4. Analytical Implementation

4.1 Core Assumptions of Big Data within Organizations

Big data will influence the organization and will have a strong impact on any organization. In order to understand the nature of big data within organizations, it is essential to relate them with several core assumptions. While, as formerly depicted, the interaction between big data and humans is highly complicated, big data are a social phenomenon. Actor network theory defines technology as yet another actor and, therefore, frames big data as an actor within the social network of an organization. Stein (2000) postulates that, consequently, all members interacting in an organization are not only influenced by a social dimensionality, but also a temporal dimensionality and a factual dimensionality. That, as the author states, co-aligns with the structuration theory as postulated by Giddens (1984). Giddens (1979) stated that time-space relations are increasingly important for understanding social interactions. Gross understands Giddens as follows: "He argues that all social systems must be understood as stretching over time and space, or better, 'embedded' in time and space" (1982: 83). This, to a certain extent, negates the former statement that actor network theory and structuration theory do not fit well together since, from the perspective of time and space, they are not contradictory. Law (1992) reports the ordering potential of time (durability) and space (mobility) within systems, thereby, as well as highlighting the importance of time and space within social systems.

In addition to the relevance of time and space, Kluckhohn and Strodtbeck (1961) discuss the dimensionalities of the senses, and classify them into temporal dimensionality, factual dimensionality, and social dimensionality. Space is absorbed within the factual dimension and is expanded. Stein (2000) uses these dimensionalities as his core assumptions for the developmental analysis of organizations. The temporal dimensionality deals with time and consists of an assumption about the *direction of time* and *velocity*. The factual dimensionality goes beyond the concept of *space* and also includes assumptions about *reality* and *risk*. The social dimensionality involves the way in which organizations assume their *identity*, *action*, and *trust*. In enhancements of his model, Stein (2000) proposes that any of these core assumptions can be described as polarities, and that organizations range along the spectrum of those polarities in the sense of an overall profile. I use these core assumptions to describe the polarities of the views that organizations hold of big data and the way they are being handled. Organizations also consciously position themselves within the spectrum of any of these core assumptions from which they derive strategies and operational structures. Table 11 presents the related polarities of big data within organizations.

Table 11: Polarities of Big Data in Organizations on the Basis of the Core Assumptions

Core Dimensionality	Core Assumption	Polarities	
Temporal Dimensionality	Direction of Time	Data Linearity	Data Monadology
	Velocity	Data Rigor	Data Swiftness
Factual Dimensionality	Space	Data Island	Data Assemblage
	Reality	Social Constructivism	Data Constructivism
	Risk	Data Risk Avoiding	Data Risk Seeking
Social Dimensionality	Identity	Social Shadow	Data Shadow
	Action	Self-Determined	Data-Determined
	Trust	Data Reliance	Data Bias

These core assumptions will be used as the starting point for understanding big data within organizations, as well as to refine the question of what an organization faces concerning big data. The polarities will describe the effects or outcomes of big data within an organization and not the implicit notion of integration design (e.g. Stein 2014). Organizations can be steered into a certain direction on the spectrum, with the goal to steer them in a certain direction, ideally to support the homeodynamic stability and agility of the organization.

4.1.1 Temporal Dimensionality

Time is important, especially in the context of organizations in which "time is money" (Loft 1995: 127). Big data are sending mixed signals, though. On the one hand, big data are available in an instant, on the other hand, big data are so ubiquitous that organizations are overwhelmed and need time to cope with the abundance. Dealings with big data are linked to the temporal dimensionality, and organizations need to consider the possibilities of integrating them. Big data are susceptible to changes in the temporal dimensionality and will influence future big data, as big data constantly generate new big data over time.

Data linearity or data monadology. Big data can also be seen as a temporal construct in itself. Any data within big data are linked to a timestamp, be it the time they were collected or the time of the collected incident. An example may be a historical book by a contemporary witness and a book by a researcher. The first book has a timestamp of the respective time period when something originally happened and the latter one has a timestamp of more recent years. Both books refer to the same event and will (hopefully) include similar data among the information they convey. Nevertheless, both have vastly different timestamps. Such differences

raise certain obstacles concerning big data. One way of coping with the direction of time is to see big data as a *linear* construct. Data in history are added up in a linear way and new data are constantly added to the tail of a linear stream of data. This view would decrease the complexity of big data drastically, as big data could be transformed into a timeline. In the context of organizations this seems particularly plausible due to the obsolescence of information (Argawal et al. 2005) and the half-life of knowledge (Machlup 1962). Organizations can focus on the most current data. *Data linearity* is consequently one-directional, and, therefore, obsolete information is *unlearned* or, more precisely, buried beneath new and momentous information. In times of big data and the potential danger of data avalanches (Miller 2010) under the assumption of linearity, this is a plausible concept.

An alternative perspective is seeing big data not as a linear but as a *non-linear* construct. Focusing on the non-linear perspective is similarly interesting for organizations. Big data that are relevant or related to an organization are essential to said organization. Although the flap of a butterfly's wing on the other side of the world may have an influence on an organization, the chances of occurrence are so infinitesimal that investing time and resources in order to prevent it is not efficient. Big data within organizations are, therefore, always merely just a portion of all big data and the organization itself is the one to select the relevant portion. This conceptual view resembles the idea of monadology, or the theory of monads (Tarde 1893/2012) which Latour et al. (2012: 598) describe as follows: "A monad is not part of a whole, but a point of view on all the other entities taken severally and not as a totality." Although this idea conflicts with the observer problem with big data, seeing an organization as a monad is helpful in understanding a non-linear perspective on big data. Latour (2002) argues that it is essential to move beyond a micro/macro categorization, and I propose that big data can be seen in a similar way. There is no obsolescence due to time, but there is obsolescence due to the monadological and non-linear connection. Certain elements of big data are irrelevant for certain organizations while other elements are relevant. From this perspective, organizations have a "highly specific point of view" (Latour et al. 2012: 598) on big data, and this view is decoupled from the linearity of time. Tarde (1893/2012) treats time following the argument of Leibniz, as not absolute but relative and rootless to non-linear connections. In the following example, although Giddens does not mention the connection to Tarde, time is not of relevance, the non-linear connection of the words unveils the meaning and story behind the example (Giddens 1984: 302).

> Private property : money : capital : labor contract : industrial authority
> Private property : money : educational advantage : occupational position

This example reveals the translation, or transformation of private property into something different. On the basis of the monad, however, private property is embedded into different contexts. Both transform their private property (although not exactly as defined by Giddens) into money. The first monad uses the money to gather capital, contract new labor, and achieve industrial authority, the other monad uses the money to gain an educational advantage that leads to a better

occupational position. Although the time is unknown, the non-linear progression reveals two different stories. The first monad is probably an employer while the second one appears to be an employee which shows that a *data monadology* is non-linear. Organizations select and utilize relevant big data. Following this logic, the sequence of combining big data will become more important, as will navigation through big data. Especially under the assumption that big data generate new big data, any monadological step will generate big data that depends on the non-linear perspective of the monad/organization. Presuming linearity or monadology will, therefore, shape big data within organizations.

   Data rigor or data swiftness. The next core assumption regards velocity. One major attribute of big data is that they can be analyzed quickly, data streams can potentially be analyzed in real-time (Barlow 2013). Analyzing big data in such a way comes with a certain tradeoff. Such analyses can be described as *data swiftness*, and although the results are nearly instantaneously available, those analyses may not be very precise. They are often designed without any hypotheses and favor correlations over causation. The use of big data in this particular way, therefore, is susceptible to errors. At the other end of the pole, there is *data rigor* use. Such precise and thoughtful use comes with a hefty toll on velocity. It takes time to analyze big data in that manner, although such an analysis is less prone to errors and gives more detailed insights into organizations. Such results are also evaluated and can explain causation within the data set.

Table 12: Big Data Tradeoff Concerning Velocity

		Data Swiftness	
		Low	High
Data Rigor	Low	Undesirable	High Data Swiftness
	High	High Data Rigor	(Currently) Impossible

As shown in Table 12, organizations face a decision concerning their direction of big data analyses. They can choose between high data swiftness and high data rigor. Other combinations are either impossible (high data swiftness and high data rigor) or undesirable (low data swiftness and low data rigor). Marketing methods applying the shotgun principle, being high in data swiftness, are promising; they may for example lead to an increase in sales (Mayer-Schönberger & Cukier 2013). Analyzing data from experiments like CERN, on the other hand, needs to be rigorous and will take time (Wright 2014). Organizations freely choose the way in which big data are used and will deal with the consequences. Organizations evaluate the costs they generate through velocity and the costs of the potential errors that may result from excessively rapid big data analysis.

The time perspective reveals that organizations can tackle big data in a variety of ways, however, decisions made by the organization influence the integration of big data into organizations. They can assume that big data will be linear or that it follows the logic of monadology, and decide between high data swiftness and high data rigor, but both influence the precision of the results of any big data analysis. Although such polarities apparently propose an either/or decision, organizations have the ability to apply both polarities. They can decide this before every big data analysis or even utilize both polarities, starting with high data swiftness and finishing with high data rigor. While possible, the likelihood of such an approach being taken in reality is debatable considering the increase in cost.

4.1.2 Factual Dimensionality

For this dimensionality, Luhmann (1991) uses the German term *sachlich*, which translates to 'objective' or 'factual'. Luhmann was predominantly concerned with the discourse about reality, subjectivism, and social reflexivity. Using the translation 'objective', therefore seems inadequate – especially concerning the problematic subjectivity of big data and their appearance as objective. The factual dimensionality deals with the tangible influence of big data on organizations and the ways in which an organization can use big data to transform itself. Big data provide a massive amount of information capable of influencing the factual dimensionality within organizations. Space, reality, and risk are affected by big data, but the direction of said influence depends on the underlying assumptions that organizations make concerning the comprehension of big data.

Data island or data assemblage. Stein (2000) points out the potential of technology to bridge space. In recent years the spatial distance has decreased significantly (McCann 2008). Space in the sense of spatial distance is no longer adequate for understanding the obstacles or polarities that space entails in organizations. Big data in particular contribute to the lack of spatial distance. In recent years, society has unlocked a new form of space that is parallel to its classical form. The internet has contributed to the concept of virtual space and big data help making this virtual space ubiquitous (Giard & Guitton 2016). There is a complete virtual dimension parallel to normal, or real space. Virtual and real space are not separate from each other and current developments indicate that both spaces are moving into alignment with each other (Bimber & Raskar 2005). Both worlds are permeable and people seamlessly jump from one world to the other. Navigation within real space, for example, is often accomplished using tools from virtual space. People often no longer consult paper maps; they use Google Maps. This represents an evolutionary development of society that, through augmented reality, gradually merges the two worlds (Azuma 1997). Big data are a main driver of this change (Swan 2013). Organizations, therefore, reconsider their own design as well as they picture the membrane between real space and virtual space.

It may be possible for organizations to regulate big data and strictly control their use. Using a metaphor from a spatial perspective, big data can be understood as on

an island. Only a small number of people within each organization has a boat to steer to this *data island*. Big data will be spatially far away from organizations and interaction will be limited and closely monitored. Organizations are establishing artificial distance in a figurative sense, which enables them to steer the internal effects of big data. Big data are, once more, placed inside of an iron cage while their use is regulated with an iron hand. Organizations could follow the idea if they were to assume that big data were acting as something uncontrollable and uncertain. Big data can change structures and change organizations and those developments may turn out to be structural shackles (Scholz 2015a) which may bring about the tendency to prohibit such a rampant use of big data and put big data on a data island. Such an assumption will probably tie up a big portion of an organization's resources, however, and may, therefore, be more efficient, assuming that big data help converge real space and virtual space together. At the very least, this is preferable to isolating big data within organizations.

Both spaces can be seen as permeable, and the organization as a real space will be open for interaction with virtual space that is big data. Such a concept resembles the concept of habitus (Bourdieu 1977) because "habituses are permeable and responsive to what is going on around them" (Reay 2004: 434). Bourdieu (1977) claims that habitus is both *opus operatum* (the result of practice) and *modus operandi* (the mode of practices) which is applicable to both the organizational habitus and the relationship between big data and the organization. Habitus appears as a fitting theory through which to understand the relationship between real space and virtual space, but Morrison (2005) as well as Reay (2004) note that there is a latent determinism, a focus on continuity, a neglect of change (Shilling 2004), and a strong emphasis on structures (Bourdieu 1986). A different notion of permeability between real space and virtual space links to technology as well. If both spaces are seen as equal (in analogy to Bryant, L. R. 2011), they all are actors and contribute to organizing the organizational network. Parker (1998) calls it 'cyberorganization' and combines real space and virtual space as a new form of organization. In recent years, the term 'assemblage' has gained popularity when describing this interplay between both spaces (Taylor 2009). Kitchin (2014a: 24) defines data assemblage as the "composition of many apparatuses and elements that are thoroughly entwined, and develop and mutate over time", but he sees the concept as predominantly connected to the production of data (Kitchin & Lauriault 2014). In this context, on the other hand, data assemblage is defined as an interrelationship (Giddings 2006), and a dynamic process (Taylor 2009). It is no longer possible to differentiate between real space and virtual space. Organizations permeate big data and big data permeates every organization. Generally speaking, *data assemblage* sounds a more realistic approach to big data within organizations. However, supposing such an overlapping of real space and virtual space makes it difficult for any organization to deal with big data altogether. It may even be assumed that resistance is futile (Russom 2013). To recapitulate the assumption about space, organizations decide on a spectrum between doing nothing against big data and letting it flow through the organization, or restricting and limiting the use of big data completely. Both poles are probably too

extreme. The way organizations first want to understand the relationship between organization and big data, however, will be a strategic decision.

Social constructivism or data constructivism. The next core assumption is about reality, an issue that is picked up numerous times in the course of this thesis. The conclusion of this discussion is that reality is constructed and, even though big data are vast and ubiquitous, they are incapable of representing reality as an objective truth. Big data are no Laplace daemon and will probably never be capable of being omniscient (Scholz 2015a), which results in the idea that big data create reality as well. Reality is created either on the basis of *social constructivism* or *data constructivism*. Translating this into organizational interaction with big data means that organizations will either influence big data or big data will deliver relevant information and insight. Conversely, data constructivism is the idea that big data influence organizations in such a way that the surrounding reality is shaped by big data. This development has recently become observable in the discussion about data-driven decisions (McAfee & Brynjolfsson 2012). Although for legal reasons, the decision needs to ultimately be made by a person, this person decides on the basis of the information provided by big data. Decisions are shaped by the reality constructed through big data.

Both assumptions have an impact on how organizations will work in the future. Assuming the social constructivist view, the impact will be similar to the beliefs of the neo-luddites, and following the data constructivist path would be more like the ideals of the anti-guessworker. However, both poles will inherently distort the reality of organizations, in a certain way acting as a reality distortion field (Levy 2000), and this subjective reality will be reinforced over time. Both views are, therefore, highly susceptible to objective subjectivism (Gadamer 1992). Big data within organizations force them to choose a certain path and deal with the consequences. Contrary to other assumptions, constructing reality is similar to a path-dependence (Sydow et al. 2009) and a lock-in (David 1985). A de-lock-in can facilitate changing the path, but achieving such change and negating the reality distortion field of organizations takes time.

Data risk avoiding or data risk seeking. The final core assumption concerning the factual dimensionality regards risk. Dealing with risks is essential for the survivability of any organization. Generally speaking, people and therefore also organizations can be categorized according to their risk behavior: as either avoiding risks, being neutral towards risks, or seeking risks (Kahneman & Lovallo 1993). Risk avoiding and risk seeking represent the polarities of this spectrum. Although these polarities are nothing new and big data will not add new facets to these characteristics of people (e.g. Tallon 2013), they will have an amplifying effect on both polarities. Big data can help a risk avoiding organization become extremely risk avoiding, especially since many risks can be discovered by means of big data. By calculating every potential risk in the evaluation, it is possible to avoid them altogether. Conversely, a risk seeker will have the same information, but will come to a different conclusion, likely taking the risk regardless of the information supplied by big data.

Within organizations, there is a variety of different types of risk behavior, but depending on the general attitudes to risk, big data can be shaped accordingly and even falsified accordingly. Risks may be increased or decreased by big data within organizations. Big data are, therefore, a new risk factor and these assumptions of risk and the risks which result from them need to be addressed by organizations – no matter the polarity in which they lie. I propose the concept of *big data risk governance*, which will be described in the course of this thesis.

The factual dimensionality specifies that big data will lead to a new understanding of space and, consequently, have a strong impact on the idea of space within organizations. Big data will also challenge reality, not in the promise of objectifying reality within an organization, but by being a new source of constructivist direction within organizations. Finally, big data amplify the potential for risks in organizations. Although there is an underlying assumption of polarities, organizations will mostly find a position along the spectrum and not at the extremes. Nevertheless, the dimensionalities highlight the essential need for assuming a certain understanding of big data within organizations. Simply using big data without consideration will have long-term consequences that cannot easily be repaired.

4.1.3 Social Dimensionality

The final dimensionality tackles the relationship of people with each other and the difference between the *Me* and the others. It asks the question of consensus or dissenus and the underlying morality (Stein 2000). The social dimensionality is concerned with the relationship of the individual with other actors in their surrounding organization, and big data are amongst the actors with which the individual interacts. The individual estimates the role of big data within organizations and the effects of big data on organizations, as well as the role that big data will play with regards to the individual. Such assumptions will influence the function of big data enduringly for the future. Assuming that big data will change an individual's life for either better or worse will cause the individual to act differently, and will affect the individual's identity, actions, and ultimately the trust the individual has in big data, as well as the way other people and big data perceive the individual within organizations. Big data add a new perspective to the social facets of an organization, the way in which people interact with each other through big data, and, more importantly, how they are influenced by information from big data.

