CrossMark

TECHNICAL CONTRIBUTION

Sensemaking in Intelligent Health Data Analytics

Magnus Boman · Pedro Sanches

Abstract A systemic model for making sense of health data is presented, in which networked foresight complements intelligent data analytics. Data here serves the goal of a future systems medicine approach by explaining the past and the current, while foresight can serve by explaining the future. Anecdotal evidence from a case study is presented, in which the complex decisions faced by the traditional stakeholder of results—the policymaker—are replaced by the often mundane problems faced by an individual trying to make sense of sensor input and output when self-tracking wellness. The conclusion is that the employment of our systemic model for successful sensemaking integrates not only data with networked foresight, but also unpacks such problems and the user practices associated with their solutions.

Keywords Artificial intelligence · Massive data · Health data · Intelligent data analytics · Syndromic surveillance · Sensemaking

1 Introduction

We argue that if the reality of clinical work is to successfully meet with new technology and methods for prevention and care, even the most intelligent analytics one could conceive of will not suffice. The interplay with a community of practice requires a process of sensemaking:

M. Boman (✉)
SICS, Box 1263, 16429 Kista, Sweden
e-mail: mab@sics.se

M. Boman · P. Sanches
KTH/ICT/SCS, Electrum 229, 16429 Kista, Sweden

finding out how to best inform a stakeholder based on data and the conclusions drawn from it, in a particular context. Involving the individuals that the data is about is crucial ("no decision about me without me"), as is clarifying the transience and dynamicity of data, and its sensitivity from privacy perspectives.

Medical health data comes in a limited number of forms (see Table 1). Non-medical health data resulting, e.g., from fitness sensor data or from syndromic surveillance are sometimes put to use in medical contexts [26]. Below, we make informed guesses about the future of intelligent analytics applied to such data. We build on current results from extensive networked foresight in the interval of 6 months to 3 years into the future, extrapolating to speculations about developments stretching to the end of this century. Artificial intelligence methods will be our main focus, and the methods useful for health data analytics in particular. The uncertainties involved are immense and, as a result, we anchor our extrapolations in observations of theory as well as practice, analysing challenges as well as opportunities.

After explaining our methodology, we present some trends and observations relevant to active healthy ageing, used as a backdrop for speculations on the future of personalised medicine, and the sensemaking that must go with it (Sect. 3). We then turn to a case study (Sect. 4), and before our conclusions we explain how to link the case study to our systemic model (Sect. 5).

2 Methodology

2.1 Data and Foresighting

The European Institute of Innovation and Technology's knowledge and innovation community on information and

Table 1 Exemplary forms of medical health data

Image data
  - Medical imaging (X-ray, MRI, PET, ECG, EEG,...)
  - Pathology microscopy
Sequence data
  - Transcriptomics (RNA)
  - Epigenomics (DNA)
Text data
  - Electronic health records
  - Scientific publications

Table 2 General trends, identified through Innovation Radar networked foresight

Ageing population
Dramatic increase in average life expectancy (co-morbidities)
Severe shortage of (informal) care givers
Rising healthcare costs for frailty
Lifestyle changes
More single households
More intense communication
Increased consumer/patient awareness
Digitisation
Massive data employment
Increased expectations on genotyping results

communications technology (EIT ICT Labs) [17] has a strategic unit that incorporates an *Innovation Radar* [10]. Across the thematic areas of EIT ICT Labs, the Innovation Radar locates ideas and concepts with high innovation potential.

In the last year, the Innovation Radar has produced three reports on ICT for active healthy ageing on mental [14], physical [8], and social [33] wellbeing, respectively. In addition, a recent White Paper [3] collects the lessons learned and provides strategic intelligence to decision makers in the partners of the organisation [16].

Besides trends with a bearing on active healthy ageing (see Table 2), the Innovation Radar team scans and investigates new commercial products in its partner countries, covers international trend perspectives, conducts patent studies, presents the *status quo* of research and development, and critically assesses new academic and science reporting.

The Innovation Radar approach is participatory, resting on the contributions of the thematic experts of the networked partner organisation. For the detailed analysis of the trends reported on here, about 20 experts plus an international network of scouts have been employed, respecting best practices of the participating organisations with respect to corporate and strategic foresight (cf. [28]).

The output examples below have resulted from networked foresight, providing value both to the partners and to the network as a whole (cf. [9]), by discovering and evaluating new technologies and opportunities. Foresight instruments that require a broad data basis have the potential to greatly benefit from a networked approach, especially in network organisations with a heterogeneous partner structure [35].

