From Health-Persona to Societal Health

Ramesh Jain
University of California, Irvine
jain@ics.uci.edu

Laleh Jalali
University of California, Irvine
lalehj@ics.uci.edu

Mingming Fan
University of California, Irvine
mingminf@uci.edu

ABSTRACT
In this position paper, we propose an approach for Web Observatories that builds on using social media, personal data, and sensors to build Persona for an individual, but also use this data and the concept of Focused Micro Blogs (FMB) for situation detection, helping individual using situation action rules, and finally gaining insights for obtaining insights about society. We demonstrate this in a concrete use case of fitness and health care related sensors for building health persona and using this for understanding societal health issues.

Categories and Subject Descriptors
H.1.2 [**Models And Principles**]: User/Machine Systems – *human factors, human information processing*; H.3.5 [**Information Storage And Retrieval**]:Online Information Services – *Web-based services*; H.3.4 [**Information Storage And Retrieval**]:Systems And Software – *User profiles and alert services*;

General Terms
Management; Design; Human Factors

Keywords
Persona; EventShop; Health and Fitness; Focused Micro Blogs; Situation Detection; Event Streams

1. INTRODUCTION
We are living in transformative times. Our society is going through a major transformation both at the individual level as well as at the global level. The most important factor in this transformation, the fuel powering this new engine, is the data; particularly the objective data collected using sensors. However, data from only a specific sensor contains only partial truth about its environment. Each sensor is designed to capture a particular attribute and most objects and events have several attributes that may change with time. The fidelity of an object or an event improves with more diverse sensors to capture their attributes. These attributes are captured by sensors recording the attributes as a time-varying signal at specific locations. Billions of sensors are now continuously measuring different physical attributes at most locations of interest in the world. Increasingly, a good fraction of these spatial-temporal measurements are being made available on the Web, giving us opportunity to create Web Observatories.

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

Each data stream observed by a sensor has valuable information. When two data streams are correlated it results in valuable insights. As more data streams are correlated, better insights are created. The fact that one can gain valuable insights by correlating multiple data streams collected for the same event or same person is well recognized, but tools and techniques for doing this easily have not been developed as well. Currently, we are collecting increasing amount of such data leading to what is commonly called Big Data. Data from different sensors is usually at different spatial resolution as well as at different frequencies. Moreover, the type of data is multi modal, making the semantics of the data very different. This poses serious problems in correlating these data streams. We need efficient tools to unify and correlate all such data streams to gain valuable insights.

Before being able to perform any computational task on heterogeneous data streams, we have to unify the collected data. In this paper we propose **events as unification mechanism.** We consider a case of such data being collected related to a person to create the *persona* that may be used in different contexts. *Persona* is the mask or appearance of a person presents to the world [2][7]. The persona models different relevant aspects of a person for interactions or for taking specific actions involving person's characteristics, professional, social, health, preferences and interests (Figure 1). In this paper, we will consider health persona of using different data streams for a specific person. The data streams that we use are related to many sensors now being used by people to collect fitness data, personal events, eating habits, sleep patterns, and every day activities using mobile phones, calendar, and social networks. Our goal is to identify challenges faced in correlating data from all these diverse data streams and develop a computational platform that allows us to plug-in other data streams to include them in refining the persona.

Such data collected for several individuals could then be aggregated to understand issues related to society. In most cases, such individual data is a rich source to do research in many aspects of individuals as well as society. We will discuss this aspect and propose an architecture to address collection and analysis of the data.

We use EventShop [3] as the platform to aggregate different data streams relevant to specifics of persona and build this into situations that help individuals and identify global situations. These global situations are then used to advise individuals in their personal decisions.

