TOOLBOX

How people are critical to the success of Big Data

A buzz has emerged around Big Data: an emerging field that is concerned with capturing, storing, combining, visualizing and analysing large and diverse sets of data. Realizing the societal benefits of Data Driven Innovations requires that the innovations are used and adopted by people. In fact like in many other technological innovations people are critical to the success of Big Data; people are in many cases where the data is collected, people are in charge of interpreting and visualising the information that can be found in the data, and people are, in the end, also the actors that need to change their behaviour in order to make a change in 'the real world'.

Marc Steen, Maurice van Beurden and Jenny de Boer

The promises of Big Data

The promise of Big Data is that one will be able to find patterns in data and that this will generate insights that will help to make better business decisions, e.g., to send police officers to a particular site, which is likely to need their assistance, based on analyses of data of past incidents, combined with real-time data (*Economist*, 2013). Data Driven Innovation (DDI), i.e. the development of new products, services, or organizational processes, driven by Big Data, is seen as a solution to a myriad of problems. The OECD (2014) proposed that Big Data can help to improve productivity especially in sectors like public administration, education, and health services.

In order to clarify what we mean with Big Data, we follow the 3-Vs definition (Laney, 2001): we are concerned with relatively large *volume* data sets, with data from a *variety* of sources, e.g., to cover not only specific events, but also variables in these events' contexts, and with data generated with *velocity*, e.g., by streams of real-time data generated by sensors. In the research on Big Data, a lot of effort is focussed at technology development, e.g., to find patterns in data produced by large numbers of real-time sensors. Furthermore, the opportunities and benefits for business and commerce are also widely discussed, e.g., for database marketing and targeted advertising. In order to complement this body of knowledge, we propose to focus on the *role of people* (to complement a focus on technology) and on the realization of *societal* benefits (to complement a focus on business benefits).

Realizing the societal benefits of DDI requires DDI to be used (by people). There is a range of methods that facilitate the design and development process (of DDI) in such a way that emphasize the importance of taking into account people's perceptions and needs in order to create a product or service that people want to use (e.g. the Technology Acceptance Model (Davis 1989)). However we argue that in DDI the role of people reaches beyond the role of the (end-)user; if the aim of DDI is to achieve the societal benefits, people also have other critical roles:

1. People (or the organizations they work in) need to be able and motivated to capture and share relevant data; A DDI needs data, and many DDI need data from people, without the data the DDI is worthless.
2. People need to interpret the data analyses' results, and develop insights (often supported by software); A computer needs to know what it is looking for in order to develop sensible and useful insights. The right questions are formulated by people.
3. People need to be able and motivated to act upon these insights; to act in the real world and realize impact; Adaptation in behaviour is often required, and is crucial to cause a change.

Let us look at the example of an Air Quality Information Service (for example Longfonds, 2015). A smart phone app is used, e.g., by people with medical conditions who suffer from bad air quality. The app shows the current and expected air quality for specific locations. Its' goal is to enable people to anticipate bad air quality situations, e.g., by postponing outdoor activities, to promote their health. The app's information is based on fixed sensors in critical locations, e.g., industrial areas, and small sensors that people carry with them. One can choose to receive basic information, based on these fixed sensors and on forecasts that are automatically calculated using other data sources, such as weather forecasts. Or one can choose to carry a small sensor and to share this sensor's data and location, which are combined with other people's data, enabling the system to provide more detailed information.

Looking at the example above we propose that the third role is most critical because the potential benefits of Big Data can only materialize if the DDI provides insights which people can follow-up practically. In order to achieve this there is a need for DDI that people are willing to use; the perceived usefulness is in balance with perceived ease of use to speak in terms of the Technology Acceptance Model. This paper explains further what specific human factors challenges are in play for the three critical roles of people in the development and deployment of DDI.

Capturing and sharing relevant data

Many DDI rely on people's willingness and ability to share data, sometimes very personal data. In the Air Quality Information app example, the advanced version of the DDI requires that people share data on their location. This introduces a range of questions regarding privacy, legal matters, ethics and trust, all which are currently under debate. Likewise, when data come from organizations, such as a municipality's data on its citizens or a company's data on its customers, there are stakes at play. Not every organization or individual is willing to share their data, e.g., because of business interests or privacy concerns. Hence if a DDI requires effort from people to generate or share the data for a DDI to function (well) this will influence the perceived ease of use of the system and the likelihood of adoption and use of the DDI.

