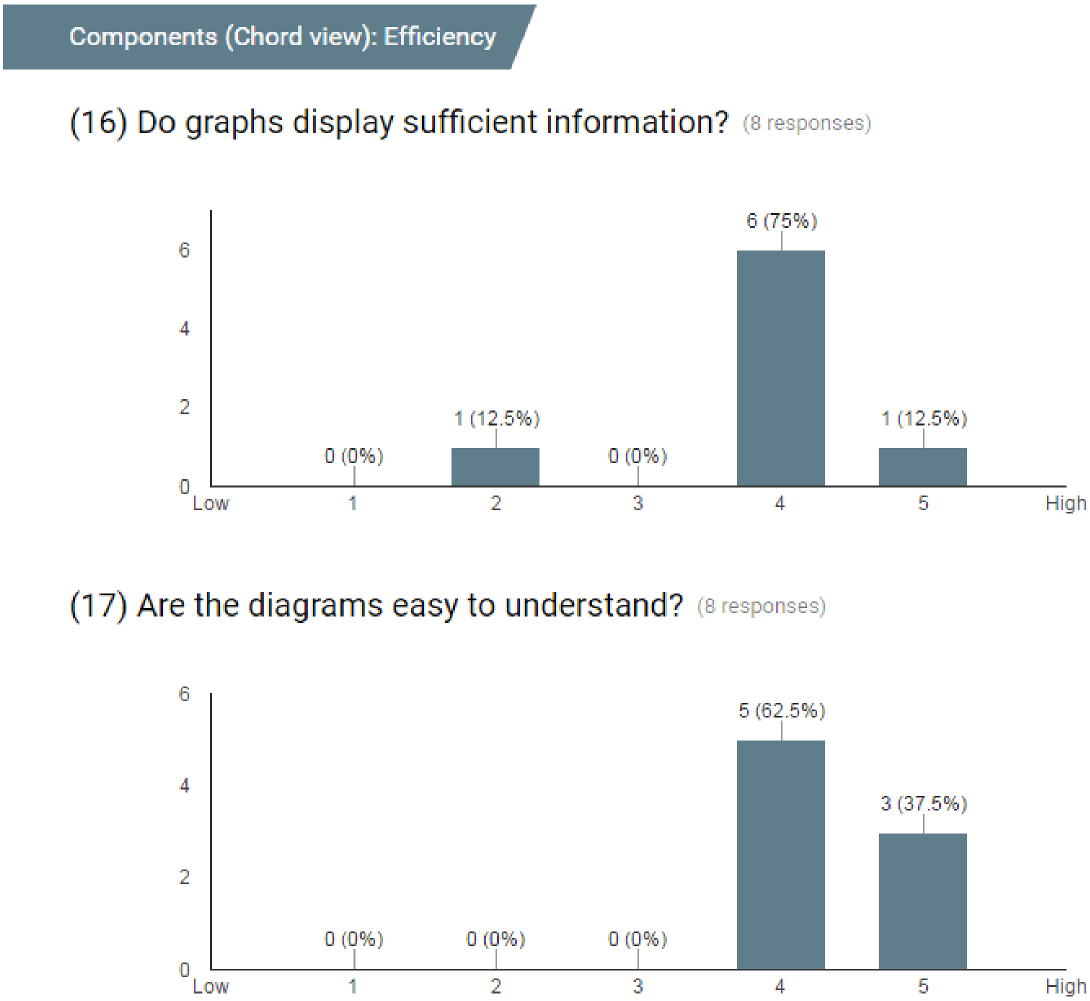


Figure 39. Survey results for the Chord Diagram

In the personal case, the diagram works even better with less elements. We received split results for the Sankey diagram, as shown in Figure 40. With one person strongly disapprove it. As we described, this is a small group of survey, but the people who participated are all experts in certain research area, so we value their opinion. Under this condition, a strong negative response needs us to take further investigation. The possible update to the graph is that we will use different colour or different thickness of the lines.