2. HEALTH-PERSONA
Until recently and in most cases even now, all information about a person is very subjective. Consider all information related to fitness and health about an individual. We all had anecdotal

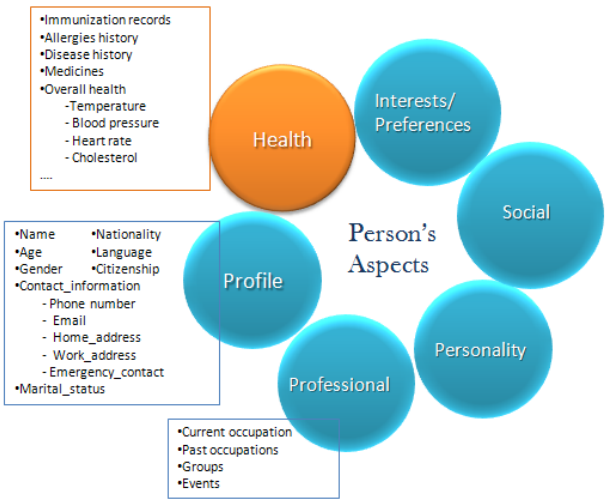

Figure 1 Different aspects of a person

information about how much exercise we do, how many hours we sleep, what we eat, and how our body is functioning. The recent information technology revolution is changing that. Using a myriad fitness and health related wearable sensors, combined with mobile phones, every minute of an individual's activities are being recorded that may be analyzed. We have seen a revolution that allows capturing information about 'self' from many sources. What we are seeing is a major shift in how we got information about individuals. In this area, we can characterize changing trends into specific stages of different types of information:

2.1 Subjective Self

All information about a person was totally subjective. This information could be collected from several sources, but it was mostly subjective interpretation by a person, including self, about the state of health.

2.2 Quantified Self

People started collecting information using different sensors. Usually sensors don't lie, they give fairly objective information. Also, many sensors could be collecting data 24/7 and storing this on a server where this data could be analyzed. This data is collected using a sensor for a specific aspect of fitness or health. Such data could be visualized and some analytics could be done to help a person.

2.3 Objective Self

Using multiple wearable sensors, people collect 24/7, many different attributes related to fitness and health. Mobile phones and other sensors also collect information related to travel and location of activities that could be converted into different events in their personal life. Information from personal calendar, social networks and other similar sources is also used. All this information can be correlated to analyze lifestyle, relate it to personal health, and help people in fitness and health.

We address more detailed aspects of observing and analyzing all data sources to get objective self health information and identify needs to build a powerful platform to accomplish this. A very thought provoking book [12] paints a very data oriented future of medicine. Future of healthcare is likely to utilize an individual's lifestyle information with personal genomics to provide individual-focused health services.

3. BUILDING HEALTH PERSONA

First step in building any user centered application is data acquisition on aspects of person's daily life. Projects such as MyLifeBits [4][5] and many quantified self apps went in that direction. But they never go beyond collecting and storing data. Many lifelogging systems lack an explicit description of potential value for users, focusing instead on technical challenges (such as data capture and retrieval mechanisms) [15].we simply lack effective techniques for Analyzing information that involves examining it in ways that reveal the relationships, patterns, trends, etc. that would be valuable for enhancing user's life.

Our goal is building the health persona, by emphasizing on analysis and correlation of different data streams as different types of events occurring in person's life. In the following, we consider correlation among 4 different types of events: life events, food events, fitness events, and body parameters events. For creating these 4 important event streams, multiple sensors may contribute to each of these streams. As shown in figure 2, life events can be derived from different sources of information such as personal calendar, Facebook, LinkedIn and Foursquare accounts, smart phone apps and GPS. Fitness event stream may combine data from NIKE Fuel, FitBit, Basis, programs like Moves on smart phones and other similar sources. These sensors can detect events that are not perceived by humans, such as alteration of oxygen levels in the blood or the amount of carbon dioxide or dangerous pollens in the air. Once these data streams are collected, they can be combined and correlated to identify patterns: for instance, they might determine which environmental conditions worsen a child's allergy or which geographical locations should be avoided by senior citizens at any particular time since they are more vulnerable to flu and other contagious diseases. This information would allow doctors to spot irregularities early, providing warnings before an illness becomes serious. Health Persona is a detailed ongoing health record which physician would have access to, and one would no longer have to think very hard to answer questions such as "How much exercise you had last month" or "When did you first start feeling this way?"