In 'Smart Dairy Farming' (Van der Weerdt & De Boer, 2015), for example, dairy farmers need to install sensors to log all sorts of data in and around their cows in order to receive cow-centric information which benefits the cow's well-being. They will only do that if they value the results the collection of those data provides them, e.g. generate cost reductions or quality improvements or help them to give the cows better living conditions that will result in animal welfare and a better product. Moreover, they will need to share their cows' data with some independent organization, and input additional data, for example on their cows, in order to help this agency to develop and fine-tune algorithms that make these DDI run. This requires an on-going effort in time and money, as data is needed real-time and continuously.

Said otherwise: the pay-off between 'gain' (what one gets from using the DDI) and 'pain' (efforts one must make to generate or share data, or to use the DDI) must be balanced. In the example of the Air Quality Information app, one can get more detailed information if one shares data - a direct, personal benefit. However in the case of Smart Dairy Farming a farmer is first required to install sensors, generate data, and only after some time will experience the benefits. If not managing these expectations about when benefits of use will occur, a farmer might discard the system in the first place.

Interpreting patterns and creating insights

A DDI typically requires that people play some role in data analyses, even though number crunching, analysis and visualizing are done by the computer. They can articulate hypotheses explicitly or implicitly, by selecting the data that goes into the analysis, or they can attribute meaning to the patterns found. These interpretive tasks rely on people, and thus involve people's perceptual, cognitive and social processes. This has pros and cons. People use their personal memories, which helps to understand matters, but also introduces subjectivity. They steer their attention, which helps to focus, but can also introduce blind spots. And people's reasoning can have diverse biases, such as confirmation bias (a tendency to confirm one's preconceptions), belief bias (to believe the logical steps to a conclusion that one agrees with) or social bias (to adhere to social norms). In very general terms, the solution for this is to enable people to do what they are good at, such as making creative leaps, to let computers do what they are good at, such as crunching numbers.

An example of how this combination can work comes from the usage of sonar to monitor fish on a fishing ship (Quesson, 2014). Normally, fishermen would interpret these images and make decisions about what to fish or not to fish, in order to reduce bycatch. They experimented with pattern recognition software, and found that this software was accurate 80% of the time, and that they could raise accuracy to 100% if they combined the software-generated advice with fishermen's knowledge (Kregting, 2015; Quesson, 2014). Interestingly, fishermen appreciate the fact that their knowledge is valued and are thus likely to adopt this system.

Another example comes from correlation and causation discussions in finding cures for diseases. Some patients capture their own personal, medical data and share this with other patients in order to find patterns, even without yet understanding the underlying mechanisms. The relations found in the data, are however not causations (read more about that in Mayer-Schönberger & Cukier, 2014). The correlations could help professional health practitioners in finding *possible* causal factors for diseases such as Parkinson (for example Michael J. Fox Foundation, 2014) or cancer (for example Bresnick, 2015), however further research should proof if the causation is indeed true.

Since so many processes are involved in the interpretation of data, one needs to identify who will be working with the data and what tasks need to be supported, and to develop appropriate software for these people and these tasks, which is a system design in itself; a system focusing on the exploration of data sources and finding possible interesting relational factors in the data. Next to that in developing a DDI the balance between the computer and human component should be found in such a way that it optimizes user's willingness and ability to adopt and use the DDI. What information is given to them and the way in which this is communicated to them is then crucial.

Acting upon these insights in the real world

People will have to act upon the information or advice produced by a DDI. The benefits of Big Data can materialize only if people or organizations are able to apply the insights produced and act upon them in the real world (OECD, pp. 18 and 23). For example, a DDI can help people to exercise more often if they use a Step Counter App on their smart phones to remind them of their ambition to walk a certain number of steps, or to drive more slowly if the traffic light systems uses algorithms to monitor and streamline traffic flows real-time. However what is much found is that people stop using the apps after some period of time, without having adapted their behaviour sustainably. This makes that societal (and personal) benefits that were intended with the app will not be achieved.

In all cases, people and organizations must be able to apply the information or advice. This will typically require a careful service design or interaction design process, in order to deliver a service or user interface that effectively informs people and motivates them to act. The information or advice must be clear and convincing. A DDI may need to, e.g., explain an underlying mechanism in order to motivate people to act.

Measures to help people to change their behaviours have been discussed in various fields. Michie et al. (2014) developed a taxonomy of behaviour influencing strategies for e-health applications. One in particular is also explained by Thaler and Sunstein (2008). They popularized the notion that one can nudge people towards specific desirable behaviours. The reasoning is that service design or interaction design entails the presentation of different options, and that by presenting option A more attractively than option B one can 'nudge' people towards choosing option A—without coercion. The example of etching the image of a housefly into the men's room urinals is well-known; these help men to improve their aim.