2.2 Anecdotal Evidence

The setting for the case study below is the online forum of the Fitbit community [12]. We perform a content analysis of forum posts, focusing only on publicly visible posts posted within the period of February and March of 2014. In content analysis, a categorisation scheme is often used in order to construct exhaustive, sometimes overlapping, categories. Since our aim was very specific, viz. to look at particular aspects of use, mainly relating to the work people do to maintain self-surveillance through these devices, we used inductive reasoning, looking for particular categories through the data but dismissing a high number of posts and threads relating to other subjects. This was appropriate because we found a majority of threads being irrelevant to most research questions, like requests for trading device straps (to get a different colour). Instead of making categories for these posts, we simply excluded them, leaving us with 29 % of the posts adhering to our initial categorisation, as determined by our initial discourse exploration combined with an initial examination of the 656 threads that we opted to categorise.

After two researchers had categorised 50 threads and agreed on the categories, one researcher categorised the rest with sample testing by the second researcher. In an analysis conducted after all threads had been categorised, 8 % contained aspects of adaptation, 10 % contained aspects of malfunctions or repair, and 6 % related to the integration of different systems. Finally, 7 % related to the user's mental model clashing with the system's actual model of calculations. A few posts contained topics of more than one category. Our data is biased since individuals who report on needing help in order to make sense of sensor data are presumable more visible than individuals who do not. Acknowledging that bias, we do not expect these results to be generalisable, but rather to open up for further lines of enquiry regarding the relationship between individuals and personal tracking in everyday life.

3 Sensemaking in Active Healthy Ageing

Eleven ICT-based trends for active healthy ageing were identified towards the end of 2013 (see Fig. 1). These trends were then analysed to determine their importance by

looking at parameters like estimated time-to-market and likelihood of turning into high impact future products and services. These trends were then linked to (in this case, ten) challenges that need to be met, and to (eight) opportunities for commercial and/or societal exploitation.

Limiting ourselves here to the trends tied to artificial intelligence and analytics, we focus on the first two (T1 and T2 in Fig. 1).

The first two trends display hope regarding what can be done to combat disease and threats to wellbeing when costs have come down enough for completely personalised medicine to be realised. This means completely integrating, at the molecular level, what goes on not only in the body but also in its immediate vicinity. To identify robust standardised molecular indicators of determining the

balance between health and disease as a function of inter-ventions is already possible, thanks to systems medicine: the application of systems biology to disease [20]. This makes it possible to define, at least in theory, at-risk parameters for wellness; letting a person know, reliably and in real-time, precisely which consequences result from which behaviours [25]. Challenges include secure and inexpensive data management at the micro level, harmon-ising not only with laws and clinical guidelines, but also with social norms.

At the macro level, health data will in this relatively far away future be used to estimate probabilities for develop-ing additional disease in a given geographic region at a given time, and given certain external factors. Two kinds of heatmap-like intelligence will be utilised. First, the global