Figure 3 shows the creation of 4 different types of events, mentioned in figure 2, in more details. We illustrate these events in different colors in each of these four streams. The original signals for the stream are continuous signals as captured from a sensor. Different types of events may be determined using a model to classify them and determine their time intervals. For example fitness events may be, no activity to vigorous activity. And each event may be determined by considering some activity per minute as determined by a specific sensor. Moreover, any of these event streams may be the result of the combination of multiple sensor streams. For example, activity level may be determined by considering Nike Fuel, Fitbit, BASIS [1], and Jawbone. Each of these measures activities differently, but these measurements are correlated to actual activity levels and are aggregated to correspond to and segment the timeline according to physical activity by a person.

With these data streams the health persona is determined by combining them over a long period to gain insights about

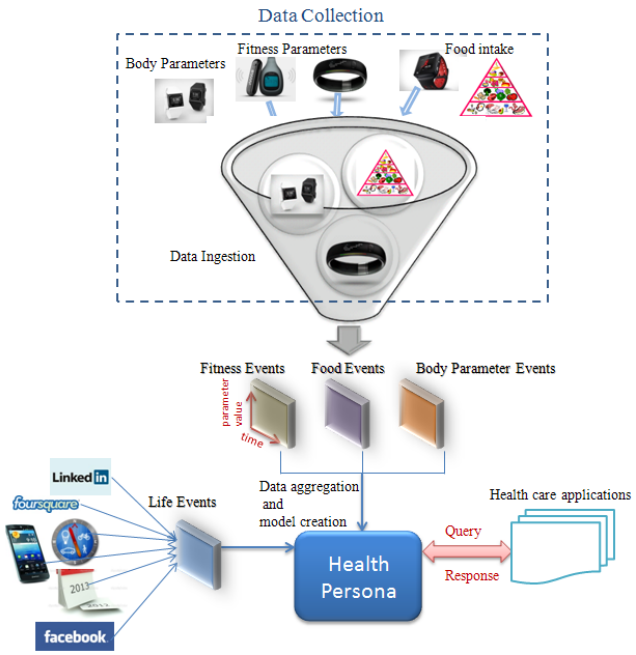

Figure 2 Building Health Persona from heterogenous sources

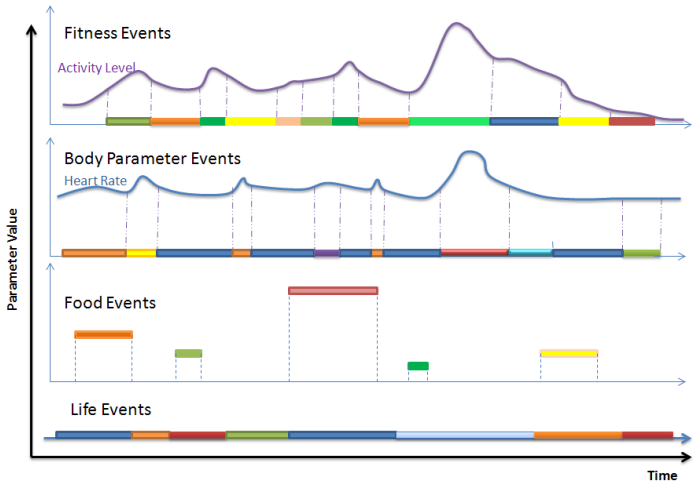

Figure 3 Heterogeneous Events streams

a person by correlating these event streams and drawing insights about a person's health. One example of such insights could be:

Calendar + GPS + Motion Sensors + Health Monitor Data

A calendar contains planned events, some of which may not happen. With assistance of GPS and motion sensors, we can make better inference of events that really happened. For instance, calendar indicates that Mr. A has a weekly group lunch meeting at restaurant B on Tuesdays. GPS data shows whenever Mr. A attends these meetings, and during that the motion sensors measure A's physical activity level, and health monitoring device BASIS records his heart rate and body temperature. On analysis of the data over last several meetings it is observed that though during these meetings Mr. A is relatively inactive, his heart rate is high and his body temperature is above normal. This correlation from different event streams suggests that there is something unusual about the meeting, or the meeting place, or the food at the restaurant.

