Big Data, the e-health bubble and its fix

Abstract
Big Data makes it possible to find patterns in datasets that would otherwise stay hidden. An interesting edge case is that of an e-health system that performs automated tailored interventions based on Big Data analysis. This case is hardly covered by the legal frameworks for treatment. That leads to some ethical questions: is such a system desirable at all? In favor of such an automated system is the increased effectiveness of the tailored treatment. But that has a downside in the form of the creation of an e-health bubble, a personalized environment that limits choices. The negative effects of that bubble need to mitigated, by new ways of protecting the patient and by new ways of empowering.

Big Data

In 1998 the biggest Dutch supermarket chain, Albert Heijn, introduced a loyalty card for their customers. For the retailer collecting personalized data about their customers was just a logical next step in their ongoing battle to optimize their processes. But on a larger scale it made clear that we entered an era where data collection and data mining are important tools for managing complex processes. It was the same year that the term “Big Data” was first introduced. John R. Masey published a highly technical paper on the ability of computers to store and process big datasets (Masey, J. 1998. Big Data... and the Next Wave of Infrastress. 1999 USENIX meeting). He predicted, remarkably accurate, for the decade to follow the growth of and the bottlenecks in the capacity of computers to store, process and communicate large amounts of data. During that decade more and more research was done on methods for analyzing such huge datasets.

By 2011 the McKinsey Global Institute published a study about the opportunities of Big Data and the techniques used for it (Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A. 2011. Big data: The next frontier for innovation, competition, and productivity). They intentionally give a moving and subjective definition of “Big Data”: ““Big data” refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.” (p. 1). A definition based on a fixed size of the data set would be obsoleted by advancements in technology and it would not be sensitive for differences between the various sectors. But this definition has a problem: what is a “typical database software tool” is not only a moving target in size of the data set that can be processed, but also a moving target in the techniques that are used. Tools for unstructured databases, for distributed storage and for analyzing data from heterogeneous sources have become more and more common over the years.

Therefore it would be better not to define “Big Data” by the size of the data set or the tools used to process the data. That brings us to the question what makes “Big Data” in the e-health different from a standard electronic health record (EHR) or standard research databases? Key here is that “Big Data” tends to deal with more data then is human comprehensible. One EHR can for example be quite comprehensible by a physician, but the EHRs of a thousand patients not. A few minutes of cardiogram readout is quite comprehensible, a 24/7 readout not. In a similar way, humans have limited capacity to link data from various sources, certainly when those sources differ in how structured the data is. So key to “Big Data” is transforming the data into something that is comprehensible.

There is a wide range of techniques available for transforming the data. (p. 27-31) These techniques can identify relations between objects or between their properties, they can identify groups of objects with similar properties or they can summarize the data to a trend. An other technique that can be used in “Big Data” is running large simulations and analyzing the results. Though the techniques can be as simple as a pivot table or a classical correlation correlation coefficient, often more advanced operations are used to find patterns. Many of them involve machine learning or

other artificial intelligence techniques. So key to “Big Data” is finding patterns that would otherwise stay hidden to the human eye. So for the scope of this paper I define “Big Data” as: “the computer assisted or fully automated processing of datasets to recognize patterns that would stay hidden otherwise”.

Big Data in e-health

In the healthcare sector, Big Data was first introduced in the fields of radio diagnostics and pharmaceutical research. But right from the start researchers have been looking at many more potential benefits of Big Data in the healthcare sector. Partly because it can potentially improve the the quality of care immensely. The ability to segment the populations to customize actions, the possibility to replace or support human decision making with automated algorithms and the perspective for innovation, new products and services (Manyika, J. et. al. p. 5-6) look very promising for the healthcare sector. The other reason why healthcare is a very attractive field for the deployment of Big Data is because healthcare is a big sector within the economy and commercial players within the sector can gain huge profits. “Big Data” gives players that are traditionally not active within the healthcare possibility to set foot aground (Engelen, L. 2014. Blog: wachten op de Uber van de zorg? http://www.smarthealth.nl/trendition/2014/10/30/blog-wachten-op-de-uber-van-de-zorg/).

Three recent developments have strengthened this trend. First of all the locale patient records are more and more shared between the various involved health professionals. With this standardizing and sharing of health records, big time data mining becomes more and more feasible. Secondly the introduction of e-health, e-mentalhealth and variants like blended care, where part of the treatment is done in e-health system and a part face to face, have vastly increased the amount of data available on individual patients, while automated tailoring of the online therapy becomes more or less a standard tool for e-health environments. Finally the introduction of wearables makes it much easier to monitor patients 24/7, with it creating vast amounts of data on the patients. Right now the use of Big Data in healthcare has become a big trend. No healthcare congress can be attended without stumbling over Big Data presentations in one way or an other. The amount of publications on the topic is hard to follow because of the sheer amount of them.

