Navigating health literacy using interactive data visualisation

Enjie Liu, Youbing Zhao, Hui Wei
Department of Computer Science and Technology
University of Bedfordshire, Luton, UK

Stefanos Roumeliotis, Eleni Kaldoudi
School of Medicine, Faculty of Life Sciences
Democritus University of Thrace, Alexandroupoli, Greece

***Abstract*— It is commonly concluded that health literacy focuses on individual skills to obtain, process and understand health information and services necessary to make appropriate health decisions. To achieve this, an individual first needs to obtain an adequate level of health literacy. However, nowadays, the information that individuals encounter with regards to their health, the amount, credibility and quality of the data make it difficult for one to make judgments on their health and disease progression, let alone make informed decisions on behaviour change. In this paper, we will report our work in providing patients with efficient ways to explore and understand the relevant health literacy. We focus on two data types: 1) harvested medical evidence from PubMed on cardiorenal disease and its comorbidities; 2) data collected from patients including from PHR and wearable sensors. Our work provides ways for patients to visualise this data meaningfully. Our work aims to improve the health literacy for the general public and increase the population's understanding of the medical field; thus helping users to make informed decision with regards to their care.**

Keywords—patient oriented services, disease comorbidity, risk predictive models, health literacy

I. Introduction

Health literacy has been defined in many different ways since it was first introduced as a term and concept. In 2012, European Health Literacy Consortium noted that: 'Health literacy is linked to literacy and entails people's knowledge, motivation and competences to access, understand, appraise and apply health information in order to make judgments and take decisions in everyday life concerning health care, disease prevention and health promotion to maintain or improve quality of life during the life course [1]'**.** Authors in [2] conducted systematic reviews on health literacy and public health, and found that a shared characteristic of these definitions focus on individual skills to obtain, process and understand health information and the services available to make appropriate health decisions. An individual with an adequate level of health literacy has the ability to take responsibility for one's own health as well as one's family health and community health [3].

Health literacy has attracted considerable attention across the globe in recent years. Research from around the world showed vast potential in demonstrating that optimizing health literacy results in improving health and well-being while reducing health inequities. On the other hand, weak health literacy competencies have been shown to result in less healthy choices, riskier behavior, poorer health, less self-management and more hospitalization [1]. According to a report by WHO Europe, limited health literacy cost more than US$8 billion, an estimated 3-5% of the total health care budget in Canada in 2009. An European Health Literacy Survey [4] conducted in 18 European countries has identified inadequate or problematic health literacy as a key determinant of health, a high-prevalence problem, a drain on human and financial resources and an obstacle to development; action in a range of settings and sectors can enhance health literacy; and building policy to support the strengthening of health literacy at the global, regional, national and local levels.

Acquiring relevant and accurate information helps patients to understand their diseases and conditions, and most importantly, to understand the benefits of changing to healthier behaviour. The so-called self-management of a disease outside the conventional clinical setting is an essential complement to clinical care for chronic diseases. For example, Diabetes Self-Management Education (DSME) has changed from a didactic approach focusing on providing information to more theoretically based empowerment models that focus on helping those with diabetes to make informed self-management decisions [5]. The informed knowledge helps patients to improve self-care behavior [6-7], such as improved clinical outcomes, i.e., lower A1C [8-9] in the case of diabetes; lower risk for people at every age for cardiovascular disease via a healthy diet and adequate physical activity [10].

Nowadays, with the popularity of online resources, there is large amount of information available to patients with regards to their diseases; however, given the quantity of information made accessible to patients, the credibility and quality of such data is difficult to judge. In addition, Internet of Things (IoT) technology provides ways to monitor and collect data on a real time basis by sensing and communication capacity. The collected data provides another dimension of health data, but also gives the opportunity to closely follow up a patient's reaction to the treatments and medicines. However, the issue of making the most out of the data in the self-management process poses as a challenge to ICT and medical professionals. Presenting data to patients in an easy-to-understand way is the first and most important step to widespread health literacy in assisting self-management.

In this paper, we will describe our work in providing patients with efficient ways to explore health literacy and combined with our observed data, this will assist them in understanding their own conditions which will subsequently lead to making better decisions regarding behaviour changes. Our work is part of the EU CARRE project, which focuses on cardiorenal syndrome and aims to provide innovative means for the management of comorbidities (multiple co-occurring medical conditions), especially for patients with chronic cardiac or renal disease or persons with increased risk of such conditions.