Next to a careful service design, also the supporting measures and the way the benefits of a DDI is communicated, sold or brought to people influences the extent to which the information provided is adopted. For example in smart industries, in which big data are used to optimize processes and service delivery for the industry, it has been identified that people need new skills to be able to adopt the DDI and adapt their behaviour (Smart Industry, 2014). These skills need to be taught and trained.

In sum, we propose that a successful DDI offers clear information and convincing advice, and motivates people to actually change their behaviours. A successful DDI will support people to make sustainable behavioural changes, beyond trying-out a DDI for a couple of days or during a specific experiment or trial project.

'Human Factors' in Big Data

Above, we discussed three critical roles of people that in order to successfully develop and deploy DDI: to capture and share relevant data, to interpret data analyses' results, and to act and realize positive, societal impact. Figure 1 gives an overview and summary of these different roles. The visualisation in Figure 1 can also be used to facilitate the discussion about the human factors that arise from the critical roles of people.

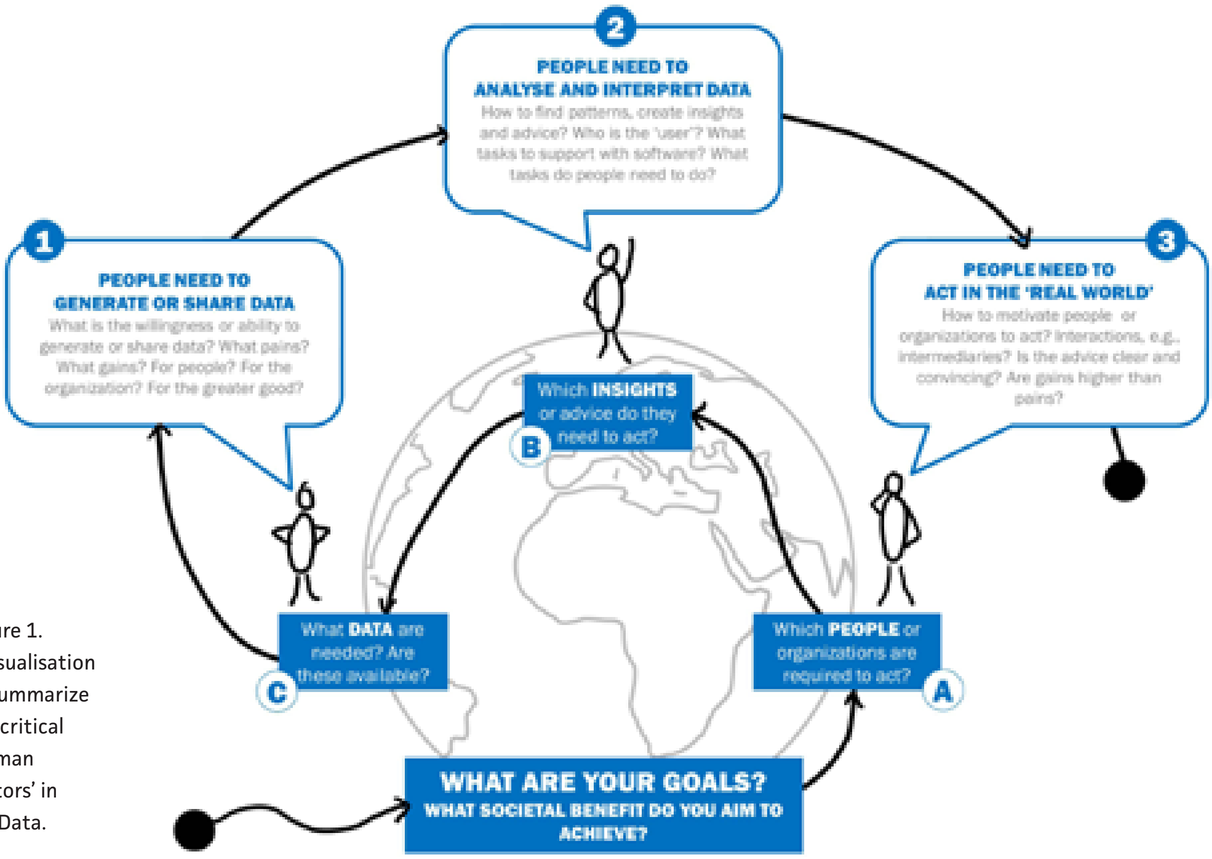

Figure 1. A visualisation to summarize the critical 'Human Factors' in Big Data.