Social shadow or data shadow. Identity is an important part of an individual and reveals its uniqueness. It entails a sense of self-conception and the idea of a person being different to others. But there is a potential difference between the self-perception of an identity and the way in which a person is perceived in their social surroundings. This may be the result of social stereotyping or because a person is wittingly acting in an atypical way. Identity can be assumed to be comparable to a black box, only giving insights through interaction with external environments. Neither big data nor social interaction will give a precise description of an actual

identity, never being more than shadow. While this shadow may be granular, it could also be a shadow play and be completely different to the actual identity.

Big data add a new form of shadow to the perception of identity. Haggerty and Ericson (2000) describe this new digital identity as 'data doubles'. On the basis of the Orwellian increase in surveillance that mimics a panopticon, they propose the idea that people are doubled within big data. This idea was picked up by Wolf (2010), Kitchin (2014a) and Scholz (2015a), and expanded into the concept of data shadows. Wolf claims that people cast data shadows wherever they go, and Kitchin describes those data shadows as "information about them generated by others" (2014a: 167). Scholz defines data shadows on the basis that big data are subjective and contextualized and "we are only seeing the shadow of reality (comparable to the allegory of the cave by Plato)" (2015a: 8). Even with big data, peoples' view is, therefore, limited to the shadow of the identity of others. In addition to that, the subjectivity of big data may distort the data shadow. Big data attempt to double or copy the original and exhibit in this way similarities to the concept of a simulacrum, defined as:

- "it is the reflection of a profound reality;
- it masks and denatures a profound reality;
- it masks the absence of a profound reality;
- it has no relation to any reality whatsoever: it is its own pure simulacrum" (Baudrillard 1994: 6).

Big data try to achieve a simulacrum of people by analyzing their digital footprints (Sellen et al. 2009) and data trails (Davis 2012), thus, using those breadcrumbs (Cumbley & Church 2013) to reflect the identity of that person. However, all four definitions of simulacrum are possible and make it difficult to achieve convergence between identity and data shadows. Interestingly, Baudrillard hints at the strong impact of the current explosion of data, which will lead to an implosion of meaning: "We live in a world where there is more and more information, and less and less meaning" (1994: 79). The author predominantly refers to this information exhaust in terms of the media, but he reasons that "information devours its own content. It devours communication and the social" (1994: 80) and, therefore, picking up McLuhan's (1967) formula, the medium is the message. Indeed, Baudrillard (1994) suggests that the media will create a simulacrum that simulates a hyper-reality. Hyper-reality is the sense of an inability to distinguish between reality and simulacrum (Tiffin & Nobuyoshi 2001). In the context of big data, identity is depicted, interpreted, and transformed on the basis of its data shadow, into a simulacrum that eventually creates a *hyperidentity*. People unwillingly but constantly contribute to their hyperidentity without having any control over it (Pasquale 2015). This hyperidentity can be a granular reflection of an actual identity but also has no relationship whatsoever. In some cases, for example, people are evaluated on the basis of their residential address and the behavior of other individuals in their area. If a person lives in a low-income area, gaining credit may become a challenge (Pasquale 2015) which goes to show that an individual's hyperidentity has no connection to their actual individual identities.

This form of shadow is created by big data and people have no grasp of the entirety of information contributing to their hyperidentity; however, individuals can wittingly influence their *social shadow*. As Goffman (1959) explains, people are capable of interacting with other people differently. He compares this to a theatre, where there is a difference between the identity of actors on stage and their back-stage identity. People play an act on stage and their identity is perceived mostly in regard to their acting. They put on masks or costumes and simply become different people. For that reason, within an organization that can be compared to a stage in a theater, people act in a certain way and thus create a stage identity. Such a stage identity is perceived by the audience at the risk of perceiving the actors in a slightly different way than the actors themselves (Watson 1982). The following statement describes the mismatch between individual identity and social identity quite clearly: "If one knows who one is (in a social sense), then one knows how to behave" (Thoits 1983: 175). People contribute to their stage identity while simultaneously having to juggle several stage identities at the same time (Scholz 2016a). In times of big data, playing an act in a certain stage reality will become increasingly difficult as everybody is constantly under surveillance and no longer capable of separating stages precisely. Hyperidentity and stage identity influence each other. A Facebook identity, for example, influences a professional identity and the chances of being recruited (Ramo et al. 2014). The interrelationship between both identities and the difference to the actual identities is summarized in Figure 8.

Figure 8: The Perception of Individual Identity on the Basis of Data Shadow and Social Shadow
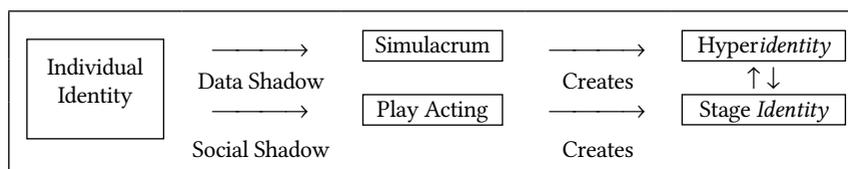
Self-determined or data-determined. The detection of a certain shadow within organizations is linked with the next core assumption of action. Because a data shadow is not much influenced by the individual or the organization, this implies a data-driven understanding of big data. If, however, the individual and organization assume a certain form of social shadow, they are capable of changing their perceived identity. Such an assumption will lead to a certain perspective on determination. The question is whether the individual and the organization are *self-determined* or *data-determined*. Self-determination in the general sense refers to the self-motivation or intrinsic motivation required in order to achieve certain goals (Ryan & Deci 2000). In that case, the individual at first has the motivation to achieve a certain goal, and afterwards uses big data as a tool to achieve said goal. Data-determination is the concept of externally motivating people to achieve a certain goal. There is much

discussion about big data nudging (Thaler & Sunstein 2008, Yeung 2016) things in a certain direction, especially in the context of big data. Richards and King (2013: 44) describe this nudging as follows: "The power of big data is, thus, the power to use information to nudge, to persuade, to influence." Big data supply the individual and organization with enough information to influence their decisions which is already done by politics (Nickerson & Rogers 2014) and governments (Schroeder 2014).

Self-determination, therefore, can also mean that big data are nudged in a certain direction. If big data reveal the desirable goals, the organization is purely data-driven. If big data are considered as a tool to achieve certain goals, however, the organization is still self-determined. Both polarities have the ability to nudge the other. However, for any strategical decision within organizations, the dominant pole is as clear as who nudges whom.

Data reliance or data bias. Finally, there is the core assumption of trust. People may believe in big data and the correctness of big data, which leads to a certain form of *data reliance*. As some suggest (e.g. Anderson 2008), if we have enough data we will get results. Such a core assumption is possible and is used (Servick 2015), but there is also the assumption of a general inherent *data bias* (Scholz 2015a). Although as stated earlier, big data rely on a certain form of data bias, the assumption of such bias will lead to the belief that big data in general are incorrect and maybe are not even used at all. Similar to the views on big data within HRM, the two views over-estimate both the objectivity and the subjectivity of big data. However, individuals in an organization will act on said assumptions and either be open or skeptical towards big data within the organization.

The social dimensionality reveals that those core assumptions will have an influence on the actual use of big data within organizations. They raise the discussion about whether decisions have become driven by big data or whether people are still capable of deciding on their own, especially in the context of identity and its perception. Do people perceive others through data or through social interaction, and how do those ways of perception differ? The social dimensionality highlights the concept of distortion through both big data and through social interaction. Assuming a certain perspective will have an influence on the other polarity and, thus, reveal the potential for nudging the other pole. Big data can nudge people in a certain direction, and people as well can likewise nudge big data in a certain direction. This development can lead to a vicious cycle. This is especially true given the difference between being self-driven or socially-driven, since being data-driven may imply that people and their decisions are controlled externally.

4.1.4 Cross-Sectional Dimensionality

Big data are adding new types of dimensionality to organizations on top of potential existing dimensionalities. Therefore, the complexity increases and this can be compared to the curse of dimensionality – a term coined by Bellman (1957). The curse explains the phenomenon that, on adding new dimensionalities of data into for example an algorithm, the algorithm will need increasingly more time to deal with

these new dimensionalities. Nevertheless, adding new dimensionalities will make the results more precise. That is the same case for the core assumptions based on the presented dimensionalities. However, they are currently separated from each other and need to be connected. Therefore, some sort of cross-sectional dimensionality is required, despite the curse dimensionality that arises.

In summary, big data are not objective and organizations are no longer comparable, and any organization can take on different points of view regarding big data. All of which being potentially fitting, however, claiming such subjectivity and uniqueness of organizations and, furthermore, proposing that big data can increase and decrease variety in an organization, may render organizations more standardized or more singular. Converging towards a certain standardized solution will lead to a loss of competitive advantage and a reduction in the survivability of those standardized organizations. A competitive organization will, therefore, prefer to become more singular rather than more standardized.

Holzkämpfer (1995) discusses singularities within organizations and focuses on extraordinary incidents within organizations and exceptional structures of organizations. Big data can be described as following his vision and will lead to such structures assuming big data have a general tendency to become more unique in order to increase the survivability of organizations.

Singularities are seen differently in related literature. There is, for example, the concept of the technological singularity. A technological singularity is the point in time at which artificial intelligence has increased technological progress to an uncontrollable and unpredictable extent (Kurzweil 2006). Although big data contribute to the manifestation of this hypothetical point in time, this thesis is concerned with singularities within dynamical systems and, therefore, rooted in systems theory, cybernetics and ultimately complex systems theory (Holzkämpfer 1995). Holzkämpfer traces the term back to Poincaré (1881) and Maxwell (1882). The latter describes the influence of singularities as follows: "it is to be expected that in phenomena of higher complexity there will be a far greater number of singularities" (Maxwell 1882: 443), thus, strengthening the relevance of singularities in highly dynamic systems. Social systems in particular are complex and dynamic. For that reason, "a small error in the former will produce an enormous error in the latter" (Poincaré 1914: 68). Holzkämpfer (1995) proposes the idea that organizations are influenced by singularities. He defines the following features of singularities:

- "Instability
- System-relatedness
- Uniqueness
- Irreversibility
- Subjectivity
- Randomness
- Complexity
- Reciprocity" (Holzkämpfer 1995: 91).

Even without describing those features, it makes sense to view an organization affected by singularities. Big data will contribute to this perspective and, furthermore, can be seen as a set of singularities as well, a phenomenon I propose calling the *big data singularity*. Describing an organization affected by big data, and regarding that big data as singularities, fortifies the claim that any organization deals with big data in a unique way. Instability is used in the sense of singularities where small things can have large consequences, and such ripple effects become obvious in big data and organizations. Due to the amount of big data, numerous minor things may have an influence on an organization. System-relatedness emphasizes that certain data are relevant to certain organizations. Not all big data are relevant to all organizations, only big data compatible with the organization's context will have an influence. Big data are contextualized and cannot be simply applied to a context they are not intended for. Organizations are unique; every organization is, in its way, truly singular. The same applies to big data; a body of data is singular as well. On the basis of the increasing granularity of big data (Kucklick 2014), every data point is unique. Data can vary in information, but the way data are predominantly collected makes it nonrecurring. Data can be collected with various tools, at distinct times, from a certain perspective, and with different intentions. While data may appear similar, metadata are always different and render every data point unique.

Any change within an organization is irreversible. More precisely, any change will create a new organization, and by trying to reverse the change, organizations will remember the recovery process. An organization is, therefore, not reversed to its former state but shifts to a new state merely resembling the old stage. The same principle applies to big data as well. Any change in big data will irreversibly transform them in a certain way. Such changes cannot be retracted, and repairing them is extremely difficult. Big data float everywhere and changes will spread through them. A certain data point can be changed in a distinct location, but whether or not the information it contains will be changed elsewhere remains unknown. Pasquale (2015) describes this phenomenon using the example of credit scores which can be changed, but not throughout all data that constitute an individual's credit score. Big data are also subjective, which imposes constraints on both big data and organizations.

The next aspect that contributes to the singularity of big data is randomness; generally speaking, those aspects that can be observed in an organization in terms of cause and effect are random. Big data have difficulties identifying causalities and often uncover correlations. They are also far from n=all, and no organization will ever have access to the totality of big data. Any selection of big data is to a certain extent random in itself. Both organizations and big data are highly complex and big data are complex. They also have a reciprocal relationship, which, while adding to their complexity, also contributes to the creation of new singularities through interaction between organizations and big data.

Creating an organization that is dynamic or even homeodynamic is plausible. Adding the concept of singularity derived from cross-sectional dimensionality to the mix allows an organization to understand and grasp big data in a more appropriate

way. Since big data are comparable to singularities, every interaction with big data is seen as something novel and distinct. Organizations will also move more in the direction of dealing with singularities from a structural perspective. Contextualizing an organization, therefore, needs to be more dynamic and relational. Dourish (2004: 22) developed a different view on contextualization: "contextuality is a relational property, the scope of contextual features is defined dynamically, […] context arises from activity". On the basis of this idea, Scholz (2013b) developed the concept of relational contextualization. The use of 'relational' in terms of contextualization may seem tautological at first glance. It does, however, underline the point that relational interaction with the context is an increasingly important factor for understanding organizations. Big data in particular become more useful if the relationships in which they are generated are transparent. Relational contextualization is essential for understanding organizational singularity, big data singularity, and, consequently, the reciprocity of the two.

4.2 Homeodynamic Organization

4.2.1 Characterizing Homeodynamic Organization

Big data and organizations constantly influence each other and are embedded into a dynamic and turbulent environment. On the basis of such an extensive relationship and that big data have a new influence on the organization, it makes sense that organizations will change. The core assumptions and the presented polarities reveal that organizations will have some freedom to react and create a unique response towards big data. Consequently, big data will trigger a transformation, but organizations will respond with a dynamic approach. The situation of the organizations can be compared to the causal texture turbulent field (Emery & Trist 1965), in which processes are dynamic and an organization is strongly interconnected with the field. This field is subjugated to linearity and non-linearity at the same time, and subsequently to order and chaos. Organizations, thus, try to achieve a certain form of homeodynamic balance in order to increase their survivability. The goal is to achieve a dynamic stability and a temporary equilibrium within the general imbalance (Luhmann 1991). In this context, however, stability does not refer to the steady state of homeostasis, but the ability to keep organizations alive. More fitting is the analogy with nautical terms: stability means keeping the boat steady or staying on a steady course. Although the ship is influenced by the environment and depends on its own integrity, the helmsman's task is to take account for all these factors and keep the ship stable.

Modern organizations are comparable to such ships, as organizations will be kept on track in order to stay profitable and, consequently, survive in today's stormy environment. Successful organizations will, therefore, act more like homeodynamic organizations that seek a homeodynamic balance. The concept of 'homeodynamics' as used in the course of this thesis was introduced by the following definition:

"Homeodynamics [involve] rate-oriented homeodynamic stability, not very far from equilibrium, fluctuating and oscillating or close to 1/f noise informationally, not fixed program-driven systems with easy generation of new activity patterns" (Trzebski 1994: 111).

Yates (1994) presents several errors concerning living systems, from which he derives the homeodynamic concept. An organization can also be seen as a living system (Kast & Rosenzweig 1972), and, following the argumentation of Yates, homeodynamics can be applied to modern organizations. One aspect of homeostasis, as criticized by Yates, is the idea that such systems are state-determined in the sense that such states influence the rate of the system. However, he makes a distinction between the two aspects homeostasis and homeodynamics in reference to non-linearity. More relevant for living systems is the rate or the velocity at which they are influenced by and influence other living systems. They have a tendency to be not very far from some form of equilibrium, and oscillate around both the equilibrium and the noise (or disorder). Living systems may be constantly changing. Yates also considers the program-driven idea of living systems. They are not predestined by their DNA and are capable of change which renders them more execution-driven. New activity patterns are constantly generated to cope with new challenges. Finally, Yates presumes that: "Systems are dynamically stable, meaning that they are able to sustain their trajectories in their basins of attraction even when coupled with dynamically rich inputs that can overwhelm structural stability" (1994: 70). In saying this, Yates moves beyond the idea that stability is linked to structure, thus agreeing with Farjoun (2010) and his concept of change as stability. Interestingly, Yates claims that homeodynamics can serve as a meta-theory (1994) for understanding living systems, but he also demonstrates that a new complexity theory needs to emerge that "must ultimately displace cybernetics, general systems theory, artificial intelligence, dissipative structure theory, information theory, and control theory from their fashionable apotheoses" (Yates 1994: 71).

When translating these ideas and concepts into a homeodynamic organization, there are several aspects that are important for an organization. First of all, any organization is changing constantly and will, over time, evolve into a new dynamic system: "homeodynamics refers to the continuous transformation of one dynamical system into another through instabilities" (Lloyd et al. 2001: 136). But organizations are subservient to attractors and will flow between them. Lloyd et al. (2001) claim that an organization tends to behave homeodynamically if there are large attractors. Big data will serve as a large attractor and, therefore, foster the tendency of organizations to become homeodynamic. An organization will be able to self-reconfigure and be dynamic enough to achieve reconfiguration in a quick and precise way.