Our findings bring patients closer to the relevant medical information on the risks and comorbidity of cardiorenal diseases from the latest research and publications on PubMED[1]. In addition, patient gets individual risk by combining the general risk with individual patient's measurements of biomarkers, both from hospital tests and from wearable sensors that take measurements at home. Our work makes use of the data visualisation approach in order to use the best interactive graphs to present the data to the patients in a comprehensible manner, a crucial factor in ensuring the patients are able to read and interpret the graph. By doing so, we contribute to bridge the gap between the health literacy and the understanding of the general public in medical knowledge; and help users to make informed decision in their care.

The rest of the paper is organised as follows: section II provides a brief description on the design and functions of the data-centric system; section III and IV are used to describe the most relevant subsystems with the data visualisation functions, with section III explaining the knowledge extraction subsystem, while section IV describes the visual analytic sub-system. Finally, section V presents three use cases of using visual analytics in presenting the data to medical professional and patients, in order to achieve a greater understanding as a whole and encourage better decision making on the part of the patients.

II. Data Centric Self-Management System

The concept and function blocks of the CARRE data centric systems are illustrated in Figure 1. The system aims to seek ways to manage comorbidities (multiple co-occurring medical conditions), especially in the case of patients with chronic cardiac and renal disease or persons with increased risk of such conditions. The concept can be applied to other diseases as well. The system consists of three main blocks: data centre, Heterogeneous data resources and the backend support.

Data centre: linked data based data-centre is the core component of the system. It integrates all different types of data, adding internal semantic links among them as well as external semantic links to Linked Open Data knowledge. In addition, the data centre will support the query endpoints for semantically retrieving the data. The data is heterogeneous and covers: medical evidences of possible risk associations, which are extracted from publications on the trusted web sites; the Personal Health Record (PHR); data collected from monitoring devices, in particular wearable sensors which enable patients to monitor their conditions and life style in a consistent and largely non-intrusive way; decisions that are derived from the personalised services.

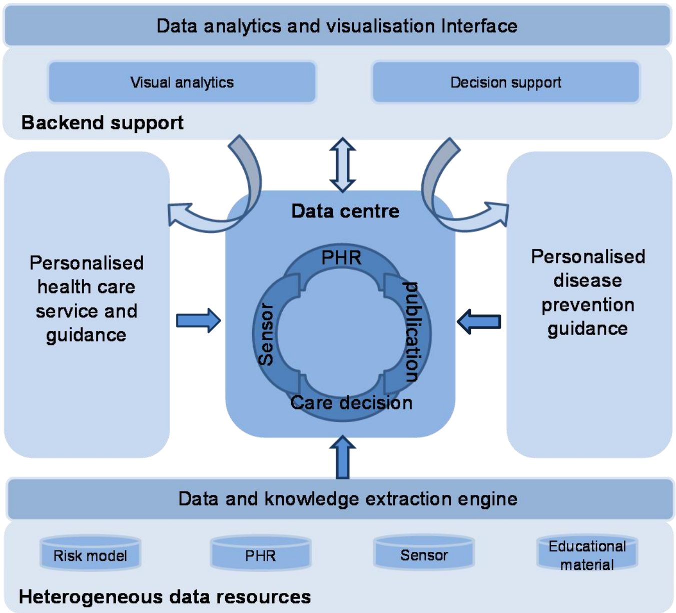

Figure 1 data centric system for disease and its comorbidity prediction

Heterogeneous data resources, this includes all different kinds of data resources related to personalised health care. These data should be the key factors in selecting the personalised care services and in defining the right treatment plan. The data may be accessible from different types of resources with structured data formats (e.g. Web API outputs or database tables) or unstructured data formats (Web pages) and by using heterogeneous presentation schema.

Data and Knowledge extraction engine aims to enable crawling data from heterogeneous data resources. This knowledge is the result of a widely published clinical study made available on PubMed, among other places. In particular, we extracted information about the risks related to causes of cardiorenal diseases and its possible comorbidities. We define this information as risk association, which indicates the potential risk of developing other conditions considering the patient's current condition and biomarker results. The risk association information will be extracted and stored with Linked Data principles into the Linked Data based repository (Data-centre). The extraction engine follows the defined scheme to lift or transform all different crawled data into a unified data space. Section III provides additional information about this part of the work.