3.1 Benefits of Health Persona

The amount of money spent for health care in United States is extraordinary. The United States spends more on health care than Japan, Germany, France, China, the U.K., Italy, Canada, Brazil, Spain and Australia combined [16]. Yet the outcome is not extraordinary. As mentioned in section 2.3 with the emergence of wearable sensors that can measure most fitness and body parameters, new data is available to us. Knowledge derived from this data can be used in lowering the cost of health care systems. Today most of medicine is not evidence-based; but based on randomized clinical trials on a limited population. So decision making in today's health care system is based on historical preferences of the physician as opposed to be data driven for each individual [21]. If physicians could have access to the history of all health related aspects of a person measured through time, the treatment process could be personalized and some costly tests and lab examinations could be waived. Here historical data is not only electronic medical records of the patient but also includes all body and fitness parameters, all the events happening in person's life and all food intakes. Health persona provides a computational framework that makes this possible.

Due to population growth, aging, and other factors; it is projected that demand for physicians will outpace supply by 2025. Simply educating and training more physicians will not be enough to address these shortages. Figure 4 shows the projected supply and demand [17]. So we need the tools, technologies and computational frameworks being developed in this space to be an *extension of a physician*. Health persona as a personal health and lifestyle platform helps individuals maintain a personal health and lifestyle record for their lifetime and obtain personalized, lifestyle-related recommendations to improve their health status. This has some life changing impacts that with only 3.9 average number of visits to medical services (physician, hospitals) [18] can never be achieved.

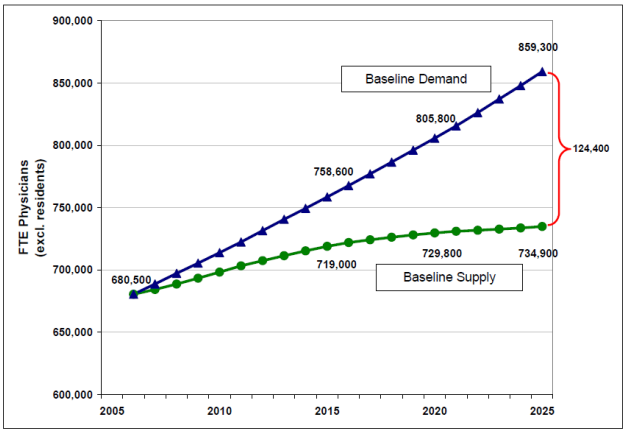

Figure 4 Baseline physician FTE supply and demand projection (2006-2025)

3.2 Ethical and Privacy Issues

Today we are living in a data age. Millions of pieces of information are generated and used by organizations every day and developments in technology have helped to make the storage and manipulation of this information much easier. Personal data will continue to increase dramatically in both quantity and diversity, and has the potential to unlock significant economic and societal value for end users.

Concerns for privacy are not new. The initial legal opinion on the right to privacy dates back to the late nineteenth century. The advent of a new technology that could permanently capture a person's behaviors and actions was the impetus for this opinion since 1890 [19]. Fundamental questions about privacy, property, global governance, human rights – essentially around who should benefit from the products and services built upon personal data – are major issues emerging due to technological advances in the last few decades. Building the legal, cultural, technological and economic infrastructure to enable the development of a balanced personal data ecosystem is vitally important for improving the state of the world and is discussed in details in [20].

4. GOING BEYOND TWITTER

Twitter data is so broad in scope that any topics can be found in tweets, but unfortunately all topics are mixed together. If we treat the relevant information as signal, and all others as noise, twitter data has very low signal-noise-ratio (SNR). Use of tools like hash-tags help, but do not really solve the problem. Given this about current Twitter like social network, we believe that they all suffer serious issue of low SNR and are less effective as information sources. Therefore, we propose the concept of Focused Micro Blogs (FMB). FMBs are on a focused topic and are targeted to a specific source. Since the topic is fixed, the information in a FMB could be easily structured and parsed. FMBs retain the important feature of sending brief real time information of Tweets, while overcoming their weakness of combining too many different topics in one place. This information is aggregated at a server designed for a specific task.