Fig. 1 Trends in active healthy ageing, as identified through networked foresight, with links to significant challenges and opportunities relevant to artificial intelligence and intelligent data analytics
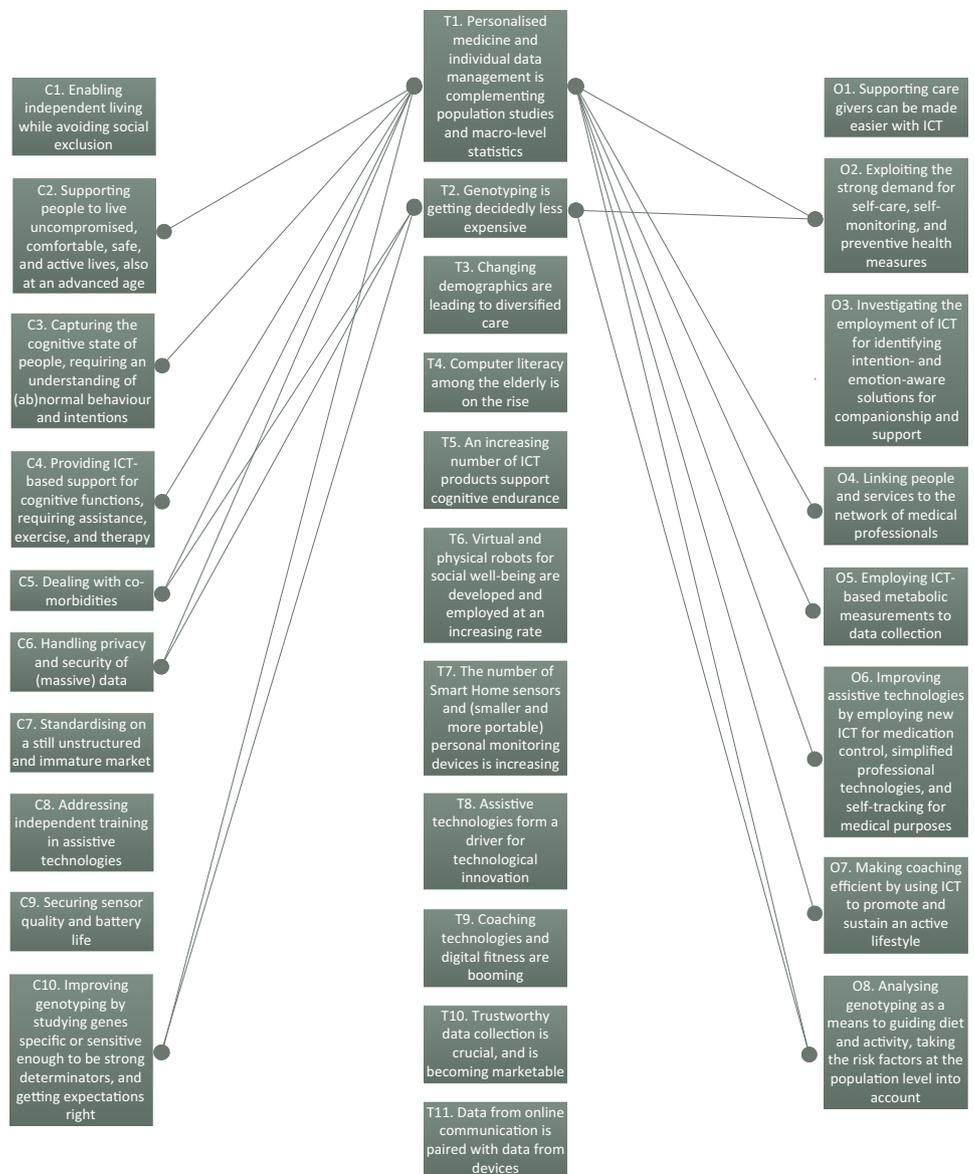
map will cover general and relatively inert properties of the environment, like air quality. It will resemble an ordinary weather map of today, but will in the future have higher reliability and detail, taking into account every new construction site in a city, every traffic accident on a freeway, and every dead deer in a stream on a mountain side. Considering the current trajectory of increasing public data availability and its benefits, it should be freely available to all, at all times. The construction of the global map, once standards are in place, will most likely be a distributed effort, involving nation states, private institutions and multi-national enterprises, as well as individuals. The map describing these pre-computed disease relationships, disease trajectories, and distances to molecular wellness define the landscape of wellness, health, and disease over time. Second, the local map will cover the rapidly changing properties of the local environment, what we call *Now!-properties*, pertaining to the customised view of the moving individual. The local map will take into account proteomics and the relevant molecular patterns in the body [23], complementing the static profile with dynamic data resulting from mobility [31].

The integration of global and local maps has been a popular hybrid model for situated cognition, planning, and localised decision making in robotics for 25 years or more (see, e.g., [24]). The methodological lessons learned from the relatively simple cyber-physical systems thus far engineered are important, albeit a meagre start. What remains is to construct a data-driven model of the interactions between the above factors, leading to efficient predictive medicine [19]. This will allow for extended rights and privileges to individuals concerning their own—now calculated, in a Leibniz sense—risks to their own wellbeing. At the societal level, policymakers are free to design targeted approaches for those individuals who are more resilient to different kinds of interventions, and inform them via official channels. The corresponding identification of associated molecular patterns facilitates the identification of risk groups susceptible to not responding to current programs for intervention. This is not merely a program for modelling disease, it is also a program for maintaining wellness that is currently being studied scientifically, in several ambitious empirical studies (cf., e.g., [13]).

Such a data-driven computational model will enable and support individuals to become informed decision makers, and support changes in their diet- or lifestyle-related choices. Whether or not this is welcome to all stakeholders (individuals, insurance companies, GPs in private clinics, etc.) is to a large extent hinging on the kind of added value this brings, and how such value chains are developed and presented. Research and development efforts are also required, in which health data will be structured and

massive data integration performed. This would allow for computation of relative risks and co-morbidities necessary, including the link to genomic and molecular information. The resulting systemic model could then be predictive: if a geographic region and a time are fixed, and a certain disease pattern presents itself, then there is set of probabilities on developing additional diseases, grounded in the co-morbidity data and in the interaction with lifestyle on molecular entities.

The stakeholders in processes concerned with making sense of data through intelligent analytics are numerous. Given that policymakers are currently the main stakeholders, others will surely benefit too from an ideal system of systems in which policymakers create fair and optimal solutions through their informed decisions. Networked foresight may include strategic as well as corporate issues: making data available to policymakers may mean informing not only government legislators or those responsible for clinical guidelines for governing data, but can also affect commercial decisions related to the resulting regulations. Hence, strategic foresighters are monitoring health data reforms, such as governmental and industrial decisions on making some health data public, continuously, to track disruptives. As a case in point, wellness tracker data is already being used in court cases, by insurance companies, and for advertising purposes (see, e.g., [27]), even as the regulatory landscape itself is still forming. This leads to a situation where the workflow between data scientists and practitioners in medicine is under rapid methodological development, incorporating critical perspectives [30], highly contingent on policy developments (cf. [7]), and linked to foresight.