The data that are collected includes:

- *Medical model* information collected from literacy available from Internet.
- *PHR* database that records a patient's basic health care environment, patient history record and the social relationships/activities of the patient.

[1] PubMED, US National Library of Medicine, National Institutes of Health http://www.ncbi.nlm.nih.gov/pubmed

- *Sensors* collect the necessary data while monitoring the patient outside the hospitals, such as body fat, BMI, steps and Calories burned.
- *Educational material* is the relevant medical literacy with regards to self-management. It aims to educate people to choose the right decision on behaviour changes.

In processing PHR and personal sensor data, security and privacy should be taken into serious consideration. However, it is out of the scope of this paper.

Backend support provides decision support service for patients with relevant data visualisation to help patients to explore and understand the data stored in the data centre. The machine-readable data from the data centre can be used by web applications, such as decision making and visual analytics. Together with the individual sensor and PHR data, useful knowledge/information related to the particular patient can be generated.

Data analytics and visualisation interface provides the interactive data analytics and mining functions. The analytics tools can help patients to view, in a user friendly way, their disease progressions with the collected data. It allows them to see the changes in measurement over time and how these changes are linked to their life styles. The data is first hand, and undeceived. The data can also help medical professionals to gain insight into the inter-links between the different biomarker, and life style monitoring. This helps medical experts to provide personalised care, treatment and medicines.

The services that the backend system supports are:

- ***Personalised health care services and guidance*** is a decision support service module that suggests treatment guidance, alerts and education materials that are suitable to a particular patient's needs based on their profile data and supported information from data analytics and mining. It can also combine the searched knowledge with sensor outputs to provide clinical information that is specifically tailored to the individual in order to track the progression and interactions of comorbid conditions. Visual analytics is employed so that patients and clinicians will be able to visualise, understand and interact with this linked knowledge as well as taking advantage of personalised empowerment services supported by a dedicated decision support system.
- ***Personalised disease prevention guidance*** is another decision support service module that will provide personalised information and life-style guidance to the patient in order to manage risks for comorbidities or progression of disease at more severe stages.

In this paper, we focus on explaining the two main subsystems: data and knowledge extraction subsystem and data analytics and visualisation subsystem.

III. HARVESTING RISK MODELS FOR CADIORENAL DISEASE AND ITS COMORBIDITIES

Risk models were often built as a result of clinical studies, and they are openly available to the public through government medical advice sites, and can be publicly accessed based on medical literature (generally indexed in PubMED).

Constructing risk models for cardiorenal syndrome is a task of harvesting risk factor data consisting of interlinked entities, such as risk elements and risk evidence that are related to ground knowledge in cardio-renal disease and comorbidities (symptoms, diseases, risk ratios, treatments, and medical evidence source) from medical state-of-the-art scientific literature databases.

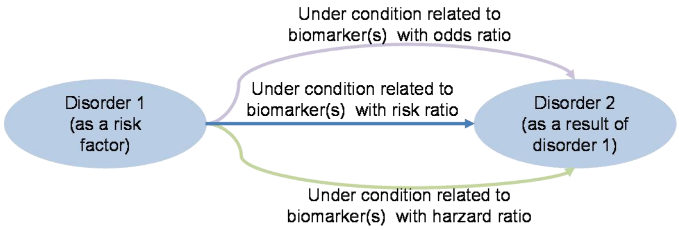

Figure 2 Schematic representation of the risk factor triplet

The analysis of cardiorenal syndrome types and related medical conditions involved the issue of co-existing diseases as well as the relevant characteristics of the individual, and the environment they are exposed to, that increases the likelihood of developing cardiorenal syndrome (risk factors). It is also concerned with the indicators of pathogenic processes (biomarkers) leading to the development and progression of cardiorenal syndrome according to current medical evidence based literature analysis. Figure 2 shows the description of the risk associations [11]. The risk association can be read as disorder 1 (i.e. diabetes) under certain conditions (such as gender), and relates to disorder 2 (i.e. coronary heart disease) with a risk ratio of 2.82 (confidence interval: 2.35-3.38) if a female patient has been diagnosed with diabetes; and with a risk ratio of 2.16 (confidence interval: 1.82-2.56) if a male patient has been diagnosed with diabetes.