In this paper I want to focus on the legal and ethical implications of this trend. But before narrowing this down, it is wise to put the use of Big Data and e-health in perspective. The best and most authoritative data in the use of e-health in the Netherlands is the yearly Nictiz and NIVEL e-health monitor. The 2014 edition makes clear that almost all healthcare professionals use electronic patient records and that a vast majority of these systems has some kind of automated alarms. (Krijgsman, J., Peeters, J., Burghouts, A., Brabers, A., de Jong, J., Beenkens, F., Friele, R., van Gennip, L. 2014. Op naar meerwaarde! E-health monitor 2014). But it is not clear to what level these alarms are based on Big Data methods: most of the described functions can be implemented without Big Data analysis. Exchanging health records in a digital way is supported by a majority of the professionals, but often the systems are not compatible and is the possibility to exchange data limited. When looking at remote treatments, online testing and remote monitoring, the picture is less clear and quite different. Not more then 10% of the treatments involve some kind of online testing, remote monitoring or online treatment. Looking solely at the online treatment, the figure is more in the order of 1%. Though it is likely that in a part of these treatments some Big Data technique is involved, it is not known at all in what percentage. But when asked what promising ICT developments are in the healthcare, 39 out of 152 practitioners answer “electronic health records”, 26 say the “introduction of healthcare apps”, 19 say e-mail consult. Only one of the responders mentions “Big Data” (p. 153). So all in all it is safe to conclude that at this moment Big Data in e-health is an hype with very limited use in practice. But given the amount research done on Big Data applications in the healthcare, it is also safe to conclude that in the future Big Data will enter the realm of e-health.

Though it is clear privacy law applies to the deployment of Big Data in general (Manyika, J. et. al. p. 11), there is no overview of the legal implications of Big Data when it is deployed inside the healthcare. Law is always based on a ethical consensus. So when looking at the legal implications of a new development, it is also good to have a look from the ethical perspective.

So, we can expect Big Data to become more and more important in the realm of the e-health. But what are the legal and ethical implications of that development? To answer this question I will take a look at the edge case of a e-health system that performs Big Data analysis and uses the results to tailor an intervention for one patient. This use case of Big Data is often regarded as a use case with a big potential in healthcare. I will explore both when that system is fully automated and when it has a role as decision making aid.

Legal framework

There are two legal frameworks that specifically regulate the processing of data for diagnosis and treatment in healthcare. The first is the law on treatment contracts (WGBO in the Netherlands), which also partly regulates medical secrecy. The second is the European regulation on Medical Devices (MEDDEV).

There are several reasons why patients need extra legal protection. First of all patients have per definition a dependent relation with the practitioner treating them, so the power between the practitioner and the patient needs to be balanced. Secondly, healthcare can involve an invasion of the physical integrity of the patients, so consent of the patient and protection of the patient are needed to protect the patients basic rights. Finally many health problems are a social taboo, so a separation between healthcare and the social circles of the patient is necessary.

To protect the patient there are generally two strategies. The first one is empowerment of the patient. By giving the patient rights, choices and the means to effect them, the autonomy of the patient can be safeguarded. The second strategy is protection of the patient from harm being done by proving a treatment is save.

WGBO

The relation between the patient and the medical practitioner is regulated in “the law on the medical treatment contract” (“Wet Geneeskundige Behandel overeenkomst”, WGBO, Burgerlijk Wetboek, Book 7, Section 5, De overeenkomst inzake geneeskundige behandeling). The core of the WGBO is that all treatments must be done with an informed consent of the patient. The practitioner must provide good care, has to be transparent and has to be able to justify his or her actions. The WGBO also regulates medical secrecy, including an arrangement for scientific research.

An e-health environment can be subjected to the WGBO as long as it working under control of a practitioner. As soon as the e-health environment starts performing automated diagnosis or automated treatment, it becomes subject to the MEDDEV directive. An e-health environment may create large amounts of data. And all data that gives direction to the treatment has to be included in the medical records according to the WGBO. So all data that is used in a Big Data analysis has to be included in the medical records, with a real risk that the size of the medical records exceeds by far the boundaries of practicality. Data of multiple patients may be analyzed when it is part of scientific research, but only after an informed consent of the patient. This also applies to Big Data analysis. But note that for each scientific research project a new informed consent is needed.