Nowhere is this clearer than in the context of massive data sets. These can be pragmatically defined as data sets impossible to analyse directly, due to their size. The vast majority of health data is structured (cf. Table 1), and while electronic health records can be long and numerous, the digital handling and maintenance of the documents is not a difficult massive data issue. By contrast, much of the data about individuals not directly but indirectly relevant to their health and wellbeing is unstructured and definitely massive [4]. An incomplete list would include data on an individual's use of transport, purchases, active memberships, phone calls, tweets, blog posts and comments, and forum posts.

Disease surveillance is performed for both communicable and non-communicable disease, and includes case and lab reports. This is the foundation of clinical work, occasionally assisted by artificial intelligence tools at the time of individual diagnoses [34, 32]. Syndromic surveillance adds data originally collected for other purposes [15], using methods that rely on detection of early signals discernible already before confirmed diagnoses are made. The
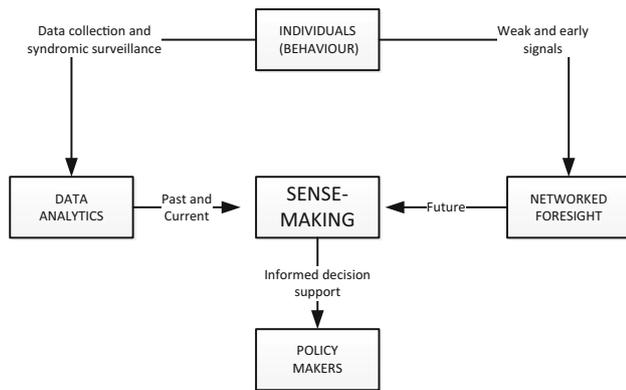
Fig. 2 A systemic model of sensemaking, with policymakers as the main stakeholders

resulting systemic model in its simplest form employs both networked foresight and syndromic surveillance (cf. Fig. 2), and to understand the sensemaking in concrete terms, we now turn to a case study.

4 Case Study

Among the many products covered in the technical foresight reports on active healthy ageing in the last year, we select for a detailed study here only one. We observe part of the process in the construction of meaning of data by observing how users of a very popular tracker, *Fitbit*, engage with each other in order to make sense of the data that they produce as a result of their own activities. We look at instances where users report having difficulties with quantification of their activities, in an online forum. We aim to illustrate some of the social practices, bodily adjustments, and technical configurations that go together with tracking.

In the Fitbit supporting material, health is described as an achievable state where humans, inhabiting bodies that can be trained to become better, consult the systems as authorities on the current state of their own bodies. The most significant claim is that access to information about how individuals live their lives: what they eat, how they sleep, and how much they move, is crucial for achieving fitness. Fitbit also emphasises partnership, "[w]e'll help you achieve what you set out to do, by sharing a full picture of your progress over time" [11]. Through such statements, entire lives of individuals become objects to be improved, and the systems position themselves as privileged observers able to see lives from a perspective not accessible to the individuals themselves in isolation.

Fitbit is able to produce this perspective on sleep and physical activity using data generated automatically by the sensors, as well as additional information entered

manually by its users. This perspective is constructed by combining different interpretations of input data into metrics. Fitbit provides a user portal where individuals can manually input data on activities such as cycling or yoga, which would not register on a motion sensor because they do not generate accelerometer readings (due to the relatively stationary position of the body). The manual logging of food intake is also a feature, translated as calorie consumption.

Using different forms of surveillance, the traces of behaviour are thus translated into calorie counts, sleep efficiency statistics, and number of steps taken per day. These are displayed to the users as user profiles that can be viewed using smartphone applications and the online portal. These profiles constitute the truth, as constructed by the activity tracking, and it is through them that individuals are encouraged to change their lives for the better. This is done by quantifying the effects of eating less or sleeping more, for example, or setting personal goals for a given time period, such as walking two kilometers in one day. Fitbit then converts a goal into a dynamic event, and starts reporting what percentage of the goal is already completed and what remains to be accomplished, e.g., 4 % of 70,000 weekly steps.