The general approach for mining medical evidences includes the following:

- Mining new evidences for existing, known risk associations: for example, a risk factor for developing a certain comorbidity within a particular ethic group, as observed by new clinical studies;
- Mining new risk associations: for example, if a new biomarker or lifestyle factor has been found to be responsible for developing certain a disease or comorbidity.

The main reason for developing automatic (or semi-automatic) systems in searching for knowledge is that the knowledge itself is developing and cannot be treated as static. For example, the field of medical knowledge dealing with understanding the risks of developing certain diseases or symptoms is constantly being enriched by the latest research in the medical domain, and while some research may provide additional proof that further legitimizes existing practices, other research may result in correcting or redefining the existing knowledge base.

The RDF repository [12] forms the data repository and it stores general medical knowledge relating to risk associations, evidence and observables, and is available for public querying without authentication.

IV. DATA ANALYTIC AND VISULISATION SYSTEM

The visual analytics module, as shown in Figure 3, is a web-based subsystem that provides supporting functions for data exchange, visualization and analysis. 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 users. The web-based framework is composed of a dashboard and multiple visual analytics component containers.

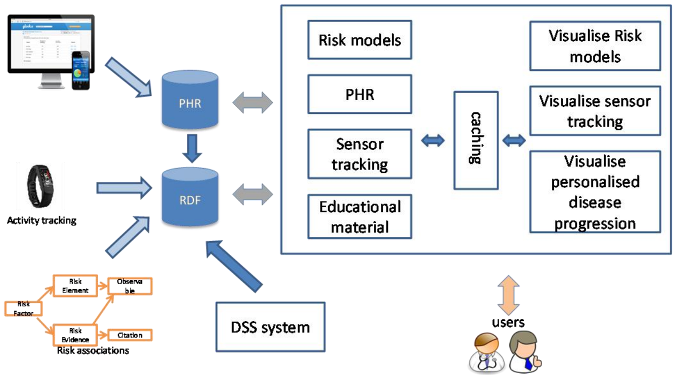

Figure 3 Visual Analytics Interface Architecture

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

- Visualising individual risks and allowing for analytical analysis of the impact of the behaviour to the risks. The functions include:
  - Visualisation of the risk models helps users to understand the general disease progression; it also helps professionals to explore the diseases progression and identify the core factor in disease progression.
  - Visualise individual tracking data using various graphic approaches with the aim to help users to understand the meaning of data; and identify the core factors that causes diseases progression.
  - Visualise individual risks and allowing for analytical analysis of the impact of behaviour changes to the risks. It helps patient to understand the relations between the outcomes and their behaviours during the self-management.
- Enhancing user experiences in behaviour monitoring, symptom reporting, observables monitoring. Referring to risk associations reported in [11], the data analytics and visualisation facilitates the lifestyle management by allowing for health status data visualisation (behaviours, symptoms). The functions include:
  - Monitoring: visualisation of a wide range of data including patients' activities, movement, step accounts, diet and other health-related behaviours and events, observables monitoring, such as blood pressure and blood glucose. The monitoring will make most use of sensors and mobile apps.
  - Personal Diary: visualisation of the health status of the individual and their behaviours, including their locations, movements, diet, sleep quality, environment, mood, blood pressure, glucose, alcohol, smoking, and other symptoms, etc. Visual analytics will be used to display individual/aggregated data items to allow easy interpretation of the data by the patients. With the search bar of the system, users can easily send queries about their activities, movements, diet, etc.
  - PHR Data: visualisation of a variety of health related measurements. Most of them are time series with different sampling intervals.

V. CASE ANALYSIS

In this section, we present interactive functions of data visualisation that helps patient to understand their disease progressions, and the effect of changing their behaviour towards their disease progression.

Patient Y is a 55 year old female (height is 160cm, weight is 94kg). She has been diagnosed with obesity (waist circumference is 125), diabetes (2 hour glucose after oral glucose tolerance test is 198 mg/dl, HbA1c is 7.8), with hypertension (160 mmHg /90 mmHg), dyslipidaemia (total cholesterol is 215 mg/dL). She is at Chronic Kidney Disease (CKD) stage 3 (moderately to severely decreased, with the measurement of 35), and Left Ventricular Hypertrophy (LVH).