Let's consider one concrete example of FMB in a popular application for traffic – the Waze [14]. Waze is an application on phones that takes GPS data periodically from the phone and uses this to compute the speed of the device. This is aggregated at the server to know traffic conditions in different areas. A user may also do a micro-blog using specific input mechanisms so that the system knows about a stalled car or a police officer at a specific location. By using sensor data from the phone and specific micro-blog-inputs, a Waze server provides very useful information to its users. This is an FMB in action.

Let's consider a Flu application. With sensors on mobile phone, information like activity level and location is easily available. Physiological data such as heart rate and body temperature can be retrieved from wearable device, like BASIS. User's overall feeling and some flu symptoms can also be collected from direct interaction between user and the application. The architecture of a system to collect different data streams and then combining these is shown in figure 5. In this architecture, Health Persona works as a server, collects and correlates health data from subscribed users. Different fitness and health related applications – like Flu or Allergy - running on user's smart phone provide Health Persona server with accurate data. This data will be used to incrementally build and update health persona. The longer user subscribed to the system, the more information gets collected.

Data from users gets collected in web servers corresponding to different apps. Each web server has a theme called T_i and a set of **S**ituation-**A**ction **R**ules (SAR). We will explain SARs in section 5. Different applications act as different FMBs feeding data into specific servers. Applications running on the phone also feed this data to Health Persona Server which stores all health related events, as discussed above, for every user. Health Persona Server not only helps an individual, but also collects community information that may be used to deduce societal health issues.

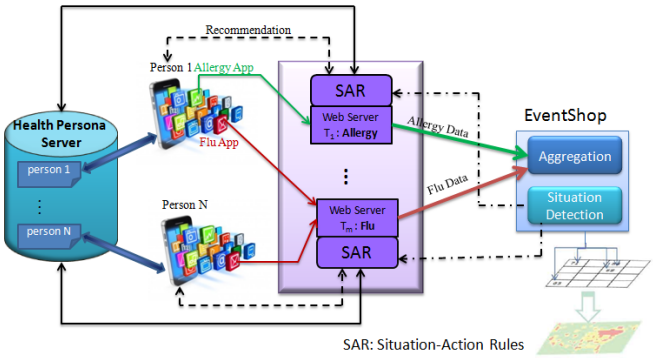

Figure 5 Phone as a source of FMB for different applications to assist its owner in different situations.

5. THINKING AT SOCIETAL LEVEL, ACTING AT PERSONAL LEVEL

It has already been demonstrated that insights from individuals can be aggregated for the society to gain societal insights. For example Google Flu Trends [6] is a web service provided by Google uses aggregated Google search data to estimate flu activity. The idea is that certain search terms are good indicators of flu activity. But can search query trends provide the basis for an accurate, reliable model of real world phenomena? Scientific hindsight shows that Google Flu Trends far overstated this year's flu season [10]. A pair of new projects [8][11] track Twitter posts to determine flu rates. However, Twitter crowdsourcing is less effective than search-term efforts, because signal to noise ratio is very low and Twitter's user-base tends to be younger and less reliable users.

Another example is traffic apps using user-generated traffic data to draw conclusions about traffic flow. Waze, mentioned above, is the world's fastest-growing community-based traffic and navigation app. It allows user to not only get personal heads-up from other users on the road but also contribute to traffic reports. This crowdsourced platform is more reliable than apps dependent on Twitter posts such as TwiTraffic[13] that looks for traffic related words in tweets.

Neither Google Flu Trends nor Twitter-based applications were successful enough in aggregating individual data to derive insight about a community. We need "Waze-like" platforms to collect and correlate accurate individual data and more importantly, aggregate these data for gaining community insights. The system

in Figure 4 uses specific applications on a phone to supply information to a situation recognition system and uses the detected situation to help the owner of the phone. It also creates health persona for the individual while creating health personas for several people that could be aggregated and analyzed for inferring societal health issues. As a result, we have two kinds of data aggregation. One that happens in each user's Health Persona and the other one that happens in EventShop.