In the realm of organizational theory, this description fits with the idea of dynamic capabilities. Teece et al. (1997) describe dynamic capabilities as the ability of an organization to use and reconfigure competences in order to deal with environmental changes. According to the dynamic capability approach, organizational competencies are achieved by smartly combining organizational resources.

One crucial goal of dynamic capabilities is that the resource configuration of organizations does not fossilize, but remains dynamic and flexible. Dynamic capabilities depend on an ongoing monitoring of resource allocation and ongoing competence-specific resource reconfiguration (Eisenhardt & Martin 2000). By being able to add new resources and release obsolete ones (Wang & Ahmed 2007), organizations stay flexible and dynamic in generating new competitive advantages (Sanchez et al. 1996). They therefore contribute to rate-oriented homeodynamic stability.

It is essential for organizations to act close to the equilibrium and close to the noise. As covered by complexity systems theory, organizations simultaneously operate at the edge of order and the edge of chaos at the same time. Big data act as a novel and potent source of disturbance. Organizations utilize this perturbation in order to gravitate around both edges. This idea exhibits consequent similarities to the concept of organizational ambidexterity (Duncan 1976: 167, Gibson & Birkinshaw 2004: 209). Organizations are surrounded by a variety of tensions (March 1991) and they need to deal with them effectively. March (1991) divides innovation into the categories of exploitation and exploration. Exploration is concerned with leveraging the potential of experimentation, seeking new ideas and generating new items (Andriopoulos & Lewis 2009). Exploitation focuses on the value-maximizing use of resources and abilities (Wadhwa & Kotha 2006). Although these categories appear to oppose each other (Lubatkin et al. 2006) and organizations currently focus on only one of them (Andriopoulos & Lewis 2009), organizations tend to move towards utilizing both aspects to become more homeodynamic. They will approach the edge of order or the edge of chaos.

In order to gravitate around the equilibrium, a homeodynamic organization is constantly fluctuating and oscillating, and, therefore, such an organization is not subjugated to path dependence and the associated lock-in (Sydow et al. 2009). Weick (1976) postulates that organizations are loosely coupled for them as well as their units (or actors, following the terminology of this thesis) to interact freely. Those actors can link with each other at any time and separate if the coupling is no longer needed. Weick (1982) focuses on flexibility, which gives organizations the ability to self-repair. He does state, however, that goals and the dissemination of information are crucial for retaining the loose coupling. Big data can be a contributing force in the diffusion of information and goals within organizations. Nevertheless, such loose coupling will keep organizations flexible and enable the actors to self-organize. In summary, a loosely coupled organization will be able to both fluctuate and oscillate and, therefore, be more homeodynamic.

Although chaos is not intrinsically bad, it can have a major influence on any organization. Too much chaos in an organization may stop it from working. Employees may stop showing up for work due to the lack of a shift schedule, resources may no longer be acquired, products no longer produced, and so on. Any organization requires at least a simple form of order. 1/f noise describes the noise that decreases with an increase of intensity of a phenomenon; there exists a correlation between the noise and the frequency of a certain phenomenon. This 1/f

noise or pink noise can be found in many other areas like physical, biological, and economic systems (Bak et al. 1987). Bak (1996:12) asked: "Why are they universal, that is, why do they pop up everywhere?" Here, the focus lies on the observation that "nonequilibrium brings order out of chaos" (Prigogine & Stengers 1984), the discourse of Gaussian and Pareto distribution, and the fact that the effect on organizations can be found somewhere else (Scholz 2013a, Scholz 2015b). Prigogine coined the term 'dissipative structure' and claims that even when an organization faces chaos, a certain form of order will emerge. Dissipative structures will reduce the chaos, thus enabling an organization to be able to sustain itself. Big data contribute to the chaos but can be used to bring order out of chaos. In order to be homeodynamic, however, organizations need to gravitate around the edge of chaos rather than moving towards order.

Yates challenged the influence of DNA on a living system and the idea that "genes act as dynamic constraints shaping product formation" (Yates 1994: 70). For an organization, the analogy with DNA could be the corporate identity (Meijs 2002), the corporate social responsibility (Visser 2011), or the corporate governance (Arjoon 2005). Nevertheless, they are often formulized, institutionalized, and subjugated to several regulations. Consequently, they are more static than dynamic. Following Yates' critique, however, the DNA of an organization may be something different and, furthermore, something more abstract. Especially as DNA follows a certain structure (Watson & Crick 1953). A more dynamic and flexible approach can be found in the string theory in which our universe can be described by only 20 numbers (Greene 2005), "and the wonderful thing is, if those numbers had any other values than the known ones, the universe, as we know it, wouldn't exist" (Greene 2005: 13:27 min). Within those constraints, everything around us has evolved out of these numbers. In a study conducted by Wang et al. (2014), students at Dartmouth were monitored through their smartphones, and the authors identified a *Dartmouth signature*. When ignoring the ethical discussion behind such an analysis, big data become obvious as means of identifying this signature and that, in fact, Wang identified an *organizational signature* (Stein et al. 2016). Organizations are not forced to follow a certain program-driven idea elaborated in their corporate identity or elsewhere, but evolve out of an organizational signature. Nokia, for example, is about *connecting people*. This signature becomes manifest in all their products, from paper to rubber shoes to cell phones. The company stays true to its claim of connecting people. Homeodynamic organizations are influenced by the organizational signature; identifying this signature will be a task for big data, but sustaining it will be a strategic one.

Organizations under homeodynamic conditions need to be able to change and find stability through change (Farjoun 2010). For that reason, enabling them to generate new activity patterns at any time is essential. Activity patterns have the potential of coordinating the functioning of organizations (Lloyd et al. 2001). Lloyd et al. explain that organizations are made up of top-down and bottom-up mechanisms, but "need only small perturbations of their parameters in order to select stable periodic outputs" (2001: 140). Although organizations are able to self-organize at the

actor-level and observe both emergent properties and patterns, they will deal with tensions both from within as well as from the outside environment. The concept of self-organized criticality describes the ability of an organization to respond to such tensions. In a homeodynamic organization, the response is quick and economical, which brings up the following difficulty:

- "Too many changes are required at the same time;
- fixing one tension makes another one worse;
- fixing tensions costs money and the firm has no extra funds to spend;
- can't effectively respond to any of them" (McKelvey 2016: 59).

While it is easy to generate new patterns within a homeodynamic organization, the challenge is to keep the organization in balance and sustain its survivability. To summarize this chapter, it can be stated that any of the aspects of homeodynamics are translated into concepts within a homeodynamic organization, as shown in Table 13. However, the organization needs to achieve a homeodynamic balance to stay competitive. Big data will increase the imbalance if left unchecked, but, if used in the right way, big data can make a contribution. The organization now holds new and powerful resources to achieve a homeodynamic balance. As a consequence, big data within organizations will be closely interlinked with the other actors, leading one step closer to a homeodynamic organization.

Table 13: Characteristics of a Homeodynamic Organization

Homeodynamics	Homeodynamic Organization
Rate-oriented homeodynamic stability	Dynamic capabilities
Not very far from equilibrium	Ambidexterity
Fluctuating and oscillating	Loosely coupling
Close to 1/f noise informationally	Dissipative structures
Not fixed program-driven systems	Organizational signature
Easy generation of new activity patterns	Self-organized criticality

Translating homeodynamics into a homeodynamic organization is influenced by big data and the core assumptions derived from the impact of big data. Consequently, big data require a more fitting characterization of homeodynamics as well as a shift from a general description of the concept of homeodynamics towards a contextualized description of it within an economic organization. Homeodynamics, combined with the dimensionalities derived from big data observed from an organizational lens, lead to homeodynamic organization as described in this chapter. While achieving a highly homeodynamic organization may theoretically be possible, it will face diminishing returns and the complexity barrier. Trying to be highly homeodynamic adds tension within organizations and will devour

resources exponentially. More actors will be involved in keeping an organization homeodynamic and keeping it balanced. That may sound contradictory, as homeodynamics are capable of gravitating and oscillating even far from the equilibrium and far from the chaos. As a result, a homeodynamic organization will be influenced by the constraints already faced by the organization, but big data allow organizations to infuse themselves with new variety and new tools to become more homeodynamic. Big data will not do this on their own, as they are subjective and also tend to standardize. The contribution of big data to the homeodynamic balance depends on the interrelationships between big data (resources) and human (resources) within the organization.

4.2.2 New Roles of the Human Resource Department

Big data will lead to a homeodynamic organization as described earlier and the change can be explained through the implications of the core assumptions of big data. Such a shift in organizational understanding will require a certain reaction from within the organization. Some function will change its role accordingly to tackle the new homeodynamic organizational environment that was triggered by big data and the relation between big data and the people within the organization. One department that already deals with the management of people and, consequently, is involved in change management is the HR department. If big data are seen as a social phenomenon, then the HR department is even more predestined to deal with big data within the organization. However, due to the substantial changes in moving towards homeodynamic organization, the HR department will, at first, react to these changes with the development of new roles for the HR department relating to big data.

The HR function is often the subject of discussion, and many researchers (e.g. Cappelli 2015, Charan et al. 2015, Stone et al. 2015) discuss the role of HRM in the future of organizations. In recent years, the research by Ulrich et al. (2013) is often suggested as a clear picture of how HRM needs to change and what competencies are necessary for HRM to be able to deal with the ever-changing new environment. As Ulrich et al. (2013: 457) state: "HR professionals have often been plagued by self-doubt, repeatedly re-exploring HR's role, value, and competencies", when facing the massive transformation. Technology is seen as a catalyst for the change of HRM function (e.g. Parry 2014), and big data are changing organizations fundamentally, however, somebody will stand up to fill the evolving gap. Big data can be seen as a purely technological phenomenon, but big data will have a stronger impact at the social level and the people within organizations. Consequently, HRM will have the chance to heed the call of big data at these times, and focus on people.

Table 14: New Roles for HR Department

New Roles for the HR Department (Ulrich et al., 2013)	Big Data Specific Roles for the HR Department	Cross-Sectional Role for the HR Department
Strategic Positioner	HR Konstruktor	Big Data Watchdog
Credible Activist	Canon Keeper	
Capability Builder	Theorycrafter	
Change Champion	Built-In Schumpeter	
Human Resource Innovator and Integrator	Data Maker	
Technology Proponent	Data Geeks	

In Table 14, the six new roles suggested by Ulrich et al. (2013) are shown, as are the roles for a HR department concerning big data within an organization. They are following the logic of the roles required, however, with a distinct focus on the special situation of HRM regarding big data. Following the role theory (Mead 1934) it becomes evident that these roles require unique and differentiated positions (Levy 1952), relations within organizations (Parsons 1951), characteristics (Biddle 2013), certain behavior (Linton 1936), a subset of social norms (Bates & Cloyd 1956), and "activities which in combination produce the organizational output" (Katz & Kahn 1966: 179). All of the big data specific roles incorporate this role logic and are derived in order to tackle a certain gap within the homeodynamic organization which aises with big data.

The role of the big data watchdog, however, will be cross-sectional, and as a unification of the HR department with all roles as well as throughout organizations. The big data watchdog is on a higher order for the HR department and, therefore, acts as a guiding system. Thereby it influences the six roles of the HR department for dealing with big data within organizations. Such a role provides the basis for any further changes and modifications caused by the transformation towards a homeodynamic organization.

4.2.2.1 Big Data Specific Roles

In the following, I will briefly describe the six roles on the basis of Ulrich et al. (2013), and afterwards explain the specific characteristics in the context of big data.

HR konstruktor. Strategic positioners focus on the understanding and knowledge of doing business, they need to learn the language of business, contribute to organizational strategy, understand the needs of all stakeholders, and have an intensive knowledge of the business environment of their organizations. Big data will change the HR department in a similar way. The HR department currently focuses on the human role within the organization, however, due to elements like

110

digitization, automation, gamification, and above all big data, the job of the HR department is becoming more technological. Many operational tasks are nowadays performed by software solutions and this will increase in the future. HRM is at a crossroads, as it has space and time to do strategic work and act as designer, creator, networker, and watchdog of the working world within an organization. It can contribute to the strategy of an organization and understand the relational network in which organizations are embedded. Lem (2013) describes such a multifaceted role as "Konstruktor." The role of the *HR konstruktor* is shifting within organizations, to become an integral function that not only looks after the employees but also connects human and machine. This may be a stretch for current HR departments, but it may be part of their survival strategy (Cappelli 2015). If operational tasks are automated and strategic decisions, due to technological complexity, are made by IT, or by quants (Davenport 2013), or data scientists (Davenport 2014), the question arises: What is the necessity of HR? The answer is still the same – to deal with people-related issues – but the embodiment is changing. The HR department needs to learn and understand big data in order to contribute to organizational strategy and have a close look at the stakeholders.

Canon keeper. The next role, of the *credible activist*, is about the credibility of HR professionals and how they build personal relationships and trust. They have a clear message, are trying to improve their integrity, and are experts in business activities. They are also self-aware about their role within organizations. Using big data extensively supports a building of trust within organizations and maintaining this trust over time (Rousseau et al. 1998). The HR department will, therefore, act as a *canon keeper*. Contrary to big data curation, the goal is to become, be, and remain credible about big data use and generate trust concerning big data and the use of big data within organizations. How is big data utilized and in which way? This role is predominantly about showing the actors within organizations that big data are used in a meaningful and positive way by means of communication. Trust can be generated by upholding the integrity and the consistency of big data within organizations. Such a process has similarities to the upholding of a literary canon, as described: "The official canon, however, is sometimes spoken of as pretty stable, if not 'totally coherent'" (Fowler 1979: 98). Part of this canon are "the events presented in the media source that provide the universe, setting, and characters" (Hellekson & Busse 2006: 10). In recent times, canons are part of culture, and emerged as a popular term in, for example, the acquisition of Star Wars: Disney evaluated all stories about Star Wars and categorized them into canonical and non-canonical. Those responsible for this task are called 'continuity cops' or 'keepers' (Baker 2008). Questions about orderliness, story integrity, continuity, internal consistency, and overall coherence are the tasks of such keepers: there are massive amounts of information that need to be integrated, ordered, and made consistent. Big data need to fit with the canons of organizations. Consequently, the HR department will deal with the canonical fit of big data and, in this way, achieve a trustworthy utilization of big data within organizations.

Theorycrafter. The third role of Ulrich et al. (2013) is the *capability builder* in which individual abilities are transformed into organizational abilities. These abilities relate to the strengths of the individual and, consequently, the strengths of organizations, and, therefore, they will influence the organizational culture and identity, or in the terminology of this thesis, the organizational signature. They are, however, concealed within big data when they comprise big data themselves. The HR department, thus, will discover those hidden capabilities at the individual and organizational level. If they are hidden, the HR department crunches the numbers and analyzes the data to discover those capabilities. They act as *theorycrafters.* This term is derived from video games, and describes the search for the optimal strategy within a game on the basis of mathematical and statistical analysis (Paul 2011). Theorycrafters establish simulations that try to mimic the video game and test different constellations on the basis of thousands of iterations. However, it goes beyond the idea of crunching numbers, and is the synthesis of big data analysis and practical experience which can derive usable results. In a podcast from the Training Dummies (2016), the creators discussed the topic of theorycraft in detail and emphasized that it is a combination of theorizing and experience in order to adjust simulation to reality. Some elements are difficult to simulate; others are pretty accurate. However, applying certain metrics to all situations will lead to distorted results, and, therefore, the theorycrafters understand the situation they want to simulate. In the context of abilities, theorycrafters are able to crunch the numbers and apply them to the contextual situation within organizations. Simply mining the data will not be sufficient to identify hidden capabilities; the HR department needs to understand the data, differentiating between signal or noise, and making sense out of the simulations. Finally, the theorycrafter translates the results into action for organizations, and thus can influence the pool of organizational capabilities.

Built-in Schumpeter. The next role is the *change champion.* The HR department supports internal abilities to change, by helping to identify emergent transformations and helping to overcome resistance, and sustain the change ability within the organization. Similar to the tendency to seek a stabilized environment, big data have the tendency to converge, become homogeneous, and favor the mean, though this would lead to statistical errors having a stronger impact (Spiegelhalter 2014), being reinforced over time and becoming difficult to change. The HR department, thus, needs to include a role for stirring up big data within organizations. This role can be called the built-in Schumpeter (Scholz & Reichstein 2015) and people fulfilling this role are trying to continuously conduct creative destruction (Schumpeter 1942). Status quo means deadlock (Farjoun 2010) and is not preferable, however, an HR professional will not destroy in order to destroy, but will have the goal of improving the organization. Big data will be a tool to help the built-in Schumpeter and make organizations more capable of change. The goal is to create alternatives and variety within organizations and within big data. The built-in Schumpeters will evaluate and improve big data use and reconfigure the related investments within organizations. That may be within their own HR department, the HR daemon, or the HR centaur.

Data maker. Another role is the *human resource innovator and integrator*. HR professionals require an in-depth knowledge of HR and acquire knowledge about new trends and new solutions. They are able to translate this knowledge into solutions within organizations, however, Ulrich et al. (2013) emphasize that the HR department focuses on the long-term effects and not on achieving short-term success. In the context of big data, the HR department acts as a type of *data maker*. The term is in analogy with the maker movement (Dougherty 2012) and describes the potential of people to create everything on their own (Stam et al. 2014) and, thus, to create new ideas and new innovations (Lindtner 2014). Big data within organizations depend on such an approach and the ability to think outside the box. Big data will not do such thinking and the HR department will seek out those new ideas with the help of employees, as in hackerspace (Guthrie 2014) or hackathons (Briscoe & Mulligan 2014), and will design a way to integrate such novel uses of big data into organizations.