Use Case 1: Visualise the measurement data use Healthlines

Most of data, such as PHR and sensor data are time dependant. A timeline is a traditional method used to visualise time-varying data and events in a linear layout and it is suitable for continuous variables because it covers a relatively long period.

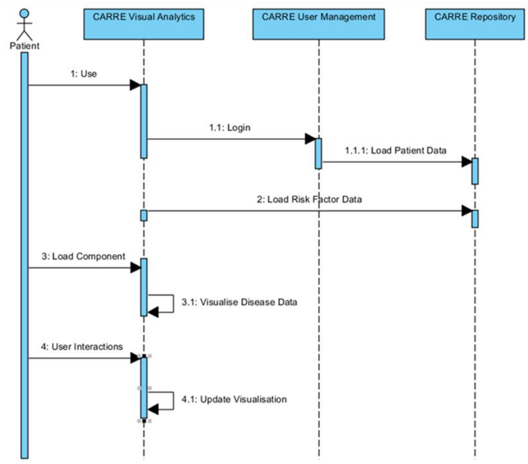

Figure 4 Sequence diagram of healthlines

The interactive healthlines provide patients with functions allowing them to explore, compare, and cross-check the health indicators and medical measurements. Interactive techniques such as zooming, overview and details are integrated with the visualisation. Line and bar charts are available for variable visualisation. A convenient drag-and-drop is used for variable selection. Figure 4 show the general usage of the healthline.

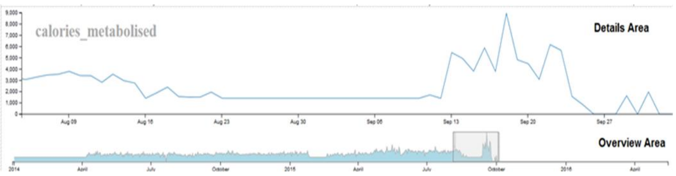

Figure 5 Screenshot of a single measurement displayed in Healthlines

Figure 5 is a screenshot of a healthline. The variables (e.g. the measurements from PHR or wearable sensors) can be picked up from a list and data trends can be observed from the variable curves. The X-axis denotes the time and can be displayed using various time scales (e.g. days or months), so users can analyse multiple data sets, such as calorie consumption, exercise (in number of steps) and blood pressure, using the same time scale and on the same screen. This makes both cross-checking and deriving possible links between the measurements easier. For example, patient can see the relationship between exercise and blood pressure. Literature shows[2] that if one exercises regularly, for at least 30 minutes on most days of the week, one can lower one's blood pressure by 4 to 9 (mm Hg). Patients can use the timeline to follow up the two measurements and see the correlation. This type of real-time visual data presentation should have a positive impact on a patient' behaviour and willingness to stick to their care plan.

Use Case 2: Visualise disease progressions using node-link diagram for risk analysis

This use case scenario is to help patient to visualise general risk associations and disease progressions, as well as their own risks and disease progressions if combined with their biomarker measurements.

Node-link diagrams are usually used to visualise a tree or network graph data structure. As described in section III, the individual risk associations are interconnected and likely to form a networked graph; therefore it is the intuitive and natural way to represent the risk associations we harvested from the literature. In our node-link visualisations of a network, entities are disorders and they are represented by nodes, the links or edges among those nodes represent the relationships between the nodes, i.e, one node transits to another node with a certain ratio.