In order to count steps, Fitbit uses personal details such as the gait length and weight of the individual, together with the sensor data. Some of these calculations are detailed in the user manual, but the discussion boards are often used to make sense of how calculations are made. We observed that specific kinds of expertise are required in order to make sense of the data created and displayed. Individuals unfamiliar with specific fitness terms and acronyms, such as BMR (Basal Metabolic Rate) or MET (Metabolic Equivalent of Task) often get confused about how calculations are made, expecting the tracker to count only steps and calories. The example below is a reply to a user who asked why the steps taken during the day were not matching her calorie count (all user names have been changed in order to respect user privacy):

> it's not the number of steps, it's the time spent at the activity detected. If Fitbit detects you walking at 3 mph for 20 mins. it assigns calories to those minutes based on the METs value of that activity, which is a multiplier of your BMR rate. If it detects you sprinting, same. So the steps count itself is somewhat incidental. (May)

Fitbit calculations are to some extent based on standard measurements, via built-in automatically activated features that users have the option to disable, if they know how. Assumptions include specific positions about when days and weeks start and end, for instance. These

positions may conflict with the individual's own perception of time:

> When I woke up this morning I did my normal morning routine of showering and getting ready and then I got a buzz on my wrist from my fitbit that says I am an overachiever and I have already logged over 1500 steps today. (pip024)

The post above appears in a thread where the user tried to make sense of how Fitbit calculates averages for her day. The day starts at 12 a.m. and, other than doing time zone adjustments, users have no option to adapt it to other custom times. Replies to the thread above suggested that maybe she was awake past midnight, and all that night activity has been counted for the next day, turning her into an 'overachiever.' Fitbit also estimates BMR using generic information such as age and gender, and so acquires the role of the expert that makes complex metrics durable and unproblematic. Such *torque* has been proposed to examine the consequences of classification in individual biographies, defined as "the twisting that occurs when a formal classification system is mismatched with an individual's biographical trajectory, memberships, or location." ([6]:223). Individuals voluntarily enroll themselves in classification schemes, but when torque becomes apparent and they no longer recognise themselves in the data, individuals seek to open the black box holding the software calculations, and the process of sensemaking is made visible in the forums.

Living with a personal tracker has the potential to make other subtle changes to everyday life. For example, one needs to remember to charge it at adequate times, such as before performing physical activity, and take care so that the tracker does not break or get lost. Failure on the part of the individual to accomplish tasks, such as correctly entering and exiting sleep mode, may result in failure to detect and measure physical activity. Regarding quantification of activity, users report that the tracker sometimes calculates steps and activity incorrectly; when sitting in a car on a bumpy road, for instance. In the post below, a user becomes aware of how the wind resistance affects the accelerometer in the tracker:

> i had the same issue with the stepcount...guess its the wind that does the damage to the count. best option is to remove it for the ride shd be the movement in the arm but the wind resistance counts—deceives it. Got to experiment with a blow dryer and c how it responds....!!!! (pani)

In order for the step count to be accurate, this user chooses to remove it when riding a motorcycle. Habits of putting the tracker on and off depending on the activity

performed become a part of the routine of the individuals quantifying themselves.

> I hit the ground running this morning. I fed the dogs and took them out, cooked an E's Radical Omelet (hubby says "Ridiculous Omelet"), stopped in the middle of that to go to the shop, open the barn, take out the golf carts, check in 4 golfers, come back, finish cooking, gobbled it down, back to the shop for more golfers (oh yes, somewhere in there I dressed!), loaded and unloaded wood (we're supposed to have ice tonight and tomorrow), AND I have no steps. Guess what? I FORGOT to exit the sleep mode. I am furious. Probably got in my 5,000 steps EASY and nothing registers. (mahaffa)

One other instance in which the work of individuals is visible is when they appropriate the technology in order to enhance comfort or accuracy of tracking. When developing an understanding of the automatic calculations performed by the software, or adopting new habits so that calculations can be performed, individuals participate in the construction of data about themselves. Often, that participation assumes a more active role, when individuals adapt the technology to their own ends.

> You might want to experiment with other options for wearing the tracker. Some people have found they can get tennis wrist band (terry cloth) and cut a slit in it to slide the fitbit in. Or they use an old sweat sock. The fitbit band itself I have always found to be uncomfortable as well. (trainee)

This post is about modifying the bracelet so that it becomes more comfortable to use during sleep. During everyday use, other users have replaced the bands with more robust ones, so that they do not slip, or worn them in other places in the body. Altering the place where the tracker is may have consequences for how the calculations are performed, and produce less accurate results.

> If your stationary bike has reciprocating handles I found that attaching my fitbit to the handlebar with a rubberband works wonderfully. I have a fitbit 'One', so it's very easy. I just put the rubberband on the handlebar, and then clip the fitbit to the rubberband. (bolf)

In the post above, a user has asked if it is possible to measure activity done on an indoor bicycle. Fitbit trackers are made to count steps and, when worn during cycling, it tries to classify that movement as steps, therefore providing presumably incorrect calorie estimation. The manual recommends that non-step based activities be manually entered in the online portal. However, some users experiment with placements that enhance the estimation of

tracker, such as seen in the post above. Even for walking, users experiment with placements on different wrists, feet, torso and hips in order to achieve estimations that they consider to be more accurate. Some of the adaptions can be also done in the software, in order to achieve more accurate quantifications. A user asking for the reason why Fitbit overestimates his walking distance obtained the following response:

> It's a known problem if you are walking fast (like 4 mph or faster): Fitbit will "think" you are running and will therefore use your running stride length to calculate distances. If you are never running, or walking most of the time, enter your brisk walking stride length in the field for your running stride length. This should give you more accurate distances (Trimique)

Individuals may be simultaneously using other trackers, for complementary features, or for connecting with another community. Sometimes, the reason for using other trackers or systems is to test the accuracy of Fitbit and its limits for representing physical activity.