Now lets have a look at the use of Big Data analysis for tailoring an online therapy. If this process is fully automated, the e-health environment is subject to the MEDDEV directive, because it automatically conducts (a part of) the treatment. When this tailoring only happens after

authorization of the practitioner, it is subject to the WGBO. But because Big Data deals with patterns that are not visible without the computer analysis, it is questionable to what extend the practitioner will be able to check the validity of the pattern the computer has found. It becomes hardly possible for the practitioner to justify his or her actions and to give transparency to the patient about the chosen treatment. And even if the practitioner is able to do so, it is likely that it would result in a too big workload for the practitioner to do that level of checks for each patient. Then likely the situation arises that the practitioner just clicks 'ok' to accept suggestions from the automated system. When that happens the practitioner doesn't comply anymore with the obligation of the WGBO to give good care. So the use of Big Data to tailor an online therapy leads the e-health environment out of the realm of the WGBO and into the realm of the MEDDEV directive.

MEDDEV directive

Compared to the WGBO the MEDDEV directive (European Union. 2012. MEDDEV 2.1/6) is much less aimed at empowerment and much more at control. The MEDDEV directive applies to all devices that automatically perform diagnosis or treatment. After the informed consent, the medical device must have an evidence based working and there must be proof that its functioning is correct and safe. Software that works under the directive must be programmed under a regime of strict quality control. If we take the example of the tailoring of the e-health intervention based on Big Data analysis, then this leads to two fundamental problems and one practical problem.

On a fundamental level there is a gap between 'evidence' in the sense of 'evidence based medice' and the kind of information Big Data analysis delivers. Big Data analysis searches for patterns and may, at its best, deliver a statistical correlations between events. But for two reasons a statistical correlation is not yet evidence for an effective treatment. An evidence based treatment assumes causality between two events. A statistical correlation is only a part of proving causality, the other parts of that are that the two events follow each other in time and that there is a logical connection between the two events. The other reason why Big Data analysis is not yet evidence is that Big Data analysis is very prone to a common statistical fallacy. How strong a correlation between two events is, is expressed by a p-value. This p-value indicates how big the chance is that the result is pure coincidental. A p-value of 0.01 indicates that 1 in the 100 results are purely coincidental. But most of the methods used for Big Data scan large amounts of variables for connections. If the threshold is set to a p of 0.01 then on average for every 100 scanned pairs of variables 1 false positive emerges. So the results of Big Data analysis have to be validated in a separate research. Big Data analysis is not enough for evidence based medicine.

That raises the second fundamental issue: Who should do the MEDDEV evaluation for the tailored e-health intervention? Can the system generating the interventions be evaluated as a whole by the vendor selling it or must the organization deploying it evaluate each single intervention? The latter is not feasible because the MEDDEV directive is aimed at large scale sale of devices that deliver a more or less uniform treatment or one kind of diagnosis. The only feasible option is to evaluate the system as a whole. But the more advanced the system is, the more variables it takes in account when recognizing patterns in the behavior of individual patients and the bigger the range of possible interventions becomes. With this increase of complexity the chance of the system doing unexpected interventions increases too, we even expect a system like that to see patterns humans would miss and react to it. But with this increase of complexity it will also become more and more difficult to prove the system will stay within safe limits or not engage in patterns that are harmful for the patient. So automated interventions can within the MEDDEV directive only be personalized to a limited extend. This limit won't prohibit the use of tailored interventions based on Big Data altogether, but it is a severe limitation to its potential.

There is a more practical issue with one of the other demands of the MEDDEV directive. It demands the software is programmed under a strict quality control regime, what is quite hard to do.

But when the software contains things like machine learning algorithms or when it contains large portions of third party software (like a database manager), it becomes next to impossible to adhere to the quality control regime. So the software used to perform the Big Data analysis probably won't be acceptable for the MEDDEV directive.

So in theory it is possible to work around the limitations of the MEDDEV directive to deliver automated tailored e-health interventions based on Big Data analysis. But in practice it won't be feasible to do so.

Ethical perspective

All in all the the edge case of this paper is not very well covered by the current legislation. So we have to raise the question if it is desirable to deploy such a system at all. If so, we should work out how to adapt legislation to regulate such systems.

The list of potential advantages of automated tailoring of e-health interventions based on Big Data is quite clear. It makes it possible to do better suiting interventions at the optimal moment. So the treatment can be much more effective. This is a huge benefit, first of all for the patient, who has to suffer less from the disease. It is a virtue to deliver the best possible healthcare. Also the people in the direct surrounding of the patient can be relieved by such a better treatment. And it also has the potential to improve the quality of the healthcare for equal or lesser costs, what would be a huge benefit for all of society. So enough reason to consider deployment of such systems.