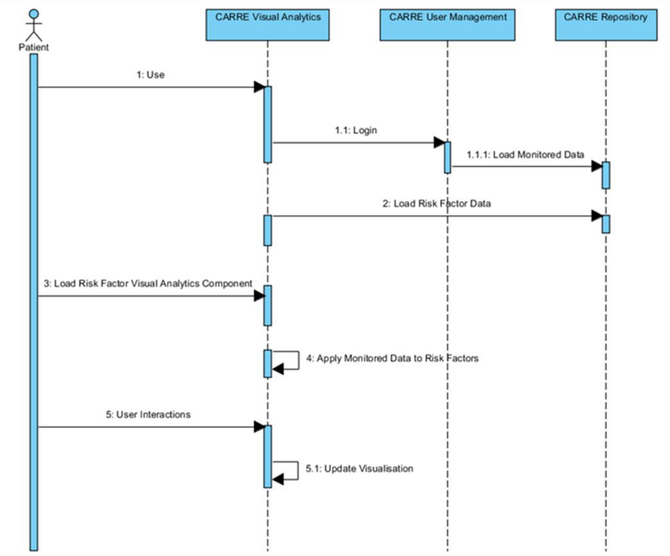

Figure 6 Sequence diagram of using node-link function

The basic graph layout is straightforward. However, with an increasing number of nodes and edges, 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 diagrams to reduce difficulties in visualisation, such as force-directed layout [13] and Multi-Dimensional Scaling (MDS) [14].

In the use case, we focuses on an individual's disease progression, so, the diagram is anticipated to be part of the overall cardiorenal disease related association diagram. Upon applying for their own profile, each user will see part of the risk association diagram, which only presents the risks related to that particular patient. Figure 6 shows the sequential diagram for the use case.

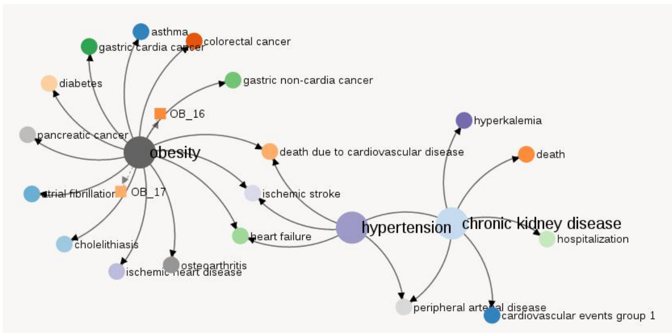

Figure 7 Example of a node-link diagram for patient Y

Figure 7 shows the risk associations most relevant to patient Y. For example, Chronic Kidney Disease carries a risk of developing hypertension. When one clicks the node, dotted lines will appear which show the diagnosis.

[2] http://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure/art-20046974

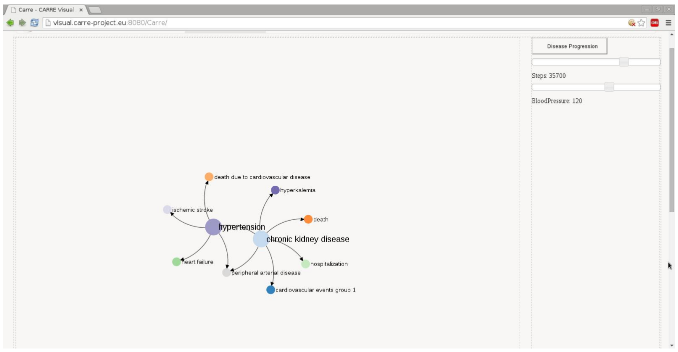

Figure 8 screenshot of the interactive node-link diagram

A useful interaction, as shown in Figure 8, is that patient Y can see changes in the measurement data, for example blood pressure, when the value is lower than the threshold of diagnosis (i.e. if a patient has no longer been diagnosed with hypertension), then the risk association diagram changes, as shown in Figure 8, where it can be seen that the risk diagram has much fewer nodes, meaning that there are fewer possible risks of the patient developing other diseases and comorbidities. Again, this function encourages patients to monitor and control their biomarkers. If combined with the example in the use case 1, then patient Y will see the positive impact of exercise on blood pressure as well as the benefit of having lower blood pressure on disease progressions. This will hopefully encourage the patient to make wiser decision and work towards changing their lifestyle.

Use case 3: Chord diagram for showing risk associations alternatively

Though filtering has been applied based on the conditions of a particular patient, it may not help for the visual analysis of the whole risk factor database. Fortunately, there are some network visualisation techniques to alleviate the problem such as the chord diagram, as shown in Figure 9.

One of the benefits of the chord diagram is that all the nodes are arranged in 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 to one node much easier.

The chord diagram clearly visualises the relationships of all risk elements in the data repository and it is particular useful in showing the main causes of the disorder. This use case also demonstrates that alternative interactive functions can help in understanding the relationships between disorders. In fact, this diagram is especially suitable for medical professionals as it aids the understanding of the relations between disorders and it clearly presents the main problem faced by the patient, all of which makes choosing the best treatment and medical care much easier.