> This seems so variable. I tested fitbit and my hrm on the elliptical a few times. It varied depending on my use. When I tried the elliptical with zero resistance, my heart rate was in middle cardio zones the whole time and I went as fast as I could. The elliptical machine gave the highest calorie burn estimate, second was my fitbit One and lowest was my heart rate monitor. To fitbit, it reported a lot of steps (understandable) and a fast speed like I was running–but it felt easier than running. Another day I tested with some mild resistance and the machine gave the most generous estimate, and my HRM and fitbit were pretty similar that time. When I tested with higher resistance—the machine gave the highest estimate, my hrm next highest and fitbit was a lot lower. I wear my One on my bra and was working the handles with my upper body, there is movement in my back so maybe this is why it did okay with the steps. (slymike)

In the example above, the user used a heart rate monitor (HRM) and a Fitbit on an elliptical trainer, a machine commonly found in gyms that with the goal of estimating calorie counting. Her post shows her varied experiments, varying different resistances in the machine and the measurements done. Often, experiments like this are done and shared between users to discuss how to combine data from different trackers, in order to estimate calories more effectively. Often, forum users recommend each other to use third-party trackers to track non-step based activities, such as yoga, and manually input the calories estimation of the other tracker.

5 Discussion

Fitbit can be pictured as a personal cloud of fitness-related data, shared with others under specific rules of how the fitness game is supposed to be played; data is shared with others in order to motivate them to exercise, and in so doing optimise certain personal metrics. We would now like to link our case study to the trends and challenges reported on in Sect. 3. Realising the idea of aggregating local clouds of personal data for wellness comes with a list of challenges [2]: privacy, technical difficulties related to the combination of disparate data sets or data collected at different temporal scales, and the need for novel analytical methods that can extract valuable knowledge from massive datasets. In the light of trend T1 and its associated challenges as identified in networked foresight above, we suggest to augment the list by the following recommendations.

First, in addition to considering privacy issues when aggregating data, we propose the notion of misrepresentation, as captured by the concept of torque [6]. Even simple classification systems like the physical activity classifiers that derive steps or calories spent per day depend on a number of assumptions, and these may contribute to misrepresentation of individuals in the system, or even contribute to making parts of their lives invisible. The difficulties experienced by individuals when they are quantifying themselves (*self-surveillance*)—the clash that may occur when their experience of the body meets normalised detection algorithms—are likely to be exacerbated by the realisation of aggregated datasets, with data from a large number of individuals (*mass-surveillance* in the form of aggregated self-surveillance). In particular, a third party looking at the data of many individuals would not have the same possibilities for sensemaking that one individual has when looking at her own data (cf. [27]). Data may not correspond to how individuals actually felt that they have moved, and if one is to take that as the ground truth, aggregated datasets would contain erroneous data. One way we suggest to overcome this challenge is to take into account qualitative data that pertains to individual interpretation of the quantified data, which entails developing novel analytical methods, or conducting studies to that effect.

Second, an additional set of problems is associated with how the need for normalisation via aggregation is likely to affect current quantification practices, as observed in the case study, as well as with the lives of individuals who do not (yet) practice self-surveillance. Maintaining the production of a constant stream of data requires a disciplining behaviour and leads to internalising certain practices. Those include the effort of carrying a device, charging it, calibrating it (e.g., entering and exiting sleep modes) and

general maintenance. For the individuals already enrolled in the practice of quantification, these efforts are already being put in, but may still be problematic, as we observed in the case study, and be a source of frustration. This can also be an issue for individuals beyond the reach of surveillance systems (cf. [22]), and for certain groups of users, such as the elderly, who might not want to or might not be capable of doing so. Any research into this issue would have to seriously consider the issues of using and maintaining wearable devices. We also observed that some individuals appropriate the system to their own ends, and share the means for doing so on the forum. This set of unscripted practices of quantification, which subvert the expected usage of the device, can be appreciated in two different ways, from the perspective of data aggregation. It is either a source of error, which would entail further disciplinary measures, such as discouraging unexpected usage, or a source of richness, in which case we would recommend means to data aggregation that allow for the individuals to define how they want to be represented themselves, thus encouraging variation.