At first sight the downsides of such systems are much less clear. But when looking at the legal issues of the edge case, it becomes clear that it moves the treatment out of the relation between patient and practitioner and into a triangular relation of patient, practitioner and automated system. The current ethical frameworks for healthcare, medical ethics and ethics of care, both presume a relation between humans and are therefore not directly suitable to asses the relation between a human and a machine. But those ethical frameworks do point to some basic values as physical integrity, autonomy, the importance of relations and the functioning of dependency. So how can we understand these relations between humans and machines?

What sets Big Data processing apart, is that it creates an interpretation of the situation and that the automated tailoring creates a personalized surrounding accordingly, all without the relation that is typical in a classical healthcare setting. In that way the edge case system creates an “e-health bubble”, the e-health equivalent of the “filter bubble” (Pariser, E. 2011. The Filter Bubble: What the Internet Is Hiding from You). Often this e-health bubble will be clearly visible: many therapeutic interventions are directly noticable when made. But the tailoring may also be used to subtly adapt some incentives to change the patients behavior. Also the Big Data analysis itself is part of the bubble. The processing makes patterns visible, but you can never be sure if, and if so which, patterns stay hidden.

Can you make your own decisions when your world is shaped for you? Is it possible to give an informed consent for entering a bubble or is it not possible to comprehend what you sign up for on forehand (if at all)? It is for sure the e-health bubble limits the choices of both patients and practitioners. Questioning the bubble takes a lot of knowledge, and if a patient lacks that knowledge he or she is easily dismissed as a Don-Quixote. Or even worse: as denying proper treatment and a danger for him or herself or as a nuisance and a cost for society. So on what ground do we justify this intervention in peoples life and the decrease of autonomy? A “better treatment” is not enough answer to justify these downsides. We need to mitigate the downsides before deployment of our edge case becomes acceptable.

Underneath is the question how we regard the patients in our healthcare system. If we see them as

beings in need of a fix, then an e-health bubble is a perfectly efficient way of fixing the patient. But if we see the patients as beings in need for autonomy and in need for room for their own decisions, then an e-health bubble is a awful distopy. Both views on patients are valid at the same time: when healthcare doesn't provide any fixes, it is doesn't quite live up to our expectations. But at the same time we want healthcare to respect our autonomy and to offer us choice. So how can we safeguard autonomy without killing the potential benefits of Big Data controlled automatic interventions?

Fixing the e-health bubble

To safeguard patient autonomy in the e-health bubble, we can fall back on two strategies: protecting the patient from harm done by the e-health bubble and empowering the patient so he or she is able to counter the bubble.

Previously we saw that the MEDEV directive, the current framework for protecting patients when they are being diagnosed or treated by automated systems, falls short. It is not flexible enough and not suited for individual tailored interventions. It is not possible to prove on forehand each bubble is safe and effective. The most easy fix would be to bring back the human supervisor. But we have seen that for practitioners it is hard, if not impossible, to supervise the outcomes of Big Data analysis. Nor is it desirable to check and authorize each intervention on forehand. So it would be better to look at the performance of the system as a whole and monitor actively for adverse events that may be caused by it. So in stead of proofing on forehand the system is effective and safe, as the MEDDEV directive demands, it would be more realistic to monitor actively and rigorous for indications the systems performance is not safe, not effective or a threat for the patients autonomy. And of course intervene when there is any kind of contra-indication. That way the safety and autonomy of the patients can be safeguarded and the power of Big Data can be used to its full extend.

Also with empowerment we run into the limits of our current practices. An informed consent looses its value if you can't fully oversee the consequences of where you are embarking in. Being able to check the files looses its value when the files are too big to comprehend. Checking the notes of your practitioner looses its value when your practitioner is a computer. So lets return to what empowerment is about: it is about enabling people to make their own choices. The classic approach is to make sure the tools are there to make a choice. But when that approach fails because the system is too hard to oversee, we better may focus on the end result of empowerment: the possibility to engage in deviant behavior. When creating e-health bubbles, questioning the system and deviant behavior become great virtues, they are signs that people behave autonomous. So any questioning of the e-health system and any deviant behavior within or against it should be actively encouraged and rewarded. There must be attention and patience when a patient outs questions or engages in deviant behavior. In the follow up of such questions of behavior the goal should be to offer the patient alternative solutions, not to push the patient back into the bubble. This means special procedures and extra means for dealing with such situations. A small effort compared to the potential gain.