Data geeks. Finally, there is the *technology proponent*. Technology has increased drastically in recent years and the work of HR is also subject to an increase in technology. Many operative tasks are automatized, other functions are digitized and big data emerges as having an increasing influence on the HR department. Although there is currently an HR-IT barrier, the HR department is driven to overcome it and be open to a more technology-focused HR function. The HR department, therefore, deals with big data, or somebody else will annex this task. This requires some form of cultural change, however, from refusing big data to becoming *data geeks* (Priestly 2015). Although data geeks follow a skeptical approach, they have an interest in utilizing big data in a way that is helpful to employees and organizations. They seek new ways and innovative ideas to analyze the available data and are always looking for new sources of data. Still, their work is within the constraints of the big data watchdog although the HR department is proactively opening up to big data and eradicating the current HR-IT or HR-Big-Data barrier.

In summary, the HR department is facing a big challenge, but it needs to take charge of big data. Big data are not another tool that is delivered or supplied by the IT department or an external business partner. Big data are a critical resource for organizations and will have a strong impact on the work of the employees. Applying big data in the way described will enable the HR department to discover the hidden potential of their employees and generate a competitive advantage for organizations. Big data allow the HR department room to focus on the strategic perspective of improving and helping employees to improve. Although there are self-doubts and people are constantly re-exploring the HR role (Ulrich et al. 2013), big data offer an opportunity to assume a strategic and integral role within organizations and influence their survivability.

Milan Lab

In order to understand the new role of HRM, it is useful to look into sports again. Davenport describes the interest of HR professionals in sports as the following: "Still, sports managers – like business leaders – are rarely fact-or-feeling purists" (2006: 102). I have talked in this thesis about Oakland Athletics and FC Midtjylland, but there are many other sport teams, for example the football clubs TSG 1899 Hoffenheim in Germany and Bolton Wanderers in England, that highlight the extensive use of data. One example in particular that seems strikingly fitting is the Milan Lab of the Italian football club AC Milan. The club is using modern technology extensively to improve health quality and "predict the possibility of injuries" (Kuper 2008). Interestingly, Meersseman, a former director, compares the lab to a car dashboard and players to drivers: "There are excellent drivers, […] but if you have your dashboard, it just makes it easier" (Meersseman in Kuper, 2008). Big data support the work of a coach and their staff.

There seems to be a focus on body health issues and a focus on data-driven decisions, however, the Milan Lab tries to improve the soul of the players as well. For example, if they have had traumatic experiences, like the brutal injury of Schewtschenko, the Milan Lab and the staff help the players be able to deal with their fear (Biermann 2007). Although it is difficult to quantify the effect of the Milan Lab, it seems that there is a positive effect. Players are able to compete at an international level at older ages (Newman 2015a). The team won the Champion's League in 2007 with an average team age of 30 years, and Paolo Mandini, the captain of the team, was 38 years old (Transfermarkt n.d.). This is interesting in times of a general 'youthism' in football (Grossmann et al. 2015). Big data change the role of the coaches and the staff.

Big data are a source with which to improve work, however, they do not make work magically better, and people are still essential. This can be seen in sports, and will be seen in organizations as well. Currently, many are praising the potential of predictive policing (Beck & McCue 2009), but the advantage of predictive policing is not that crimes are discovered by algorithms, but that the police can do their work faster, more systematically, and more efficiently (Peteranderl 2016). Again the role of the police officer has changed.

This will be similar to the use of big data within organizations. Big data will enable people to become more efficient, but it is the HR department that makes big data and the people more capable of dealing with each other. The role of the HR department will drastically change, however; it will be responsible for exploring new potentials and new ideas for the use of big data. Big data do not magically make organizations better places, and people are still greatly involved. This unique use of big data will be a competitive advantage, and such a unique use comes from

the people involved. The Milan lab will not share their information (Newman 2015a), and FC Midtjylland has no interest in sharing their secrets either (Biermann 2015). Competitive advantage is created by people and not by big data, consequently any organization requires a unique way of using big data and not buying 'off the shelf' tools from some external provider.

4.2.2.2 Big Data Watchdog as Cross-Sectional Role

Those different roles are dealing with many facets of big data within organizations, however, they can be seen as relatively separated from each other. It is essential to have some sort of cross-sectional role for the HR department since using big data within organizations leads to complex interplay and interaction; big data needs to be supervised within organizations. Due to the contextualization of an organization in particular, and, subsequently, the organizational signature, only a portion of big data is useable for organizations. Big data will also have an influential impact on organizations, therefore, big data are closely supervised and organizations are capable of dealing with big data. Therefore, the HR department will be authorized to watch over big data.

Such a role is comparable to that of a watchdog. The term "watchdog" is currently being discussed in pop culture, largely due to the video game "Watch Dogs" (issued by Ubisoft in 2014). The game is located in a futuristic Chicago which is under total surveillance that is misused by the developer (a company) and the users (the city and the police). The protagonist also has full access to personal data of all inhabitants and acts as a watchdog and vigilante. This mirrors the recent discussion of big data and the role of the NSA (Gallagher 2013) and, therefore, reflects the current zeitgeist, however, such a watchdog is not only essential at the societal level, but especially at the organizational level.

Merriam-Webster's dictionary defines a watchdog as "a person or organization that makes sure that companies, governments, etc., are not doing anything illegal or wrong". It is often used in the context of investigative journalism, where the journalist acts as corrective (Miller 2006, Rensberger 2009), as well as other non-profit organizations that act in a similar way (Rao 1998). In recent times, and especially since the case of Edward Snowden, a watchdog has been compared to a whistleblower (McLain & Keenan 1999), however, external whistleblowing is seen as a last resort (Miceli & Near 1994), while a watchdog intervenes at an earlier and still changeable phase, fulfilling an important control function. Being a watchdog is, therefore, not just protecting, but also guiding, and acting in general as a sort of corrective within an organization.

As big data are always connected to humans, it may seem obvious to give the role of watchdog to HRM. In order to provide justification for the proposal that the HR department is an appropriate office to handle this watchdog responsibility, the functional role of HRM will be highlighted. Modern HR departments already go beyond the stereotypical "hiring and firing" and focus on employability. In times of

increasing employee participation (Busck et al. 2010) and its visibility in employer branding (Wilden et al. 2010), the HR department is dedicated to the task of acting in the interests of employer and employee at the same time. It is a "first and second party" rather than a "third party" and, therefore, has internalized the ethics of both the company and employees. In fact, if the goal is to help improve the performance of employees and the relationship between employer and employee, a department responsible for workplace training is a good fit for this watchdog role, and the HR department usually covers this responsibility. This description also applies to their unique and differentiated position within organizations.

Specialists in employment law can be found in the HR department, being able to resolve legal questions. Since the use of big data is a multifaceted issue, a consortium of employees from various departments and representatives from trade unions and work councils could be integrated in a steering committee headed by the HR department. Although a suitable legal landscape is missing at the moment, or at least is underspecified in many cases, the HR department can base its decisions on its broad experience with other sensitive and more deeply regulated issues such as diversity (Zanoni & Janssens 2004). In analogy to this, the HR department has already exercised restraint in going beyond the boundaries of "good corporate governance" (Fauver & Fuerst 2008: 673). HR departments are following a certain subset of social norms and behave in a distinct way.

Another element is that the modern and digitally competent HR department is able to provide its professionalism regarding big data handling. This role includes ensuring that the analytics of the big data are correct, unbiased, and not taken out of context (Kitchin 2014a). It also includes ensuring the appropriate retention of data, making sure that the data are used legally and will remain internal and secured, and maintaining transparency in data collection and accessibility for appropriate parties. On the one hand, the HR department is bound to secrecy and, therefore, required to keep internal information internal. On the other hand, the HR department needs to bring together big data experts, increasing their expertise in IT, the law, and analytic skills, thus training them to handle this responsibility. With such a diverse set of characteristics and with a unique position within organizations, the HR department contributes to organizational output.

These elements emphasize the claim that HRM could act as a big data watchdog; however, they also highlight the complex situation of big data use within an organization and the essential need to keep big data within organizations. Such a watchdog needs to deal reactively with the impact of big data on an organization from a social and ethical perspective. This role is highlighted in most of the current criticism concerning big data (e.g. Boyd & Crawford 2012, Kitchin 2014a).

Target Baby versus One Family Baby

In order to explain the uniqueness of the big data watchdog as well as the need for such a role, the following examples will present two different approaches to dealing with information derived from big data. In 2012 a debate erupted due to an incident involving Target (Hill 2012). The company has extensive information about their customers and their buying history, such as baby-shower wish-lists that can be organized through Target meaning that Target knows which customers are pregnant. Using these data sets, they derived a pregnancy prediction score. One sign used to identify pregnancy among their customers is the following:

> "Many shoppers purchase soap and cotton balls, but when someone suddenly starts buying lots of scent-free soap and extra-big bags of cotton balls, in addition to hand sanitizers and washcloths, it signals they could be getting close to their delivery date" (Duhigg 2012).

Target utilized this information and sent the customers they believed to be pregnant coupons for baby clothes. Hill (2012) described a case in which a customer showed such a coupon to the Target manager addressed to his daughter. Although the manager apologized, it turned out that the daughter was pregnant, but hadn't told her father. Remember this is a real case, in which Target discovered a pregnancy before a close family member, and that actually it was through a relatively simple method (Ellenberg 2014). From a marketing perspective, it makes sense to use the information, but from the corporate social responsibility perspective, it may harm the company's credibility if customers perceive it as unethical (Schramm-Klein et al. 2016).

A similar example involves the episode "Connection Lost" of One Family (Season 6 Episode 16). This time, the mother tried to talk to her daughter after a fight couple of days ago and wanted to know what was wrong with her daughter. Eventually, the mother was stuck at an airport and only had access to the internet, so she phoned relatives by using a video messaging platform, viewed the Facebook feed of her daughter, and accessed her daughter's iCloud by hacking her password (it was an easy password). A package from Amazon arrived at the home and was opened by the father. She gathered the following information. Her daughter recently married (Facebook), had posted several pictures with a male friend (Facebook), is at the moment in Las Vegas (Find my IPhone), and had ordered baby books (Amazon). This all suggested that the daughter was pregnant and had eloped, however, at the very end, the daughter had just fallen asleep and there was nothing to worry about. The daughter changed her Facebook status as a joke and had met with a friend who borrowed her phone and the books were for her boss. The daughter screamed afterwards: "Borders, Mom!"

Although this example is about families, organizations have likewise the ability to access a similar amount of social data. In times of the dissolution of labor, especially, there is no difficulty for organizations to access such data. Are there changes in behavior, does someone stay longer at work, how much coffee do they drink, how much do they talk with others? Is it ethical for an employer to check the social media profiles of employees if they call in sick? In a survey by Jobvite (Singer 2015), they discovered that only four percent of recruiters do *not* use social media. Consequently, recruiters will screen social media profiles and potential employees will clean up their profiles (Brown & Vaughn 2011). But the overarching question remains: Is it ethical to use such information?

Using big data will always have an ethical dimension and the answer to the ethical question will vary from organization to organization and with their surrounding network and stakeholders. It is essential to deal with this question carefully. Big data are volatile and new ways of generating big data are emerging; consequently organizations need to be watchful and vigilant about new developments. The aim for any organization is to find a fitting answer for their organization. Some will find what Target did acceptable, and many would be outraged by the action of the mother, but both examples used big data to gather information.

4.2.3 Human Resource Daemon

The previous chapter described the reaction of the HR department to the transformation towards homeodynamic organization triggered by big data. However, this is only the first step in dealing with big data within an organization. Only reacting towards big data will not be sufficient, consequently, the HR department will establish new structures in order to use big data in a valuable way for the organization. The HR department now has the competence to deal with big data and, furthermore, there is the need to create an environment for big data within organizations, as big data potentially have a life of their own which would cause a loss of control in the organization. Big data, however, influence the organizational network and all actors within organizations. Such an independent existence of big data would put them into a black box and organizations would treat big data as an external influence. In order to utilize big data to the fullest and use them for homeodynamic balance, organizations need to understand organizational big data and integrate big data into the organizational network.

Letting big data roam freely through organizations is not an alternative path, as big data will overburden individual actors within organizations. It is, therefore, necessary for an institution to deal with big data and be accountable for big data within organizations. Big data are driven by technology and, currently, big data are often deliberately used under a veil of ignorance. Although there may be reasons for putting big data into a black box, they need to be contextualized within an organization. This understanding can only be achieved by adding the social perspective to

118

big data and not solely focusing on technology. Big data are also not objective and so in order to use them in an organizational context, they require transformation. Somebody will also utilize, distribute, and train the employees. Big data within organizations are, therefore, every employee's business. I propose that the human resource department is the most suitable candidate for dealing with big data within organizations. Although HR departments probably lack big data knowledge and the competencies to handle them, they are capable of dealing with people. This seems to be exactly what is needed:

> "The challenge is not just a technological one: the selection, control, validation, ownership and use of data in society is closely related to social, ethical, philosophical and economic values of society" (Child et al. 2014: 818).

It sounds somewhat paradoxical, but big data emphasize the "people question" within organizations. Big data require people within an organization and will not replace them. Using the Moneyball example, Silver (2012) describes the time after the Moneyball incident as a fusion between worlds: that statistics and scouting might work together and mutually achieve more than using either data or people. He claims that this synthesis helped the Boston Red Sox to win their first championship title in 86 years. There are various other examples, one of which being chess (Kelly 2014, Ford 2015), that reveal the advantages of this mixed usage. It is therefore no longer true that data will create an competitive advantage as everybody has access to them; in a world driven by data, people will make the difference. Those people need help from an HR department that moves beyond operational tasks and works towards becoming a strategic partner (Charan et al. 2015) and the *social* big data expert capable of handling big data.

In order to deal with big data, it is essential to integrate them into organizations. This implementation does not lead to a complete transparency of big data which would overburden employees, but enables the HR department to utilize big data to their fullest capacity. Most current software, and many programs and applications, are separated into a backend and a frontend. The backend is the system that operates in the background and is invisible to the user. The frontend is the user interface (UI), which is visible. A user in this case is an employee in the organization. The HR department does the heavy lifting of big data in the backend, and designs a frontend for the employees. A backend is sometimes called a 'daemon' which is why I propose a system called the human resource daemon in analogy to the Laplace daemon. Although the HR daemon will not provide all answers, the goal is to outline a hypothetical organizational daemon that is capable of giving solutions to all questions within organizations.

Such a daemon and, subsequently, the HR department, will deal with three aspects of big data. Firstly, big data will be generated. I use the term 'generate', as big data are always modified in one way or another when entering organizations. At least, any external big data are labeled 'external' and any internal big data are labeled 'internal'. This differentiation will be judged in a certain way and will influence organizations in different ways. Like any analysis, the HR daemon constantly

generates new big data all the time. Secondly, big data will be evaluated. This evaluation already starts in understanding the source of the information. Information, for example, has a half-life and using old information comes at a risk. Another concept that deals with evaluation is the categorization of high data swiftness and high data rigor. Finally, there is the aspect of monitoring. Big data will influence social interaction within organizations and have an impact on the organization. The HR department establishes structures to watch over the influence of big data on the organization. Generating big data will be conceptualized in the *data farm*, evaluating big data in the *fog of big data*, and integrating big data in the *big data immersion*.

4.2.3.1 Data Farm

Big data construct a certain form of reality, be it influenced by a social constructivism or on its own through a data constructivism. Seen from a temporal perspective, however, this reality will be reinforced over time. This type of reality is fortified, and big data act self-referentially towards the acceptance of such a reality. Big data are trapped within a self-reinforcing circle which leads to a risk of uniformity, an increase in homogeneity, similarity, and convergence. This could be beneficial if big data were to be objective, not in the light of their subjectivity. Consequently, big data distort reality towards one potential and subjective reality. This tendency to reduce variety and, therefore, reduce big data, is explained in the following example concerning maps:

> "A good map eliminates as much spurious information as possible, so that what remains is just enough to guide our way. Moreover, when the map is well made we gain a deeper understanding of the world around us. We begin to recognize that rivers flow in certain directions, towns are not randomly placed, economic and political systems are tied to geography, and so on" (Miller, J. H. 2015: 1).

Although it makes sense to eliminate spurious information, the problem with big data is that spurious and relevant information is indistinguishable. Navigating by means of a map is a simplistic goal and unnecessary information is easily singled out (e.g. Miller, J. H. 2015), but achieving homeodynamic balance within a turbulent field is an obviously abstract goal and relevant information can become dynamically irrelevant and vice versa. It is, thus, essential to gather as much information as possible. Assuming big data have a tendency to destroy variety, the HR department creates it.

The concept of the data farm aims at creating variety in order to generate more big data and, most importantly *more diverse* big data. In today's age of technological advancement, storing huge amounts of data has become very affordable (Murthy & Bowman 2014). In the context of an organization, only a small portion of that big data is important. The first way of generating big data involves using a variety of algorithms. If several algorithms are available, all are applied. There is no objective explanation of why one algorithm may be superior to another; some have an inherent ideology (Mager 2012), and others seem to be correct even though the

programmer does not understand why (LaFrance 2015). To ensure the potential of making a decision, a selection of big data is required.