With the goal in mind you first follow steps A, B and C. Then the three critical roles of people are listed clockwise (1, 2, 3). These roles can also be understood as chronological steps in a process of service development and deployment: (1) data are collected; (2) data are analysed (either during development, when people develop and tweak algorithms to model these patterns, or during deployment, when data are analysed real-time using these algorithms); and (3) data-based advices lead to behavioural changes. Moreover, the ultimate goal to generate societal benefits can be a starting point for a series of questions (counter-clockwise in Figure 1): What are your societal goals? Which people or organizations are required to act? Which insights or advice do they need to act? What data are needed? Are those available?

Conclusion and recommendations/relevance

The field of Big Data gives a lot of attention to technology and to business applications. In order to compliment this focus, we discussed the potential societal benefits of Big Data, e.g., for people's health, for sustainable development and the roles of people to reach the societal benefit. Insights from the human factor community are essential to further develop and optimize DDI such it will benefit performance *and* well-being. It requires diverse approaches in order to deal with the critical roles of people in the right way; for technology adoption (to balance pains and gains), behaviour adaptation (to achieve impact), and experts on human behaviour (to interpret patterns).
Although the field of Big Data is dominated by data mining algorithms, machine learning, and many reports emphasize the superiority of computers, for example during decision making, we emphasize that finally people are responsible for the outcome and apply results in real life, or alter their behaviour. Since data is seen as the 'new oil' it becomes more important in our life, both at work and at home. In the coming years many new innovations will be introduced (e.g., internet of things, decision support systems) that affect the way we life's and how we perform our jobs. Therefore a new discipline within the Human Factor community can arise that focuses on how data alter our behaviour, performance and well-being and how to do that in the right way to achieve the societal benefits it promises it can achieve.

References

Bresnick, J. (2015). 'Big Data Analytics Research Projects Target Cancer, Genomics' HealthITAnalytics.' Available at: http://healthitanalytics.com/news/big-data-analytics-research-projects-target-cancer-genomics/. Retrieved December 23, 2015.
Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly 13(3), 319-340*.
Economist (2013). 'Don't even think about it - It is getting easier to foresee wrongdoing and spot likely wrongdoers'. Available at: http://www.economist.com/news/briefing/21582042-it-getting-easier-foresee-wrongdoing-and-spot-likely-wrongdoers-dont-even-think-about-it. Retrieved January 21, 2016.
Kregting, A. (2015). In Depth: Combining the Data to Gain the Most. *World Maritime News*. Available at: http://worldmaritimenews.com/archives/169197/combining-the-data-to-gain-the-most/. Retrieved December 23, 2015.
Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. META Group (currently Gartner). Retrieved August 28, 2015.
Longfonds (2015). 'App "Mijn luchtkwaliteit" informeert snel over smog'. Available at: https://www.longfonds.nl/nieuws/app-%E2%80%98mijn-luchtkwaliteit%E2%80%99-informeert-snel-over-smog. Retrieved December 23, 2015.
Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, John Murray Publishers, London, UK.
Michael J. Fox foundation (2014). 'The Michael J. Fox Foundation and Intel Join Forces to Improve Parkinson's Disease Monitoring and Treatment through Advanced Technologies'. Available at: https://www.michaeljfox.org/foundation/publication-detail.html?id=555&category=7. Accessed December 23, 2015.
Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W. et al. (2013). The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions. *Annals of Behavioral Medicine, 46*(1), pp. 81-95.
OECD (2014). *Data-driven Innovation for Growth and Well-being*. OECD, Paris, France.
Quesson, B.A.J. (2014). TNO 2014. R11772 The SOFIC (Sonar for Fish Classification) project phase 3, TNO Technical Sciences, Den Haag, the Netherlands.
Smart Industry (2014). *Smart Industry – Dutch Industry Fit For The Future*, pp 33-35.
Thaler, R.H. & Sunstein, C.R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. London, England: Pinguin.
Weerdt, C. van der, & Boer, J. de (2015). Focusing on behaviour to ensure adoption of Big Data information services in Precision Livestock Farming. *Proceedings of European Conference on Precision Livestock Farming 2015*. Milan, Italy.

About the authors

M.G.D. Steen, PhD
Senior Research Scientist in Human-Centred Design
Open Innovation and Applied Ethics
TNO, The Hague
marc.steen@tno.nl

M.H.P.H. van Beurden, PhD
Research Scientist in Human-Technology Interaction and Human Factors
TNO, Soesterberg

J. de Boer, MSc
Medior Consultant Human Factors in Data
TNO, Soesterberg