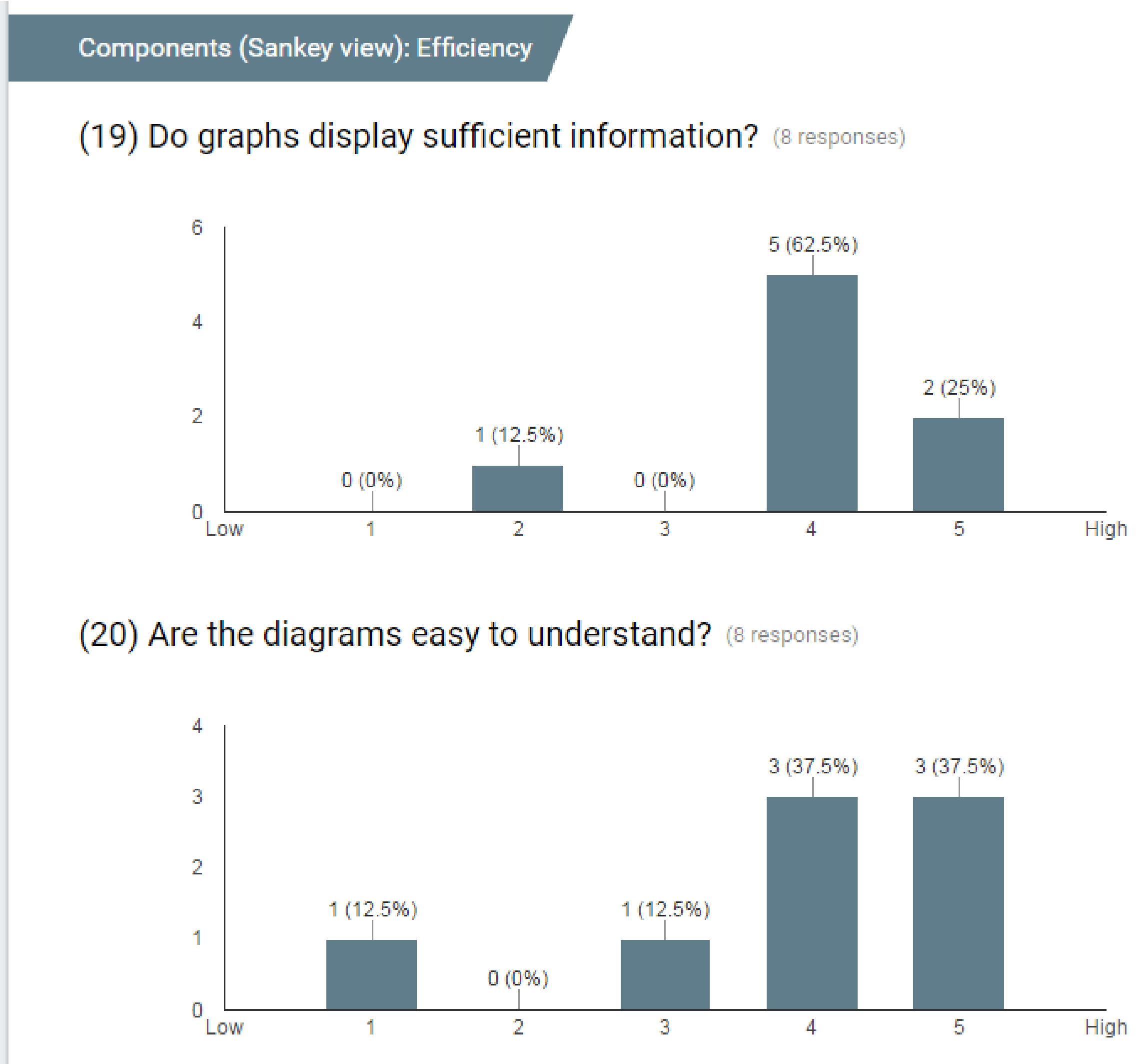


Figure 40. Survey results for the Sankey Diagram

The study results show that the visual informatics tools effectively visualise the data and are easy to understand and interact. The user interface provides most of the needed information. The evaluation results also remind us that we need to make the visualisation tools easier to use to meet the user's expectations.

Annex 2

Visual Analytics Software

What is CARRE: Visual Analytics Module?

The focus of the CARRE **Visual Analytics Module** is to cover the following main challenges:

(a) scalability to large graph data sets

(b) clarity, usability and aesthetical aspects

(c) maximum portability using Web technologies

Additionally, it includes the design of the graph visualization and description of scalable implementation of the graph visualization incorporating lessons learned from the summative usability evaluation.

Download

Software of Visual Analytics Module:

- **v1.0** (Released 15 May 2016, Deliverable 5.3)
- Source code: CARRE_Visual_Analytics_Module_v1.0.zip (JavaScript code)

The CARRE Visual Analytics Module is Open Source