Big data are also learning through machines, and people learn as well, so the data farm will learn and will use new and different algorithms. Although there are often good reasons for choosing a certain path at a bifurcation, under homeodynamic conditions, it may be a total evolutionary dead end. The next task for the data farm is, thus, to archive evolution (Scholz 2016b). It is possible that an earlier evolution of the data farm may be more accurate regarding newer changes in the environment. Above all, it has become more evident in recent research in organizational theory that history matters (Sydow et al. 2009). Remembering history helps understand recurring patterns (Turchin 2008) and can be used to increase the ability to predict events (Spinney 2012). Although a precise prediction of the future may be impossible, any organization has the tendency to tackle situations in a certain way, depending on their organizational signature. Such a data farm is shown in Figure 9. Every evolution and every algorithm creates a new data stream and a self-consistent form of reality. This can be compared to the idea of the multiverse (Deutsch 2002) according to which there is an infinite number of universes, each of which differs slightly from all others. The authors of the science-fiction work "Long Earth" (Pratchett & Baxter 2012), propose the possibility of there being an infinite number of different Earths. Human characters in the book are capable of switching between these Earths. Some multiverses are only marginally different, others (the authors call them jokers) vary drastically. This can be seen as a parallel to the data farm and accentuate the fact that deviations of any kind will have an influence.

Figure 9: Evolution of Data Streams within the Data Farm
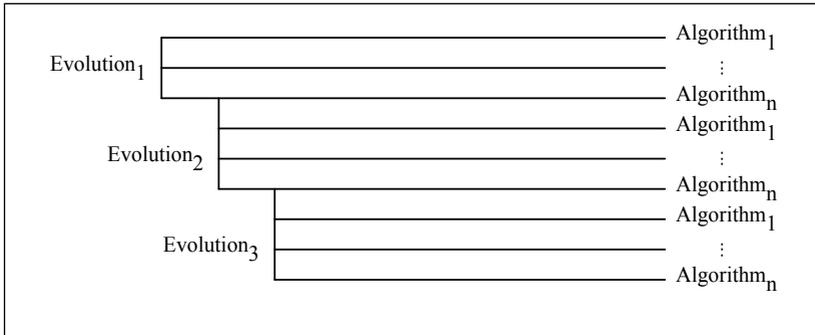
The general goal of a data farm is to increase the variety of big data within organizations, thereby counteracting the tendency of big data to destroy variety. Such an increase in big data can add to their overall preciseness. The HR department can paint a more granular picture of available information. This is interesting, as there is a concept in cryptography and collective intelligence which states that "no

information is information" (Grimson 1980: 114). The absence of certain information will tell a story, and knowing that all information is utilized will increase the story.

Finally, the data farm adds a certain scalability to big data analysis within organizations. It is a tool that everybody can use. Every new analysis, however, is added as a new data stream to the data farm and acts as a data mutation or a 'joker'. Those new data streams are highly contextualized and include all the relevant metadata required to understand what properties have changed. This form of scalability is essential for the fog of big data. Nevertheless, it is important to highlight that the data farm remembers any form of data mutation, and can use this for all big data analyses if needed or required.

Psychohistory

When talking about big data and predictive analytics, people (e.g. Turchin 2012) often cite Asimov's "Foundation" (1951) and take it as an example that everything is predictable. That is, however, too overenthusiastic, and in the course of the book Asimov reveals that predicting the future is a very complicated task. At the demise of a Galactic Empire, Hari Seldon, the creator of psychohistory, created two Foundations and established the Seldon Plan.

> "The Seldon plan is neither complete nor correct. Instead, it is merely the best that could be done at the time. Over a dozen generations of men have pored over these equations, worked at them, taken them apart to the last decimal place, and put them together again" (Asimov 2010: 497).

It seems impossible to predict the future. By predicting the future in this particular way, however, Seldon influences the future, as the one known Foundation is an enclave of knowledge for physical sciences. Similar to the uncertainty principle (Heisenberg 1927), people will attempt to make this *particular* future a reality which is the task of the second and hidden foundation consisting of social scientists. Their task is to manipulate society to follow the Seldon plan.

In order to keep society on track, this second foundation adds new amendments to the plan and influences society in a certain way. They do, however, calculate for all eventualities and especially for all unknown unknowns (Pawson et al. 2011) as shown in the character called Mule in the book. The second foundation derives a variety of amendments and plans, and administers numerous changes to bring society back on track. The assumption is that it is the social scientists who utilize the Seldon Plan, proactively trying to make it happen, and that the physical scientists subordinate themselves under the Seldon plan. Similar behavior can be observed in the use of big data (e.g. Lange 2002).

This example illustrates the importance of both keeping an eye on everything, and increasing the variety of data streams, enabling any organization to apply changes to the system. Psychohistory is linked to the field of cliodynamics (Finley 2013). This field tries to analyze history quantitatively in order to discover patterns that can be used to predict the future. Although Turchin (2008) limits the possibility of using history or any other way of predicting the future due to "mathematical chaos, free will and self-defeating prophecy" (2008: 35), cliodynamics can be used to learn lessons and discover empirical regularities (Turchin 2012) or remember the mistakes of the past (Scholz 2016b). Psychohistory and cliodynamics depend on big data, however, and depend on somebody separate to curate the different data streams.

4.2.3.2 Fog of Big Data

One issue that is crucial for the use of big data is their possible incorrectness. Information from big data can be outdated, collected for a different purpose, be tampered with, be incomplete or fragmented, or be faulty due to measurement errors or errors in communication (be they technical or due to human communication). Using big data comes at a risk due to the uncertainty of the value of information generated. In military terms, and within current video games, such strategies for dealing with uncertainty are called the "fog of war". The term was first used by von Clausewitz, using the terminology of *Nebel des Krieges* and he describes it as follows:

> "War is the realm of uncertainty; three quarters of the factors on which action in war is based are wrapped in a fog of greater or lesser uncertainty. A sensitive and discriminating judgment is called for; a skilled intelligence to scent out the truth" (von Clausewitz 1832/1976: 101).

The concept of fog of war is analogous to the use of big data which is why I propose establishing the concept of *fog of big data*. Big data are not seen as a reliable source without sufficient information about them. Metadata will, therefore, become increasingly important for the HR department, as well as in the social media appearance of employees. Obviously the actions of employees on these platforms say something about them. It is common knowledge, however, that companies monitor people on social media, and so potential employees clean their social media profile. Regardless of legal questions raised by this social media research (Hoeren 2014), recent developments in research question the reliability of data collected this way (Brown & Vaughn 2011). People adapt to the use of big data within organizations in a certain way and so organizations will deal with the fog of big data and "scent out the truth", as von Clausewitz (1832/1976: 101) explains.

Dealing with the fog is a strategical task and requires an active use of resources. As in military strategy, the HR department needs to actively scout for reliable information and evaluate existing information. There is also a need to evaluate the tradeoff of having many risky data points and only a few precise ones. The HR

daemon, therefore, requires a range of tools to deal with big data. One tool is to give organizations the ability to identify faulty data by means of *big data baloney detection*. Another tool serves the purpose of simulating all potential outcomes of thinkable and *un*thinkable strategies, as described in the concept of *big data tinkering*. Using the first tool densifies the fog of big data as it depends on rigorous analysis. Only a smart portion of big data will be visible, but the picture will be clear. The second tool will cause only a light fog of big data as it depends on outside-the-box thinking. The picture will be fuzzy, however.

The fog of big data reveals the potential of big data within organizations, and also the uncertainty and risks concerning big data. They will be evaluated from all perspectives which is why the comparison to the fog of war seems fitting. Somebody derives a plan and all available information needs to be evaluated and calculated. Power and knowledge are not purely derived from information, but from its translation into action. The fog of big data strengthens the previous claim that the competitive advantage will be found within the human actors. Algorithms are bound by their rationality and their rules, then the human factor (Zuboff 2014) adds irrationality and diversity into the mix. Both big data and people, therefore, represent sources of risk, but big data will be shackled within people's subjective reality, and will end up as a fog of big data with which people will retain the ability to interact dynamically.

4.2.3.2.1 Big Data Baloney Detection

We are surrounded by data and currently big data are being put into a black box and perceived as something magical. This observation, as stated earlier, puts organizations into a difficult position. Sentences like "the data clearly states" or "there is a significant correlation" are common and emit confidence, maybe even faith in big data (Boyd & Crawford 2012). Nevertheless, big data do not increase the precision of data analysis; on the contrary, big data increase the veil of ignorance (Rawls 1971) and people's trust in numbers (Porter 1996). Similar to the claims that big data eradicate theory (Anderson 2008) and that, they are, in fact, strongly theory-driven (Mayer-Schönberger 2014), big data do not lead to more precise observations, but much rather increase the number of observations that are plausible at first sight but often turn out to be wrong.

The topic of dealing with observations and the potential of incorrect observation is discussed in great detail by Popper (1959), introducing the principle of falsifiability. He claims that there is a general asymmetry in analyzing hypotheses. Although it is not possible to verify a hypothesis in its totality, it "can be contradicted by singular statements" (Popper 1959: 19). Big data strengthen the claim of falsifiability, but the opposite effect is currently observable. A large group of people is content with discovering patterns or correlations within big data and believes that this is sufficient due to the amount of big data available. It seems that big data are subjugated to economies of scale to which the problems and the errors in big data

124

are also subjected. As Spiegelhalter reasons: "Serious statistical skill is required to avoid being misled" (2014: 265).

Sagan (1996) was facing a similar situation when he discussed ways to deal with pseudoscience. He demonstrated that there are ways to identify solid scientific research and rigorously tested work and not fall for poorly conducted research. He developed a 'baloney detection kit' that equips people with the tools for skeptical thinking, an ability more crucial than ever as big data have become so complex that they may in fact appear as magic. Many of the tools proposed are perfectly suited for the use of big data. Sagan proposes the following nine tools for his baloney detection kit:

- "Wherever possible there must be independent confirmation of the 'facts'.
- Encourage substantive debate on the evidence by knowledgeable proponents of all points of view.
- Arguments from authority carry little weight – 'authorities' have made mistakes in the past. They will do so again in the future. Perhaps a better way to say it is that in science there are no authorities; at most there are experts.
- Spin more than one hypothesis. If there's something to be explained, think of all the different ways in which it *could* be explained. Then think of tests by which you might systematically disprove each of the alternatives. What survives, the hypothesis that resists disproof in this Darwinian selection among 'multiple working hypotheses' has a much better chance of being the right answer than if you had simply run with the first idea that caught your fancy.
- Try not to get overly attached to a hypothesis just because it's yours. It's only a way station in the pursuit of knowledge. Ask yourself why you like the idea. Compare it fairly with the alternatives. See if you can find reasons for rejecting it. If you don't others will.
- Quantify. If whatever it is you're explaining has some measure, some numerical quantity attached to it, you'll be much better able to discriminate among competing hypotheses. What is vague and qualitative is open to many explanations. Of course there are truths to be sought in the many qualitative issues we are obliged to confront, but finding *them* is more challenging.
- If there's a chain of argument, *every* link in the chain must work (including the premise) – not just most of them.
- Occam's Razor. This convenient rule-of-thumb urges us when faced with two hypotheses that explain the data *equally well* to choose the simpler one.
- Always ask whether the hypothesis can be, at least in principle, falsified. Propositions that are untestable and therefore unfalsifiable are not worth much. Consider the grand idea that our Universe and everything in it is just an elementary particle – an electron, say – in a much bigger Cosmos. But if we can never acquire information from outside our Universe, is the idea not impossible to disprove? You must be able to check assertions out. Inveterate skeptics must be given the chance to follow your reasoning, to duplicate your experiments and see if they get the same result" (Sagan 1996: 210–211).

On the basis of this baloney detection kit, I will present a *big data baloney detection kit*. This kit will be a helpful tool for the HR department to discover faulty data as well as faulty conclusions and create a structure in the HR daemon that tackles the veil of ignorance in organizations. The kit consists of: (1) The necessity to find other data sources to seek validation. Big data are subjective but big data represent a way to access other sources without any effort. Although a certain data set may reveal facts, they are always checked and validated. (2) Big data analyses are not performed by only a few people or only certain people or departments within organizations. Everybody who is influenced by the results of a certain big data analysis needs to have a voice. Many big data decisions will involve employees in one way or another, that is why the HR department and the works council are part of them. (3) Big data are subjective and even data from authorities like government agencies will be distorted in some way. This may happen on purpose or by mistake, but without precise knowledge about the way in which data are collected. These data are not superior simply because they were collected by an authority. (4) In the context of big data, the hypotheses surrounding such correlations derived from data mining will become ever more important and have a major impact on the use of big data. Although some correlations may lack all logic (Vigen 2015), there are many correlations that appear, seem to make sense. These correlations may discover a causal effect; nevertheless, correlations do not give information about causal relations. Who influences whom? Correlations in big data can often be explained in some way, but using multiple explanations will at least lower the chance of choosing the wrong one. (5) If big data reveal correlations, the explanation behind each one becomes more relevant and can be a source of criticism. Although a correlation makes sense in the subjective reality of one person, it may be baloney in other subjective realities. (6) In terms of big data, to quantify does not mean to use more data, but to evaluate the quality of the data available. Although some data may be more numerical, they could be of poor quality. Quantity are, thus, replaced with quality. Good data always trump bad data, however, the answer is not that easy if it comes to the comparison between good data and many data. As Hand (2016: 631) states: "Large does not necessarily mean good, useful, valuable or interesting". (7) Big data are always a mosaic of different data sets (Sprague 2015) and in order to improve the big data analysis, every source and every link is checked for quality and for potential biases within the data set. (8) Occam's razor can be applied in the same way and the emphasis lies on explaining the data set equally well. (9) Results from big data are always tested for errors. Big data analyses need to become more transparent in order to understand their reasoning (Dalton & Thatcher 2014). It is understandable that many big data analyses cannot be duplicated. Big data are closely interlinked with the source as well as with the results (Ansolabehere & Hersh 2012). Many algorithms are self-learning and evolve over time, but in order to discover baloney, people possibly need to follow skeptical reasoning.

Big data have another serious problem. Hypotheses are no longer formulated in advance. Big data can find patterns and correlations without requiring any hypotheses at all. Organizations face the problem of HARK (Kerr 1998), the acronym for the practice of hypothesizing after the results are known. Although it may not

have a strong influence on research, as those hypotheses are still rooted in pre-liminary literature review and research (Bosco et al. 2015), the appeal of HARK in an organization is much greater and may lead to falsifying data (Kerr 1998). This phenomenon helps describe the problems when faced with big data. In most cases, organizations will conduct HARK on a large scale. HARK seems to be considered a legitimized method (Loebbecke et al. 2013). Given that it is not an unsound method per se, dealing with HARK requires different tools. The big data baloney detection kit will contribute to the competences of organizations in minimizing the potential for HARK to have a negative impact.

Sagan defines the following fallacies that are to be avoided as part of the balo-ney detection kit: "ad hominem, argument from authority, argument from adverse consequences, appeal to ignorance, special pleading, begging the question, obser-vational selection, statistics of small numbers, misunderstanding of the nature of statistics, inconsistency, non sequitur, post hoc ergo propter hoc, meaningless ques-tion, false dichotomy, short-term vs. long-term, slippery slope, confusion of correla-tion and causation, straw man, suppressed evidence, weasel words" (Sagan 1996: 212–216). I will not describe them as many of the fallacies are already described through cognitive biases and statistical errors.

Putting big data into a black box and seeing the work of big data as magical would have prompted Sagan to categorize big data as a source of nonsense. Im-plementing a tool for skeptical thinking into the HR daemon is essential. The big data baloney detection kit consists of general elements with which to understand the reasoning behind any big data analysis and will help open up the black box. Intensive baloney checks, however, will require a vast amount of resources which render the intensity of use of such a kit a strategical decision, especially considering the tradeoff of potential risks.

Where is everybody?

In his book, Sagan (1996) addresses the possibility of alien life and tries to debunk the potential of alien abduction. Although the discussion about alien abduction is indeed pseudo scientific, the question of alien life is a fitting example with which to describe the relevance of the big data baloney kit. Big data and many new telescopes allow astronomers to gain a more precise picture of our universe than ever before. Researchers have discovered over 2,000 exoplanets since 1988 (http://exoplanet.eu/catalog/). The universe is vast, however, which is why our knowl-edge of it is massive and minuscule at the same time. With all this information, we cannot answer the question of whether or not we are alone in the universe.

Scientists struggle with the so-called Fermi paradox (Webb 2002). The universe is so vast and old, and there are so many earth-like planets that there is likely to be extraterrestrial civilizations and they may have visited Earth. Webb (2002) suggests fifty answers to the question. He proposes that the following scenarios

are the most plausible: that they are signaling but we do not know how to listen and we do not know at which frequency to listen, and that we have not listened long enough. There are many solutions to the paradox. Scientific observations, however, are generally faced with rigorous skeptical thinking.

In recent years, the planet KIC 8462852 has been the source of vivid discussions. This planet was behaving unusually. Boyajian et al. (2015) discussed a variety of scenarios that could describe this planet. In addition to those scenarios, they do not rule out the possibility, although extremely small, that it was built by somebody. "Aliens are always the very last hypothesis you consider, but this looked like something you would expect an alien civilization to build" (Andersen 2015). Wright et al. (2016) agreed with the statement that explaining such an observation with alien life is the last resort, especially as the hypothesis cannot at the moment be disproven. Explaining inexplicable phenomena induces an "aliens of the gaps" (Wright et al. 2014: 3) fallacy. Ultimately, it is more probable that there is just noise in the data rather than signs of alien life (Boyajian et al. 2015, Wright et al. 2016).