Third, we observed that individuals are often found busy completing the patchwork of combining data from multiple data sources. Additional studies that look at how that effort is spent by the technologically minded individuals whom engage in such practices could shed light on how data may be combined in a micro, meso, and macro fashion. Learning from, and interacting with, the individuals already engaged is likely more beneficial than ignoring them altogether, which seems to be the current research practice.

Going back to Fig. 2, we may view the case study as an example of how to move from individual behaviour, via weak and early signals, to foresight. Customisation of trackers, i.e. personalising the process of logging (self-surveillance) data leads to a number of issues related to design, manuals, appropriation. All of these issues, in an ideal world at least, feed back to the developers of future releases of software and hardware; closing the gap between product and user experience. Foresight goes further, however, also notifying stakeholders about when and where this ideal process may go out of bounds, e.g. when proprietary product development is forced to move to open source, or when the user community directly assists in funding further development through crowdfunding. Since networked foresight relies on experts, this translates to employing the skills and experience of early adopters and prosumers: it is their ideas that are contributing towards the sensemaking. Looking instead at the left path in Fig. 2, we see that non-health indicators like how we travel, socialise with friends, and consume perishable goods feed into data analytics. Much of this is in the past tense, helping policy makers to understand trajectories at the macro (aggregate)

level, and so grasping state-of-the-art. Most of this data can be dealt with using ordinary statistical methods; which tracker is the most sold being one data point. Perhaps even more important is the grade of use, which is a more qualitative affair: statistics on which trackers people actually use and which ones are just left in the drawer, for instance [21]. Recently, much of the data has been too large to allow for ordinary statistical methods to be applied [1]. This includes data harvested from tweets, blogs, and posts; analysed for context and sentiment [29]. Here, big data analytics and novel mining methods to go with it are often required [18].

Finally, we have elsewhere explained how data science and computational approaches can influence as well as learn from health care policymaking [7]. We have here largely dealt with the individual as a future stakeholder, always keeping in mind the fact that the policymaker is the current one. We have also recently showed [4] how artificial intelligence can be a part of this future, and that human-machine interplay can then be a key to improved sensemaking [5]. The bi-directionality of the power of decision making is important, since the underlying data and the intelligent reasoning based on that data must serve both policymakers and individuals. In the future, general practitioners and other care givers may also seek to exploit data as a part of their daily clinical work. Most of our conclusions also apply to such a scenario, with its multiple uses and users of data.

6 Conclusions

We presented a systemic model in which networked foresight complements intelligent data analytics, resulting in a kind of requirements specification for sensemaking needed for their integration. We believe that data can serve by explaining the past and the current, while foresight can serve by explaining the future. Our proviso was first that policymakers are the chief stakeholders. We then broadened our view to include individual stakeholders, presenting a case study of one tracker for wellbeing. Measuring physical activity via trackers or other gadgets is a process of co-construction. Users of these systems actively engage in the interpretation of the data that is produced about them. They appropriate and they collaborate in a sensemaking effort that is centred around individual experiences and a willingness to share, often in the hope of getting something in return that 'makes sense'. Through the kind of anecdotal evidence we presented in our case study, we can begin to understand how far off autonomous sensemaking is, and what is required in terms of artificial intelligence to get there. We loosely outlined a vision of the future of medicine where all sorts

of data analysis is integrated, in an intelligent fashion, to assist the individual, as well as other stakeholders, in such processes.

Our main conclusion is that the same kind of mundane problems that meet a user of a tracker today will have to be solved and taken seriously in future integrated solutions for personalised medicine. If engineering and economics issues are solved, there is still the end user to consider, to whom added value must be explained. Those end users will typically not be patients, but people in good health. The wellness trajectories constituting the network of networks will then have to be understood, all the way down to the molecular level. Even if such a bold vision of the future serves only as a metaphor or a grand challenge, it is highly stimulating driver of innovative ideas and concepts since the end result could mean increased health and wellbeing for all.

References

1. Committee on the Analysis of Massive Data, Committee on Applied and Theoretical Statistics, National Research Council (2013) Frontiers in massive data analysis. The National Academies Press, Washington, DC

2. Barrett MA, Humblet O, Hiatt RA, Adler NE (2013) Big data and disease prevention: from quantified self to quantified communities. Big Data 1(3):168–175

3. Boman M (2014a) Active healthy ageing. Innovation Radar, White Paper TR2014-004, EIT ICT Labs

4. Boman M (2014b) Who were where when? on the use of social collective intelligence in computational epidemiology. In: Miorandi D, Maltese V, Rovatsos M, Nijholt A, Stewart J (eds) Social Collective Intelligence, Springer, Switzerland, pp 203–225 (Computational Social Sciences)

5. Boman M, Gillblad D (2014) Learning machines for computational epidemiology. In: Proc 1st Workshop on Computational Epidemiology, IEEE Big Data, pp 1–5