EventShop [3] is a situation detection framework that makes data aggregation possible. It's a generic approach to select, import, and combine real-time data streams and operate on them to detect real world situations for generating appropriate information and actions. EventShop supports data from different sources. Any sensor data can be associated to a stream based on its location and frequency of creation. Aggregated data can also be visualized on a map. As a part of EventShop, Situation Detection unit detects real world situations (e.g. flu activity in a region) and based on that provides SAR units with appropriate information. Based on that, SARs send personalized recommendations to end users.

This approach doesn't suffer from high noise to signal ratio in flu-related tweets or flu-related search term indicators. More importantly, Users both contribute to the system and gets benefit from it. Precise community health insight naturally follows precise individual health insight.

6. CONCLUSION

Lots of personal data is being collected from various sources. By collecting and analyzing this data we can build effective personas that may in turn be used to help individuals. Moreover, by aggregating this data, we can gain insights at the societal level. We are working towards this. Also, we introduce the notion of FMBs that may be very effective in collecting and analyzing meaningful data in Web Observatories.

7. REFERENCES

[1] BASIS. http://www.mybasis.com/tour/

[2] Erving, G. 1959. The Presentation of Self in Everyday Life, *Anchor Books*, (New York, 1959).

[3] Gao, M., Singh, V., Jain, R. 2012. Eventshop: From Heterogeneous Web Streams to Personalized Situation Detection And Control, *ACM WebScience* (WebSci-12), 105-108

[4] Gemmell, J., Bell, G., Lueder, R. 2006. MyLifeBits: A Personal Database for Everything, *Communications of the ACM*, Vol. 49, (Jan. 2006), 88-95.

[5] Gemmell, J., Williams, L., Wood, K., Bell, G., Lueder, R. 2004. Passive Capture and Ensuing Issues for a Personal Lifetime Store, *Proceedings of 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences* (CARPE '04), Oct. 15, 2004, New York, NY, USA, 48-55.

[6] Google Flu. http://www.google.org/flutrends/us/

[7] Jung, C. G. 1953. Two Essays on Analytical Psychology, *Jung's Collected Work #7*, (London 1953).

[8] MappyHealth. http://mappyhealth.com/

[9] Pollen Density. http://www.pollen.com/allergy-weather-forecast.asp

[10] Resnick, B. 2013. Why Google Flu Trends Will Not Replace the CDC Anytime Soon, *National Journal*, (Jan. 25, 2013).

[11] Sickweather. http://www.sickweather.com/

[12] Topol, E.2012. Creative Destruction of Medicine, *Basic Books, 1st edition*, (New York, 2012).

[13] TwiTraffic. http://twitraffic.co.uk/

[14] Waze. http://www.waze.com/

[15] Sellen, A., Whittaker, S.2010. Beyond Total Capture: A Constructive Critique of Lifelogging, Communications *of the ACM*, Vol. 53, No. 5, (May 2010), 70-77.

[16] Steven Brill. 2013. Bitter Pill: Why Medical Bills Are Killing Us, *Special report TIME magazine*, (Mar. 04, 2013).

[17] Dill, M., Salsberg, D. 2008. The Complexities of *Physician*. *Supply and Demand*: *Projections* Through 2025, *Center for Workforce Studies*, (Nov.2008).

[18] Janice Lloyd. 2012. Study: Doctor Visits Decline, Uninsured Less Healthy, *USA Today*, (Oct. 2012).

http://www.usatoday.com/story/news/nation/2012/10/01/doctor-visits-decline/1605987/

[19] Warren, S.D., Brandeis, L.D. 1890. The Rights to Privacy, *Harvard law Review*, Vol. 4 No. 5, (Dec. 1890), 193-220.

[20] Schwab, K., Marcus, A., Oyola, J.O., Hoffman, W. 2011. Personal Data: The Emergence of New Asset Class, *An Initiative of the World Economic Forum*, (In Collaboration with Bain & Company, Inc, Jan. 2011).

[21] William Burnett. 2009. Evidence-Based Medicine: Is American Medical Care Based on Science or Politics?, *The Student Doctor Network*, (Sep. 2009).

http://studentdoctor.net/2009/09/evidence-based-medicine-is-american-medical-care-based-on-science-or-politics/