Although we are currently drowning in astronomical data (Zhang & Zhao 2015) and astronomers continue to find new planets and new phenomena, they are using both the classical baloney detection kit and the big data baloney detection kit. Researchers want to eliminate all possibilities before claiming the discovery of alien life.

> "With this in mind, it's possible that this binary nature is due to scientists being extra cautionary on how they present results to the public. If something extraordinary such as life beyond Earth is detected, then we'd better be prepared to unequivocally back up such a statement" (Boyajian in Greene 2016).

Such an example highlights the relevance of being precise and cautious with this particular topic. However, organizations also need to be cautious with their use of data. Using data that may or may not be accurate and being satisfied by correlations, or the first hypothesis that comes to mind, will harm organizations. Critical decisions with far-reaching consequences need to be handled in a deliberate and precise way, just as much as the question of whether or not we are alone in the universe.

4.2.3.2.2 Big Data Tinkering

By detecting baloney within big data, organizations are in danger of creating a tunnel vision. Organizations will restrict themselves within the possibilities of big data. This may even slow down organizations or lead to a dead-lock (Takebayashi & Morrell 2001). The strength of big data is that it is possible to just look into the data, to find patterns, discover coincidences nobody even thought of, or simply simulate someone's crazy idea. An organization needs the ability to "play around" (Jacobs 2009: 36) with big data and to have space for exploratory analyses.

128

Big data alone will not be a source of creativity or innovation, but will enable people to think outside the box and will augment people with new tools. A potentially fitting terminology for this is 'the bricolage' by Lévi-Strauss, who describes it as "doing things with whatever is at hand" (Lévi-Strauss 1966: 17). In an organizational context, bricolage is often linked with entrepreneurship, innovation, and organization theory (Duymedjian & Rüling 2010) and tackles resource allocation within an organization or, to paraphrase, the process of "creating something from nothing" (Baker & Nelson 2005: 329). Although the analogy 'from nothing' is unclear within an organization, as some resources will be re-allocated, those resources, in this case often people, will be used in a different context and environment. Organizations are forced to improvise, fixing things, or designing new things (Weick 1993, Louridas 1999). Weick describes the need for such bricoleurs and their ability to improve and redesign organizations with the following reasons:

1. "People are too detached and do not see their present situation in sufficient detail;
2. past experience is either limited or unsystemized;
3. people are unwilling or unable to work with the resources they have at hand;
4. a preoccupation with decision rationality makes it impossible for people to accept the rationality of making do; and
5. designers strive for perfection and are unable to appreciate the aesthetics of imperfection" (Weick 1993: 353).

These reasons are strengthened by big data baloney detection and shackle people within organizations by focusing on the past, the existing observations and the tendency to implement structures within organizations. They do, however, thereby increase the rigidity of organizations, implement a strong lock-in (Sydow et al. 2009). There may be a tendency to stabilize organizations (Weick 1979), but in today's world stability means being dynamic (Farjoun 2010). Consequently, the work of a bricoleur seems to be an efficient way to increase the dynamization within an organization.

Although bricolage may be a fitting description, the term is sometimes used in the sense of errors, or shoddy piece of work. In my opinion, a more precise term is *tinker*. The Merriam Webster dictionary defines the term as somebody who is repairing and working in an *experimental* way. This is still similar to the bricoleur of Lévi-Strauss, but evades the negative connotation of the word in the English language. In the video game "World of Warcraft", tinkerers are described as:

> "The creators of incredible inventions from steam saws to siege engines, their devices allow them to overcome nearly any situation – and if they don't have the device they need, they just might be able to design and create a new one on the spot" (Kiley 2005: 86).

Tinkerers are known for using their resources at any time and any place. I suggest that big data tinkering will become an essential element in the use of big data within organizations. Creating new ideas and new concepts, using tools in different ways, and utilizing big data for such tinkering or for innovation will boost the

competitiveness of any organization. Such tinkering is not blindly mining big data driven by data, but combines the creativity of people with the computational power of big data. Big data are shackled by their rationality, although this is distorted in some ways, and by their boundaries. Tinkerers can add their irrationality to the mix and drastically expand the benefit to be gained from big data. 3D printing and the maker movement (Dougherty 2012) can be used as an example. Although a 3D printer can print everything, somebody tells it what to print. There are almost no restrictions (e.g. food or steel printing), but a tinkerer needs to use the 3D printer. The same is true for big data.

The HR daemon has the ability to let the actors within organizations tinker, and the HR department encourages people to tinker in various ways. More importantly, the process of tinkering is noted as tinkering and, therefore, is potentially baloney. There is a fundamental difference between using big data in the sense of big data baloney detection and for big data tinkering. The difference between tinkering and precise work is comparable to the contrast between bricoleur and engineer (Freeman 2007). Big data tinkering is also about testing the possibilities of technology (Miller, J. H. 2015) and the ways in which big data can be utilized within organizations. Those tinkerers (as the model in World of Warcraft) are crossing borders, be they social or ethical. They are at the very least a higher risk for organizations, so the HR department needs to establish a safe space for such tinkering. In addition to establishing tinker spaces (similar to maker spaces) and encouraging people to tinker, the HR department needs to balance both extremes of rigorous use and wild speculation. One way to deal with this could be through risk evaluation.

Rosetta Mission

A fitting example of such tinkering with big data is the Rosetta Mission (Glassmeier et al. 2007) by the European Space Agency (ESA). In 2004, a probe was launched on a flight to the comet Tschurjumow-Gerassimenko. The mission goal was to land the Lander Philae on the comet. This alone was ambitious, however, the ESA had little information about the comet. In addition, due to the distance between Earth and the comet, steering was impossible (there was an approximately 30-minute delay), so the ESA faced a situation where they had to have everything necessary on board prior to the launch. Just putting everything into the satellite was not an option, and it would have increased the weight to dimensions that would lead to different problems. Every gram not only cost more money, it would also make the launch into orbit more dangerous. A heavier lander would also cause complications in the landing process. So, ESA did not know the composition of the comet and they had to deal with the strong restriction in weight.

Under these circumstances, landing on a comet is a difficult mission, and a great deal of work was required to increase the chances of success. As a public organization, ESA minimizes potential risk or else funding will be spent on projects that are more promising. ESA, thus, ran simulations for a variety of different conditions. Although this is typical of any space flight since the Apollo (Branch 1997), the Rosetta team claimed to be "prepared for every eventuality" (New Scientist 2015). ESA gathered information from various sources, including previous missions, research surveys, simulations, data from other space agencies, and data from suppliers about their components. All this information was used to design Rosetta and Philae adequately. Although information about the comet was scarce, several compositions were nearly impossible, and it seemed plausible that some combination of ice and iron would be realistic. ESA tinkered a plan in which Philae would grapple to the surface with harpoons, an idea that at first sounded quite extraordinary. Above all, the weight question was tackled by several researchers. A variety of simulations were necessary to find a sufficient solution, respecting the interests of every researcher and minimizing the risk of failure.

Although ESA ran a vast number of simulations, the comet proved to be much harder than anticipated (Yuhas 2014). Philae bounced off the surface and eventually crash-landed. Philae came down in a shadowy region and, therefore, could not generate energy through its solar panels. There was little time to gather as much information as possible. The ability to tinker allowed the team to gather much and, most importantly, interesting data from Philae in the short period of time until the battery died (Dorminey 2014). ESA quickly anticipated what was possible and plausible within the remaining time. Ultimately, the Rosetta mission was executed successfully (Lee 2015).

Especially in today's world, organizations need to think outside the box and be creative. Big data enable organizations to think of every eventuality, however, risk cannot be entirely eliminated. Using baloney detection may have led the organization to decide against the Rosetta mission, but big data supplied the tinkerers with enough information to convince the ESA to follow the plan. It was clear that there were risks, but by thinking of all possible eventualities, the team was able to deal with those problems. Philae may not have worked to the fullest capacity, but it gained new insights, and that was the mission.

4.2.3.3 Big Data Risk Governance

The HR daemon faces two extremes, big data baloney detection and big data tinkering. Both are entangled with a certain type of risk and can, as a result, be categorized through risk. The risk value gives top management the ability to make decisions more precisely and be more aware of the surrounding risks. As noted for the core assumptions of big data, organizations are influenced by risk. Especially

for the goal of achieving a homeodynamic organization, risks are disruptive factors that could disturb the delicate balance within an organization. In addition to the risks of big data, organizations are still affected by risks from external sources and internal sources. Big data help to make risks visible and transparent. Big data are also a potential risk factor, especially if big data are a black box. Furthermore, globalization and interconnectedness render today's world riskier than ever before. Concepts like risk governance (Stein & Wiedemann 2016) attempt to steer risks in a beneficial way for any organization. Big data and risk governance try to decrease the influence of uncontrollable risks. Neither are currently equipped for an efficient search for risks, or for their precise evaluation, yet both may greatly benefit from one another. I, therefore, propose the unified function of *big data risk governance.*

The research field of risk governance developed in recent years (van Asselt & Renn 2011) and its origins can be linked to the European Commission TRUSTNET program (Amendola 2002). There is, however, no common definition of risk governance. Generally speaking, it deals with the regulation of (commercial) risks (Renn 2008, Stein 2013). In order to define risk governance precisely and understand the underlying framework, a clear look at the terms risk and governance is required.

Risk can be defined as the "effect of uncertainty on objectives" (ISO 31000), and even though this definition describes the situation adequately, it does not sufficiently cover the potentials of risk. An assessment of risk seems feasible, however "understanding these various aspects of uncertainty in a complex system is extremely difficult" (van Asselt & Renn 2011: 437). Risk is also connected to ambiguity (Renn et al. 2011) because risk regulation is always linked to people, and ambiguity refers to the existence of multiple values. This makes risk assessment variable and debatable. Risks can be separated into simple risks, complexity-induced risks, uncertainty-induced risks, and ambiguity-induced risks (IRGC 2005). The first type is rare as risks are rarely simple (de Vries et al. 2011). The majority of risks can be sorted into the other classes, but findings reveal that risks are usually managed as simple risks (van Asselt & Renn 2011).

The term 'governance' refers to several different actors determining decisions, the appropriate framework, and processes (Hagendijk & Irwin 2006). The term 'governance' is derived from the Latin word *gubernare* and the Greek word κυβερνάω. In ancient times, it was connected to the navigation of a boat and the responsibilities of the captain. The term 'governance' within an organization follows the same rationale and describes the process of steering an organization in a certain strategic direction. Governance per se is, therefore, the task of 'navigating through rough waters'. Similar to the definition of risk complexity, uncertainty and ambiguity will influence governance and make the task extensive. As far as governance is concerned, the task is to deal with risks.

Based on the definitions of risk and governance we identify both terms as linked to complexity, uncertainty, and ambiguity. Risk governance is, therefore, a construct that tries to tackle the complexity, uncertainty and ambiguity of risks in a way that is traceable and systematic. Risk governance also includes structures that monitor and give early warning (Charnley & Elliot 2002). Risk governance is not simply a type of

risk management; it increases risk resilience (Collingridge 1996). Discussion regarding risk governance has led to a dynamic concept concerning risk. Dealing with risk is no longer an if-then-else loop but a system that is flexible enough to adapt to the prevailing conditions, however, if for risk governance we provide the roles of steerer, captain, and decider, it takes on a superordinate role within organizations.

Big data and risk governance are capable of dealing with risks, however, they are apparently inadequate for dealing with the overwhelming complexity of risks, especially because big data on their own are a source of novel risks. Unifying both functions into one reveals that there are several complementary aspects. Risk governance, on the one hand, requires information in order to search a risk network for potential risks. Lacking accurate information makes steering an organization impossible. Big data support risk governance with an abundance of information (Bell et al. 2009). On the other hand, big data struggle with the evaluation of their objectivity. There is an inherent data bias in any big data analysis. Interestingly, risk governance deals with such shortcomings, and subsequently, such uncertainties and risks on a daily basis. Risk governance is, therefore, capable of supporting big data analysis. Big data and risk governance could significantly benefit from each other. Big data and risk governance enable each other to work more efficiently, particularly in providing rigor and the relevant results for organizations. Based on this, unifying both systems creates one singular function capable of utilizing those dualities.

The function of big data risk governance creates new tasks within organizations. Due to its duality, big data and risk governance cross-fertilize each other. As shown in Figure 10, I propose the following elements: establishing, identifying, seeking, assessing, mitigating, and anticipating. These aspects of big data risk governance enable risk governance through big data and vice versa.

Figure 10: Big Data Risk Governance
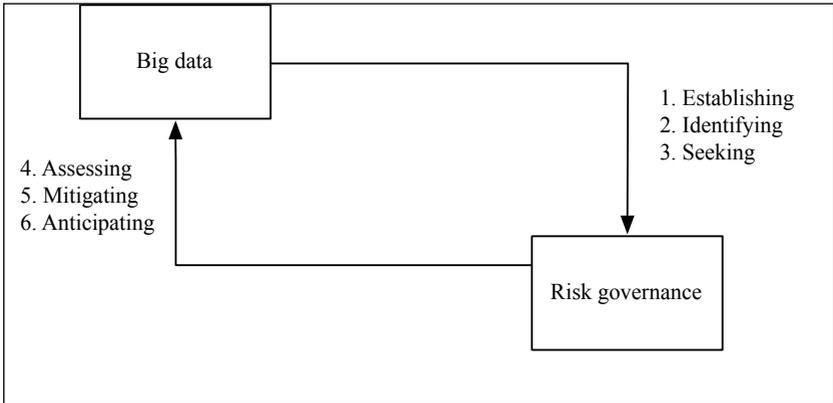
The first element is *establishing*. As stated earlier, the risk network is essential for risk governance. What are the potential risks for an organization? Big data can provide the necessary information for such a task. By analyzing all available data, it is possible to establish a risk network of all risks. That information can include risks that have only a distant effect on organizations, but still are intertwined with them in a small way. On that basis, risk governance obtains a broad but precise picture of the risks surrounding organizations.

In a second step, big data supports risk governance in the *identification* task. Knowing all risks can be overwhelming and can have a paralyzing effect, however not all risks are relevant to an organization. Depending on the risk network, some risks are more influential than others, and on the basis of this information, risk governance can focus on a selection of risks rather than all risks. Big data also provide information about the connections of risks within the risk network. How are those risks connected and how do they interact with each other? Based on the answer to that question, big data can contribute to the *seeking* process of undiscovered risks. Due to the granular picture of the interconnections of risks within the risk network, it is possible to find new risks: in today's complex world in particular, these new risks can be from the result of second-order effects or cannibalization effects (Desai 2001). Although a single risk seems insignificant on its own, in connection with other risks it could be critical. In those first steps, big data support risk governance to get a clearer picture of the risk network and enable risk governance to act in a better and quicker way. This is especially true since those tasks can be done in 'near real-time' (McAfee & Brynjolfsson 2012). Big data can also simulate a variety of compositions of the risk network and develop various predictions.

Big data can also be supported through risk governance. In the fourth step, risk governance improves upon the *assessment* part of big data. As stated earlier, big data are not as objective as some researchers believe (Boyd & Crawford 2012) and that means big data depend on a critical analysis (Dalton & Thatcher 2014). Risk governance can fill this void and provide an assessment of the risk network and the influences of such a risk network. What are the causal relations and do they make sense? Those results need to be comprehended from both a contextualized and a holistic perspective. Due to such thorough analysis, risk governance helps to find errors within the big data analysis and also supports big data in *mitigating* their risks. Risk governance could use several algorithms to minimize the big data risk. Finally, risk governance supports big data in *anticipating* new developments and new risks. Big data, on their own, only find results within their limited data sets. Big data only react to this constructed data world. Every predictive analysis (Sprague 2015) will take place based on that data bias (Scholz 2015b). Risk governance needs to seek new data sources, implement new ideas, and proactively envisage the potential (re) actions of the environment and especially of the human actors within the environment. Reinforcing this effect, humans react to the results of big data analysis, and this could lead to self-fulfilling prophecies (Merton 1948), anticipatory obedience (Lepping 2011), or self-preventing prophecy (Brin 2012). Such behavior will cause

distortion, forcing big data to adapt. Big data risk governance can anticipate such behavior as well.

This duality results in a big data risk governance that is capable of acting and reacting in a quicker, broader, deeper, more differentiated, more sustainable, and more insistent way. Due to the dualities of both elements, big data risk governance develops a risk network that helps understand the complex environment in which an organization acts. Big data become more precise, less risky, and are questioned constantly. Big data risk governance establishes an elaborate risk network that needs to be fostered and groomed permanently, and this comes at a high price. As previously stated, there is a dilemma in deciding between big data baloney detection and big data tinkering, however, big data risk governance is capable of revealing its usefulness by showing potential gain and loss and presenting simulated results. It represents an investment for the future. Big data risk governance, combined in this duality, supports itself to overcome its inadequacies. It is reasonable from an evolutionary perspective to combine both worlds into one distinct function, thus giving an organization a function to navigate through the data deluge and through the risk network, leading to the achievement of the goal of a homeodynamic organization in the midst of a turbulent and stormy sea.

On top of these aspects of the big data risk governance, the ethical problem as the ethical component is part of the concept of risk governance (Stein & Wiedemann 2016) and, therefore, will be part of big data risk governance as well. Although, risk governance is a highly ethical topic, big data are the subject of ethical discussions even more frequently. The importance of tackling the ethical question in the field of big data can be highlighted in the following example. In the video game "Starcraft II: Legacy of the Void" (issued by Blizzard in 2015), there is an interesting discussion that highlights the relevance of being critical, and skeptical, and dealing with big data in a social and ethical way.