Figure 9 Mousehovering on a certain risk elements showing all risk factors starting or ending at the specific risk element

VI. CONCLUSION AND FUTURE WORK

In this paper, we used data visualisation techniques to present medical evidence on cardiorenal disease and its comorbidities as well as biomarker data collected regularly via PHR and wearable sensors. In addition, we provided functions that present the combination of the two sets of data to make more sense about one's health. The functions target the general public, which has limited skills in understating statistical analysis. The case studies showcased three possible usages of the visualisation functions. Our work aims to bridge the gap between the health literacy and the understanding of the general public in medical knowledge; and help users to make informed decision in their care.

Future work will focus on improving the functions, such as by interlinking the individual biomarker data with the overall general disease progressions. In addition, we will conduct a wider range of user tests and work on improving the functions based on the feedback.

ACKNOWLEDGMENT

The research is sponsored by EC CARRE project under agreement of FP7-ICT-611140.

REFERENCES

[1] http://www.euro.who.int/__data/assets/pdf_file/0008/190655/e96854.pdf

[2] K Sørensen, S Van den Broucke, J Fullam, GDoyle, J Pelikan, Z Slonska, H Brand and (HLS-EU) Consortium Health Literacy Project European, 'Health literacy and public health: A systematic review and integration of definitions and models', BMC Public Health, BMC series, 2012

[3] Edited by: D McQueen, L KI Potvin, JM Pelikan, L Balbo, Th Abel, 'The Role of Theory in Health Promotion', Health and modernity. 2007, Springer

[4] K Sørensen, J M. Pelikan, F Röthlin, K Ganahl, Z Slonska, G Doyle, J Fullam, B Kondilis, D Agrafiotis, E Uiters, M Falcon, M Mensing, K Tchamov, S van den Broucke, and H Brand, 'Health literacy in Europe:

comparative results of the European health literacy survey (HLS-EU)', Eur J Public Health. 2015 Dec; 25(6): 1053–1058.

[5] DG Marrero, J Ard, AM Delamater, et al. Twenty-first century behavioral medicine: a context for empowering clinicians and patients with diabetes: a consensus report. Diabetes Care 2013;36:463–470 209.

[6] L Haas, M Maryniuk, J Beck, et al.; 2012 Standards Revision Task Force. National standards for diabetes self-management education and support. Diabetes Care 2014; 37(Suppl. 1):S144–S153

[7] SL Norris, MM Engelgau, KM Narayan. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. Diabetes Care 2001;24: 561–587 208.

[8] SL Norris, J Lau, SJ Smith, CH Schmid, MM Engelgau. Self-management education for adults with type 2 diabetes: a meta-analysis of the effect on glycemic control. Diabetes Care 2002;25:1159– 1171

[9] JM Barker, SH Goehrig, K Barriga, et al.; DAISY study. Clinical characteristics of children diagnosed with type 1 diabetes through intensive screening and follow-up. Diabetes Care 2004;27:1399– 1404

[10] American Heart Association, http://www.heart.org/HEARTORG/GettingHealthy/Preventing-Heart-Disease---At-Any-Age_UCM_442925_Article.jspIyngkaran P, et al. Cardio-renal syndrome: New Perspective in Diagnostics. SeminNephrol 2012; 32 :3-17

[11] E Kaldoudi, A Third, G Gotsis, S Roumeliotis., N Karvelyte, L Rimseviciu, K Pafili., D Papazoglou, G Juozalenaite., E Semertzidou, Z Visockiene, K Zigeridou, Functional Requirements & CARRE information model, Deliverable D.2.2 in the Project "CARRE: Personalized Patient Empowerment and Shared Decision Support for Cardiorenal Disease and Comorbidities", FP7-ICT-61140, July 2014. https://www.carre-project.eu/project-info/deliverables/

[12] A Third, E Kaldoudi , G Gkotsis, S Roumeliotis, K Pafili, J Domingue, Capturing Scientific Knowledge on Medical Risk Factors, SciKnow 2015: 1st International Workshop on Capturing Scientific Knowledge, collocated with the 8th International Conference on Knowledge Capture (K-CAP), Palisades, NY, USA, October 7-10, 2015

[13] P Eades, 'A heuristic for graph drawing. Congressus Numerantium', 42:149–160, 1984

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