6. Bowker GC, Star SL (2000) Sorting things out: classification and its consequences. The MIT Press, Cambridge

7. Cakici B, Boman M (2011) A workflow for software development within computational epidemiology. Comput Sci 2(3):216–222

8. Casati F, Kosman R, Baker K (2014) Physical wellbeing for active healthy ageing. Innovation Radar, Foresight Technical Report TR2014-001, EIT ICT Labs

9. Cowan R, Jonard N, Zimmermann JB (2007) Bilateral collaboration and the emergence of innovation networks. Manag Sci 53:1051–1067

10. EIT ICT Labs (2014) Create business intelligence of the future. http://www.eitictlabs.eu/about-us/strategy/innovation-radar/ Online; Accessed 19 Jul 2014

11. Fitbit (2013) Fitbit tracker product manual. https://www.fitbit.com/manual, Online; Accessed 19 Jul 2014

12. Fitbit (2014) Fitbit official site: Flex, one and zip wireless activity and sleep trackers. http://www.fitbit.com/home, Online; Accessed 25 Mar 2014

13. Gibbs WW (2014) Medicine gets up close and personal. Nature 506:144–145

14. Hallberg J, Casati F, Hedman A, Plomp J, Synnes K (2014) Mental wellbeing for active healthy ageing. Innovation Radar, Foresight Technical Report TR2014-002, EIT ICT Labs

15. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D (2004) Syndromic surveillance in public health practice, New York City. Emerg Infect Dis 10(5):858–864

16. Heger T, Boman M (2014) Networked foresight: The case of EIT ICT Labs. Technological Forecasting and Social Change In press. Available: 10.1016/j.techfore.2014.02.002

17. Heger T, Bub U (2012) The EIT ICT Labs—towards a leading European innovation initiative. it - information technology 54:288–295

18. Holst A, Ekman J (2011) Incremental stream clustering for anomaly detection and classification. In: Proceedings SCAI'11:100–107

19. Hood L, Auffray C (2013) Participatory medicine: a driving force for revolutionizing healthcare. Genome Medicine 5:110–113

20. Hood L, Heath J, Phelps M, Lin B (2004) Systems biology and new technologies enable predictive and preventative medicine. Science 306:640–643

21. Ledger D, McCaffrey D (2014) Inside wearables: how the science of human behavior change offers the secret to long-term engagement. Endeavour Partners, white Paper

22. Lyon D (2007) Surveillance studies: an overview. Polity Press, Cambridge

23. Marx V (2014) An atlas of expression. Nature 509:645–649

24. McDermott D (1990) Planning reactive behavior: a progress report. In: Sycara KP (ed) Innovative approaches to planning, scheduling and control. Morgan Kaufmann, San Mateo, CA, USA

25. National Research Council (2011) Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. Committee on a Framework for Development a New Taxonomy of Disease, Washington DC

26. OECD (2014) Unleashing the power of big data for alzheimer's disease and dementia research: Main points of the OECD expert consultation on unlocking global collaboration to accelerate innovation for alzheimer's disease and dementia. OECD Publishing, digital Economy Papers No 233

27. Olson P (2014) Fitbit data now being used in the courtroom. Forbes, TECH 11/16/14

28. Rohrbeck R (2011) Corporate foresight: towards a maturity model for the future orientation of a firm. Physica-Verlag, Springer

29. Sahlgren M (2006) The word-space model. Ph.D. thesis, Stockholm University, department of Linguistics

30. Sanches P (2015) Health data: representation and (in)visibility. Ph.D. thesis, Royal Institute of Technology, KTH/ICT/SCS

31. Sanches P, Svee EO, Bylund M, Hirsch B, Boman M (2013) Knowing your population: privacy-sensitive mining of massive data. Netw Commun Technol 2(1):34–51

32. Shortliffe EH, Buchanan BG (1975) A model of inexact reasoning in medicine. Math Biosci 23:351–379

33. Synnes K, Reithinger N (2014) Social wellbeing for active healthy ageing. Innovation Radar, Foresight Technical Report TR2014-003, EIT ICT Labs

34. Upbin B (2013) IBM's Watson gets its first piece of business in healthcare. Forbes, TECH 2/08/13

35. Van der Duin P, Heger T, Schlesinger MD (2014) Towards networked foresight? exploring the use of futures research in innovation networks. Futures 59 (in press)

Magnus Boman Computational epidemiologist and a professor in Intelligent Software Services at KTH, Stockholm. Senior expert researcher at SICS (Decisions, Networks and Analytics lab), Stockholm. Leads the Innovation Radar of EIT ICT Labs since 2011.

Pedro Sanches Ph.D. student at KTH, Stockholm. In the finishing stages of his doctoral thesis on health data analytics, financed by FCT.