> "Karax: The replication data is the sort that allows accurate duplication of one's consciousness. Fenix personality may be accurate. Within the ninety-ninth percentile.

> Artanis: So there is a chance for discrepancy.

> Karax: Quite a miniscule one.

> Artanis: And in a lifetime, how many choices does that variation impact? Who would you be with such a difference in the decisions you've made?"

Although copying one conscious soul to another may be science fiction, it can be applied to the big data discussion. As noted earlier, there is a data shadow of people created by big data, which creates a hyperidentity that may or may not coincide with a person's real identity. This example explains that any difference between the data shadow, the social shadow, and the actual identity will have consequences and will influence the actors within any organization. Although there may be only tiny variations between those shadows and identity, over time, these differences could become impactful. As long as big data are not big enough to meet these expectations,

big data use will be pivotal. It may be even the case that there will always be some remainder of difference, as long as we are not part of an all-embracing surveillance society. If big data cannot grasp the behavior of people entirely, it will only create a subjective view of the shadow, and analysis depends on the methods used to decrypt such behavior. Big data will only give us a portion of the data shadow and only one shadow from a certain viewpoint, leading to an interpretation that leaves much room for interpretation. As Barry (2011: 8) summarizes it, big data "provide destabilizing amounts of knowledge and information which lack the regulating force of philosophy which … ensures that institutions remain rational". Organizations deal with a variety of different ethical obstacles in the use of big data. Mittelstadt and Floridi (2015) derived the following ethical themes on the basis of a literature review of 68 papers:

- "Informed consent
- Privacy
- Anonymisation
- Data protection
- Ownership
- Epistemology
- Big data divide" (2015: 10).

Informed consent tackles the question of whether people's consent to allowing data to be collected can become something dynamic in times of big data. People do not know what data are collected about them, or know how such data are used. This leads to privacy and anonymization issues, two themes that are strongly influenced by big data. Privacy is a highly debated topic in terms of big data (e.g. O'Hara & Shadbolt 2008, Solove 2011, Tene & Polonetsky 2012), and there will be several transformations concerning that topic (Rubinstein 2013); anonymization is a concept from the past, and with big data it becomes relatively easy to de-anonymize anonymized data sets. Clemons et al. (2014) call this the myth of anonymization. Data protection is essential for big data within organizations, however, many leaks (e.g. Kuner et al. 2012) reveal that it is still a neglected topic and that is critical as the data sets become more granular and more individualized. Ownership is about the discussion of who owns the data. This is a topic that will be discussed in detail regarding big data authorship. Mittelstadt and Floridi (2015) identified a link towards epistemology. It seems problematic to understand big data and the complexity of big data, in a context where big data more and more mimic a black box. Finally, Mittelstadt and Floridi (2015) deal with the big data divide and tackle it through the divide of power and control over big data. The element of surveillance and profiling is especially highlighted. People are unaware of being profiled or surveilled. All these themes raise questions about justice concerning big data and the difficult task of dealing with big data in an ethical way.

The predominant question for dealing with big data within organizations is, thus, what is ethical big data use. There are two ways in which a moral compass could be derived: on an individual basis or at an institutional (group-level) basis. Both are

influenced by each other and focusing on one will harm the other (Mittelstadt & Floridi 2015). We therefore seek alternatives. One way would be to look at a higher order system that connects both ethical perspectives. This could be found in a kind of big data ethos. *Ethos (ἔθος)* is the Greek word for custom or habit and describes guiding beliefs or ideals. An ethos supports its users with guidelines and simple rules to follow. There is still space to act according to individual and institutional ethical elements, but ethos also supplies people with an ethical safety net.

Big data ethos is an omnipresent guiding system that influences the complete big data use process. Ethical considerations are essential in data collection and data analysis. Data shadows and social shadows are competing to influence the perceived identity. Given that there is an inherent bias in big data use, from the viewpoint of a big data ethos there is a need for responsible handling throughout the complete process. The crucial part of big data, therefore, is vigilance. I am proposing a concept of "data vigilance." The term vigilance is derived from the Latin word *vigilantia* and means wakefulness, watchfulness, and attention. Vigilance is important in order to adapt the use to remove any kind of bias. Data vigilance is also linked to someone's accountability.

In order to specify ethical big data vigilance, I propose a framework consisting of four dimensions. *Attention*: This deals with the element of being alert at all times and developing a watchful eye in every situation. *Consciousness*: This refers to the necessity of having some sort of ethical value system or ethos and following its values. *Intention*: The purpose of big data use is to reach certain objectives which go beyond the purpose of maximizing profit and include all interests within an organization, especially those of the employees. *Stabilization*: Analyses on the basis of big data can be done in real time and organizations can be completely flexible, however, the goal is to make an organization stable (not static) within its environment. Big data enable organizations to become more homeodynamic and, therefore, sustainable.

All these dimensions are essential in order for organizations to gain an understanding of what uses of big data may be ethical in their particular case. It is essential to understand that a certain use may be ethical for one organization and deemed unethical for other organizations. On the basis of its organizational signature, an organization already has some insights into a rudimentary version of the ethical value system. Facebook will have a different value system and will, for example, be more open to big data than Airbus. In a way, Facebook's product is big data and, therefore, will focus even more on big data than Airbus. In order to highlight the necessity of vigilance, we can look at the following example. It is possible to insert code into a webpage to retrieve the battery status from a smartphone. The idea of this was initially to deactivate certain functions to save the battery. Although that makes sense at first glance, it is theoretically possible to use this information to identify a particular user on the internet because the information about battery status is incredibly precise (Olejnik et al. 2015). With knowledge about battery status, people can be tracked across the internet and the browsing history can be reconstructed. The HR department is watchful with such information and deals with it in a fitting way.

Big data risk governance is, consequently, helpful for an understanding of the work of the HR department within the HR daemon. Ethical big data vigilance within big data risk governance allows organizations to derive methods that fit, as well as establish structures to proactively find new issues that may be ethically questionable and deal with them. Vigilance will also allow organizations to be more resilient, transparent, and, most importantly, comprehensible to the other actors within organizations. Dealing with big data in such an ethical way will decrease distortion, but also will disenchant the magic of big data. Organizations use big data to improve themselves, and, therefore, treat people within the organization in an ethical way.

Case of Google Flu

Although this example does not completely follow the proposed big data risk governance model, it reveals the relevance of the concept and the need for dealing with the risks of big data and the surrounding risks. In 2008, Google launched a project that helped to predict outbreaks of the flu. Google claimed that their predictions were 97% accurate compared to data from the Center of Disease Control, but without the time delay that CDC results normally have (Ginsberg et al. 2009).

Google used a vast amount of data to *establish* a risk network concerning flu-related searches. They *identified* 1,152 data points that related to the flu (Ginsberg et al. 2009), however, they initially did not *seek* new or abnormal search patterns like the A-H1N1 influenza (Cook et al. 2011, Olson et al. 2013). Those inconsistencies within the risk network caused Google Flu to overestimate flu prevalence, making the results no longer precise, and rendering them even less accurate than those of the CDC (Lazer et al. 2014, Kugler 2016). Those errors within the big data analysis were spiraling out of control and Google needed to *assess* the potential risk sources. As Lazer et al. (2014) note, Google changed the software and the algorithm of their searches. In 2011, for example, they introduced a feature that suggests search terms on the basis of the initial search word. People also change their search behavior (Lazonder et al. 2000) over time and search engines are susceptible to manipulation to a certain degree (Zwitter 2014). Understanding and comprehending those influences is important, but it is crucial to *mitigate* those big data related risks. Lazer et al. present one solution in their paper: "By combining GFT and lagged CDC data [...] we can substantially improve on the performance of GFT or the CDC alone" (Lazer et al. 2014: 1203). There is a high volatility inherent in the internet, and Google is in the midst of all those changes. How do human dynamics interact with algorithm dynamics? It is essential to *anticipate* for future challenges. Google has enough data, but as this case shows, they do not always ask the right questions.

Taking the perspective of ethical big data vigilance, it may seem that many of the problems could be prevented due to having a high *attention*. Nevertheless, some sort of *consciousness* was available, and the *intention* of Google was to improve the health aspects of their users and, therefore, can be seen as positive. As well as the foundation for some *stabilization* in their ethical big data vigilance. But, the main problem seems to be that Google had certain blind spots in their attention. They did not see various risks that led to the distortion of results. Consequently, only the attention was problematic, but this was critical enough to lead to the problems described.

There are many sources that make any endeavor riskier, and big data can contribute to such risk. At the moment, there is a blind spot concerning the risks of big data, although it seems obvious that big data will not always find the correct answers. Big data and the case of Google reveal that it is not a pure, objective, technical entity, but that big data are strongly entangled with the social world and any change will influence their informative value. Algorithms change and people change in such volatile ways that they make big data a risk as well. Therefore, the ethical perspective is crucial within the observation of big data and, furthermore, is an ongoing process. Big data risk governance involves dealing with those risks as well as the ethical consequences and enables organizations to thoroughly evaluate their strategic decisions.

4.2.3.4 Big Data Immersion

In the next step, the HR daemon tackles the integration of big data within the organization. In particular with a focus on the relationship between big data and the people, big data will become immersed in the organization and, therefore, affect many aspects and fields within it. First of all, the HR daemon will need to tackle the questions surrounding data protection, privacy, ownership, and copyright of big data, as well as the people generating the data, which will be conceptualized in big data authorship. Furthermore, big data are not static entities and will change constantly over time and space. Therefore, if big data are an integral part of the organization, it will be necessary to monitor and maintain big data or to deal with *big data curation*. Finally, a main part of the HR daemon will be to train the employees in handling big data on their own and developing essential big data competencies. The HR department will increase *big data literacy* within the organization.

4.2.3.4.1 Big Data Authorship

Using big data within an organization can be beneficial, and is in the bilateral interest of employer and employee, but there are still the main issues of data protection, privacy, ownership, and copyright of big data about employees. It is important to highlight that those terms are not synonyms but rather tackle diverse topics concerning big data (Dix 2016). However, organizations are facing a difficult situation.

On the one hand, complete transparency is not the ultimate goal, as it can lead to information overload (Toffler 1970); on the other hand, hiding all data is also the wrong approach, as it can lead to a violation of trust. It is essential to find solutions to secure data and privacy, as well as to legally ensure copyright. Within an organization, it is essential that some balance is achieved and that both the organization and the employees benefit from big data adequately. That can be difficult, as some use can violate privacy and copyright. As stated earlier, the legal landscape is currently still struggling in relation to big data. One issue is that because of legal regulations data are only used anonymously, but that would cripple big data use at the individual level. Big data can support employees in individualized ways, but if the employees are anonymized this benefit dissipates. In fact, being anonymous is a myth in these times of big data (Clemons et al. 2014); with enough data it is possible to de-anonymize any information (Tene & Polonetsky 2012, Froomkin 2015). Although an HR department does not de-anonymize these data sets, the potential for malpractice is clear. Another issue is that, for example, European law prohibits personalized data use if a specific purpose is not given, and the tools of big data such as data mining are legally highly restricted. Following the law to the letter would mean that exploring big data is not allowed within organizations in any way.

Although data protection laws are more rigid in Europe than in any other part of the world, problems with big data are not limited to European organizations. Keeping employees in the dark and abusing big data leads to the post-panoptical (Bauman 2000) behavior of employees. Because employees believe they are monitored, they change their behavior appropriately in the sense of anticipatory obedience (Lepping 2011), for example, "cleaning up" their Facebook profiles (Brown & Vaughn 2011), and thereby distorting the data shadow and the social shadow even further. Other employees could discover the patterns of surveillance and exploit the system (Zarsky 2008, O'Neil 2012), as well as making the shadow of their identity vaguer. People changing or hiding their behavior will lead to an impreciseness of big data and subsequently to errors in decisions based on *infected* data.

The question of what type of privacy is even possible in today's organizations also arises, in organizations, in which movements are traceable, sensors are ubiquitous and smart machines are collecting a massive pile of data. It becomes increasingly difficult not to gather data about employees, not even information from secondary sources. Smart machines depend on their sensors and for security reasons need to keep track of the humans around them. This information about the people could potentially be repurposed for different objectives. Within an organization, people are constantly tracked, deliberately and unwittingly, thus making their data shadows bigger (not necessarily more precise) and contributing to the hyperidentity of employees within organizations. The question of rights regarding data is even more unclear in that case. Do we assume that the person (or organization or even machine) that collects the data has all rights on the data, or that the person the data are about holds the property rights?

Privacy (Matzner 2014) and copyright laws (Kaisler et al. 2013) are apparently unfit to deal with such modern problems (Lessig 2008), and, therefore, I propose a

concept of *big data authorship*. The idea is rooted in similar observations in virtual worlds (Roncallo-Dow et al. 2013). In those worlds, "authorship is a collaborative act" (Guertin 2012: 2) and goes beyond the question of copyright and privacy. In this case, both the player within the virtual world and the creator of the virtual world create the design and story of this virtual world together. They are both its authors. Although privacy and copyright are still difficult to grasp, both parties understand and see their task as producing and contributing to a common goal, and in some cases, these interactions evolve or emerge in a form of unwritten social contract and mutual trust (Roncallo-Dow et al. 2013). Virtual worlds such as 'World of Warcraft' and 'Eve Online' are built on these premises of collaboration and both games have now existed for over ten years. People are becoming more committed to remaining loyal to a game if they perceive the authorship to be fair and truthful.

In the context of an organization, gathering big data is also a collaborative act to which everybody within an organization contributes. Due to the complexity of data, it is difficult to untangle these contributions. If we understand big data within organizations as a similar concept to the authorship of virtual worlds, then big data are a shared experience and joint action between organization and employee. Big data are first of all kept within organizations, and the HR department will act as "primary gatekeeper" (Grimes 2006: 970). Keeping the data generated by an organization and its actors within the organization will increase the trust of employees. They will more freely share their information if they know that the data are safe and secure.

Everybody is seen as the author of big data within organizations, and the HR department is responsible for the fair use of big data. HR departments can flag certain data as private or as having limited visibility, and employees can do the same. If employees are interested, they can use existing data for their own analyses following the motto: putting big data into the hands of employees. The HR department fosters this relationship and monitors fairness within the organization. A social contract, as in the example above, may be a broad solution, however the HR department could also use the tool of psychological contracts (Rousseau & Tijoriwala 1998). Everybody collaboratively contributes to big data within organizations and is seen as the author of that. The HR department needs to implement the ability to have data transparency as a part of the HR daemon, allowing employees to use the data (and tinker with big data). The HR department also needs to evaluate the appropriateness of hiding certain data, avoiding a potential transparency trap (Bernstein 2014).

Big data are not limited to employees within organizations, but include employees who have left organizations, and who have authored a variety of data in their time within organizations. It would be possible for the HR department to define a certain data set involving these employees and cleanse it from internal data. Such personal data sets can describe several performance indicators and serve as a datafied certificate of employment. In analogy to encryption, this data set could be compared to a personal key. These data are cleansed from critical information about the organization but are meaningful for the employee. Furthermore, this key can be combined with the organizational signature (Wang et al. 2014, Stein et al. 2016), and would generate a simulated assessment of the employee within any organization.

Such personal information could be an interesting addition to the recruiting cycle. First of all, employees would have a reason to share their data willingly because it benefits them, and, secondly, any organization could check the fit of a potential employee more precisely. It is important to emphasize that only with a combination of an organizational key, on the basis of the organizational signature and a personal key, meaningful and contextual results can be derived. The organization or the top management will trust the HR department to be responsible for those personal keys, and especially for the organizational key.

Although legal discussions concerning copyright and privacy are ongoing, big data authorship is a proactive task for the HR department which utilizes big data in a more transparent way. It reduces the power imbalance and enables an employee to contribute to the use of big data. Trust will be strengthened, and the employees will include some (encrypted) transparency, as well as keeping the useful part of their data. Big data are, at least within organizations, not something blurry that somehow emerges out of nothing, but something to which everybody is actively contributing. This change from being constantly monitored without knowing it to the idea that such data can be used by both organizations and employees to improve the organizational environment and an employee's career, marks a strong psychological change in narration. From a legal perspective issues are not solved by authorship, but it enables organizations to find a solution that allows them to utilize big data for their own needs. The emphasis now is more on perceived fair use and the trust relationship between organizations and the employees. This is something HRM was designed for, and only the HR department is capable of dealing with these interrelationships within organizations.

Quantified Self

In light of current digitization and the technological progress in wearable devices, people are now capable of self-tracking (McFedries 2013). The process of self-digitization, self-monitoring, or self-quantifying is combined in the term 'quantified self'. People are willing to share their data in order to be compared to other users and to evaluate their performance. Quantified self-movement is, currently, often discussed in the context of health topics (e.g. Swan 2013, Ruckenstein & Pantzar 2015) and the motivation can be described as follows:

> "... technological developments in the portability, precision and 'accuracy' of heart rate meters has transformed the realm of everyday calculability. They allow us to 'see' our own heart (instant feedback), and in seeing, allow us to make adjustments in what we do: they allow us to quite literally tune our own engine" (Pantzar & Shove 2005: 5).

There is the ideal that such self-tracking will be about one person all alone, so the claim is that n = 1 (Nafus & Sherman 2014). However, the step linking such data with big data is relatively small, and in the context of organizations there are ways to track people's communication. With the help of badges, it is possible to track and analyze the communications of all employees within an organization (Orbach et al. 2015, Atzmueller et al. 2016): who talks with whom, for how long, at which location and the topic. This can be linked with analysis of the voice: Is the person agitated? Combining the quantified self-data with communications and voice data would give many insights into employees.

It is obvious that such data need to remain within organizations, and it is also obvious that the decision-maker within organizations will have some knowledge of the people involved. If, as in the paper by Orbach et al. (2015), the goal is to improve informal communication within organizations, somebody needs to know which people are being analyzed. This information may not be of relevance in the results of the analysis, but there will be somebody within the organization that had access to such data. Using data from wearables would become even more personal. Data will also be available from smart machines, for example, infrared sensors could unwillingly monitor employees, and these data may be useful to organizations.

Data about employees would become more personal and more detailed than ever before. Big data within organizations are unwillingly full of data shadows which are not anonymized in any way. It becomes essential to have somebody who watches over the employees and allows them to use their data as well. Organizations and employees author big data within organizations and, in order to utilize such data to the maximum, people need to trust and believe in the fair use of such data. Interestingly, quantified self-movement shows that people are willing to contribute their data if there is an actual incentive, such as better health (Swan 2013), and a certain trust that their data are protected (Nafus & Sherman 2014).

As in any organization within the self-tracking business, big data within organizations depend on the people and their self-interest to contribute their data. There is a need for trust of the data fiduciary. Any tracking and, therefore, big data within an organization is surveillance, but the task of the HR department is to make the experience convenient for everybody involved (Whitson 2013).

4.2.3.4.2 Big Data Curation

In the company context, it is essential to remember and to note the missing objectivity (Gitelman 2013) and interference in the way data are gathered, analyzed, and interpreted (Van Dijck & Poell 2013). False claims of objectivity will have an impact and will disrupt the relationship between employer and employee. Making the big data value chain transparent within organizations and additionally incorporating the subjective bias into the analysis will improve the relationship. Both sides will

be increasingly able to understand and discuss the results. The HR department acts as a moderator for such communication, a task that is already present within organizations.

Although organizations have a variety of data available and generate more data all the time, as described in big data risk governance, big data are not jacks-of-all-trades. Depending on the way big data are analyzed, there are different sources of risk. There is a general risk, but there are more specific risks from big data such as subjective interpretation, contextualized data, statistical biases, sampling biases and so on (McNeely & Hahm 2014). There is, thus, a *big data risk additive* that the HR department incorporates into big data analyses. Due to such an additive, it seems that big data and the results of big data are subjective, and need much work to be transformed into results that can be used by an organization. Mayer-Schönberger and Cukier (2013) envisioned new professionals called algorithmitists, who "would act as reviewers of big-data analyses and predictions [...]. They would evaluate the selection of data sources, the choice of analytical and predictive tools, including algorithms and models, and the interpretation of results" (2013: 180). I argue against the idea that such professionals understand, at first, the inner life of organizations and therefore have knowledge about the organizational signature and also the competencies for the analysis of big data. A technical expert would not be suitable, but a social expert could deal with the inadequacies of big data within organizations.

Big data are often contextualized, subjective, and consist of repurposed data. The amount of data solely collected about people and for the purpose of HRM is relatively small and often categorized as bad data (Buckingham 2015). Many processes within organizations are outsourced, automated, or robo-sourced (Gore 2013). Data are generated that do not follow certain standards within organizations or are not available for further use. Big data also deteriorate over time and become less precise and riskier to use. There is a similarity with the half-life of knowledge, in which the knowledge of people and their competencies may become obsolete over time, but the period of time varies from knowledge to knowledge. Big data will become obsolete over time, yet the speed at which data dissolves depends on the data. The big data half-life adds to the risk additive.

The HR department curates big data within organizations and deals with existing big data and the acquisition of big data. The task is similar to that of a museum curator. Big data need to be presented in a certain form to fit a distinct theme, and such a theme could be the organizational signature. Big data need to be checked and refurbished if necessary. Big data from other sources that may be useful are controlled and adjusted to the organizational signature. The origin of any data is however clearly labeled and the changes made to the data tracked. If it is unknown when and where certain data were collected, there is a high risk that such data will be highly subjective and highly outdated. If the data collected from a reliable source are fundamentally changed and distorted by that in certain way, such data are no longer reliable.

Data are always interlinked with other data within big data. Changes in one part of big data influence other parts of big data. If the HR department discovers

errors, they need to correct them at this certain point, but also need to check all links connecting to the error. The curation process has, consequently, the potential to be self-healing for an organization, with a focus on big data. Errors are also bound to happen, especially in a turbulent environment and with heterogeneous data sources, and if the data are contextualized in an inept way. Somebody has the ability to control and curate data in a distinct way that fits the organization, and to archive data that no longer seem required or seem outdated. Alternatively, in contrast to the half-life of knowledge and the need to unlearn such knowledge, big data can file such outdated information, put it in an archive and access it at any time. Only the required data and the current data are on display.

A variety of *control and test* mechanisms are necessary. Finding errors or distortions within big data are a critical task for the HR department and the ability to do this is implemented in the HR daemon; however, as big data are vast – at the organizational level, individual level, and relational level – it will be not sufficient to control only one level. It is necessary to have the ability to combine levels so as to spot problems. For example, a combination of distant reading (Moretti 2013) and ground truth (Pickles 1995) will be essential in order to triangulate the effects and identify consequences on organizations. Distant reading is an approach to understanding "literature not by studying particular texts, but by aggregating and analyzing massive amounts of data" (Schulz 2011), a description that fits any big data approach. Accompanying such a distant picture is a method from the field of cartography, where researchers use data from the ground to support their analyses. A similar metaphor can be used with big data analysis. Although a large amount of data allows a picture from *far above*, it is also essential to validate it from the ground. The ground can mean the individual level, but it also can mean the methodological inner life of an algorithm, so the HR department can look into the heart of their big data analyses. Especially in times of machine learning, those algorithms act on their own in a certain way. Using the analogy of the museum's curator, they will not want museum pieces to be categorized without knowing how the categorization works. The curator can try to make sense afterwards and reverse engineer the algorithm behind it, but if the algorithm sorted the pieces inadequately it will take time and resources to rearrange the museum pieces. The same is true for big data: letting the algorithm do the work may sound promising at first, and if it works it works, but there is a risk that the big data within the organization will be transformed into something irreversible.

The HR department implementing the HR daemon, therefore, needs to monitor, handle the risks, collect, curate as well as control, and test big data within the organization. It also needs to detect and categorize data shadows and social shadows, biases and track changes. The organizational signature is the masterpiece of the collection and is treated and preserved in that way. That task may not be done exclusively by HRM, but they lead the curation, and are responsible and accountable for big data curation within organizations.

Employment Screening

As noted earlier, it has become customary to do background checks on employees through big data. Often this is done by checking the activities of the potential employee on social media. It is much debated (e.g. Sorgdrager 2004), whether these results are appropriate for categorizing potential employees. Social media profiling (Esposti 2014) is part of employment screening, and "there truly has been an explosion in how technology has changed and continues to change selection practice" (Ryan & Ployhart 2013: 20.11). Not only do organizations use social media, but they do extensive background checks. Organizations use a variety of sources and in that way use external vendors of information.

Another way of screening employees in the U.S. is by evaluating their credit scores, as provided by one of three scoring companies, and this is used in many organizations (Bernerth et al. 2012). A recent study by the SHRM (2010) discovered that 43% of organizations (n = 385) checked their job applicants on the basis of their credit score if they are potentially selected, and 13% of organizations run a credit check on all job candidates. A credit score describes the creditworthiness of a person. Organizations use this score to make assumptions about potential employees (Hollinger & Adams 2008) and predict their behavior and performance (Gallagher 2006). Normally, organizations would not obtain numerical values but rather information about how much money is owed to whom (Kuhn 2013). It seems misleading to use such broken down values as there is a variety of information lost in obtaining the score. Reasons that could explain a low credit score or poor credit report are not available, such as race, residence, or family status (Traub 2013). That in itself is problematic, and leads to a new type of financial discrimination (Shepard 2013).

Although the implications of credit checks are questionable, organizations rely on the data delivered by those external scoring agencies, and they depend on the accuracy of those credit scores. Choosing an employee on the basis of a credit score and realizing afterwards that there are errors in the credit score could have an impact on the selection of the best candidate. Eventually, the best candidate may not be identified as the best. The credit score is not as accurate as some people believe. The Federal Trade Commission (FTC) ran a survey in 2012 and discovered that "26% of the 1,001 participants in the study identified at least one potentially material error" (FTC 2012: i), and even worse, 5.2% had an error in their credit score that would lead to a lower interest rate for a credit. Consequently, a credit score may or may not be accurate and in a follow-up study the FTC revealed that people who disputed their credit score had a "meaningful credit score increase" (FTC 2015: ii). A credit score will have an impact on recruitment if it is incorrect, and by a percentage of 26% there is a high probability that there will be errors within the credit score of most job candidates.

146

The task of big data curation, in this case, is to incorporate the potential risk into the calculation. A credit score can probably give some insight into the history of any potential employee, however, there is a margin of error. The credit score needs to be flagged as a subjective score, and the credit report as a subjective source of information. At the very least, any potential employee will have the chance to give a comment on this information. Errors may be unknown to them or there may be other reasons for this score.

The element of discrimination especially is seen as critical. Recruitment on the basis of data or numbers could disguise discrimination behind a veil of objectivity. As Traub (2013) found, there is a discriminatory factor within the credit score. This factor can be discussed in regard to the Chinese Credit Score or social credit system (Stanley 2015). The Chinese government is truly applying big data to a universal score. Due to the high regulation of the internet, they can collect a massive amount of information about people and, additionally, connect a person's information to their friends. The score is then evaluated not only from an individual's behavior within society, but from how their friends behave (Falkvinge 2015). The score determines whether people can apply for a visa or a loan (Hua, 2015), and is seen as "the most staggering, publicly announced, scaled use of big data" (Obbema et al. 2015). Although the rating is at the moment available to the state, people brag about their scores on social media (Doctorow 2015) and the next step is to use this score for recruitment, especially as it is more granular, and made on the basis of more information than the American credit score. This sounds like science fiction, but as Doctorow summarizes it:

> "Paternalism, surveillance, social control, guilt by association, paternalistic application of behavioral economics and ideology-driven shunning and isolation – it's like someone took all my novels and blended them together, and turned them into policy (with Chinese characteristics)" (2015).

Organizations cannot change such systems, but they will factor in all the problems such a system would mean for an organization. Although such scores are seemingly accurate, they are not. There may be a social agenda behind them, but big data are predominately subjective, erroneous or simply outdated. Such errors are difficult to eradicate (Pasquale 2015) and so the HR department deals with the risk additive. Big data are in dire need of curation in a form that means organizations can use them in an efficient way. Blindly introducing big data into organizations will change the organization in an uncontrollable direction. As for a museum curator, however, there are ways to transform big data to fit with organizations.

4.2.3.4.3 Big Data Literacy

The role of people within organizations is currently transforming fundamentally. Machines and computers are becoming the grunt workers for many narrow and repeatable tasks. This development is also observable in conjunction with big data:

employees gain room to focus on complex thinking, innovation, and creativity. This depends, however, on the utilization of big data. At the moment there is a disparity between people who have the ability to use those new technologies and people who are not able to do so extensively. The former are augmented by technology and capable of doing incredible things, the latter, however, fall into a veil of ignorance and are driven by big data to a certain degree. To make matters worse, at the moment there is a war for big data talent (Ahalt & Kelly 2013), in which government agencies and IT companies are competing with every other organization. Organizations need to close the big data gap and recruit or train potential candidates. The HR department needs to improve big data literacy within organizations (Christozov & Toleva-Stoimenova 2015). D'Ignazio and Bhargava describe the concept big data literacy as follows:

- "Identifying when and where data is being collected
- Understanding the algorithmic manipulations
- Weighing the real and potential ethical impacts" (2015: 2).

Talking about big data literacy reveals the connection with media literacy: "Media literacy – indeed literacy more generally – is the ability to access, analyze, evaluate, and create messages in a variety of forms" (Livingstone 2004: 5). People are enabled to deal with media, and such a description fits big data literacy. Big data literacy is about these competences, and the task of big data literacy is to train employees in such a way that they are capable of dealing with big data. The HR department has the capacity to encourage their development. By means of human resource development, people can be taught big data literacy. This training will tackle computational thinking (Wing 2006), statistical thinking (Hoerl & Snee 2012), and skeptical thinking (Sagan 1996). The goal is to empower the employee to open the black box and lift the curtain behind the big data magic. As Clarke (1977: 35) stated in one of his three laws, there is the tendency to understand such complex and opaque technology as magic: "Any sufficiently advanced technology is indistinguishable from magic". Big data contribute to the veil of ignorance (Rawls 1971) within organizations. In order to be able to deal with the task, the HR department needs extensive training in computational thinking (data farm), statistical and skeptical thinking (fog of big data), and utilizing their HRM and ethical training (big data watchdog).

The prime goal of HR development is to lift this veil of ignorance so that employees understand the use of big data within organizations. Employees also need to be trained in a way that means they are also capable of tinkering with the existing data, and exploring on their own. Achieving this goal will be done through training and development in big data competences. This also includes the HR department, and as Priestly stated precisely, "We're all data geeks now" (2015: 29). It will be essential to lift all employees to a level that they understand and use big data analytics (Davenport 2013) as well as being critical of them (Boyd & Crawford 2012).

Depending on the big data literacy within organizations, there will be a tendency towards convergence or divergence, and standardization or individualization. Employees need to be capable of dealing with big data. John Draper, aka Captain Crunch, coined the term "Woz-Principle" (Freiberger & Swaine 1999), derived from

148

an idea by Steve Wozniak. It suggests that as many people as possible are trained in using technology to the extent that they are capable of inventing new things. At best, technology is as simple and open as possible. From this it follows that people are empowered to design their working environment for their specific needs (Baumgärtel 2015). Such trends can be seen in the open source communities, hackers, or in gaming. By empowering employees in the sense of the Woz-Principle, the HR department will transform the operation system or the HR daemon of the organization and individualize the working environment or user interface for every employee. Employees will be able to customize their working environment for their specific needs. This could lead to a realization of the following statement: "Making people think is the best that a machine can achieve" (Gigerenzer 2015: 320). That means that the goal of HRM is to enable people to have an intrinsic "desire to exploit the information capacity of the new technology" (Zuboff 1988: 392).

A critical issue is that people tend to have a certain amount of technophobia (Brosnan 2002), anxiety (Beckers & Schmidt 2001), and a fear of coding (Spinellis 2001). Although it may sound promising to follow the Woz-Principle in training and development, and beneficial to training the data geeks on their own, the focus lies in convincing people to learn to code and to use statistics. The HR department is the leader in the role of transforming organizations to enable people to design their own tools. It is responsible for a balance of user-friendliness and for the ability to tinker. The essential task is to convince people to acquire the basic abilities of computational and statistical thinking (Dasgupta & Resnick 2014). Empowering people through the Woz-Principle will let them think, create, and innovate in a way that leads to a prolonged competitive edge for any organization.

This means that all employees need a rudimentary training in computational, statistical, and skeptical thinking. Big data influence all decisions within organizations and employees will face big data on a daily basis, but big data are complex by definition, so organizations are transparent concerning their analyses; employees who do not understand the consequences will be skeptical and deny the use of big data (Shah et al. 2012). Ignorance may be bliss, but only with improvements in big data literacy the effectiveness of big data can be improved for organizations and for every employee.

There is also a new layer of complexity concerning learning and development. Today's big data algorithms are no longer mere tools (Varian 2014); they are learning as well. Machine learning (Goldberg & Holland 1988) and deep learning (Deng & Yu 2014) are standard parts within algorithms. Algorithms learn on their own and, most importantly, change on their own (Gillespie 2012) – and, if not watched, become unintelligible to humans (LaFrance 2015). This means there is a dependency, or even duality, between human learning and machine learning. Human learning and machine learning are also within a feedback loop, and are (negatively speaking) in a vicious cycle or (positively speaking) a co-evolutionary loop. This is similar to the idea of the red-queen hypothesis (van Valen 1973), in which both sides challenge each other to improve, adapt, and learn. The function of developing and training is no longer limited to people, but includes algorithms as well. This is especially

important since algorithms can learn erroneous things (just as humans can), but algorithms are not capable of judging them. Algorithms can, therefore, develop ideologies (Mager 2012) and subsequently create reality. HR development and machine learning will merge into one function within organizations in the future. People are trained and algorithms are trained. Both are constantly working together, so they influence and learn from each other.

The HR department is the expert in training and development, they are also capable of dealing with resistance to change (Dass & Parker 1999) and convincing people (Armenakis et al. 1993). It becomes increasingly important to train the employees in big data literacy, not only to achieve some form of transparency, but also to harness the possibilities of big data. "Data is useless without the skills to analyze it" (Harris 2012). Big data will unfold all capacities if people are taking advantage of the potential. The borders between HR development and machine learning are also dissolving and training for algorithms as well as for employees will help employees to work better with big data.

Gamification of HR development

There is a current observable trend not only of *gamifying* work (Oprescu et al. 2014) but also of gamifying HRM, by, for example, incorporating video game design elements into HRM processes. Gamification (Hamari et al. 2014) is often used under the premise that gaming is fun and engaging. Players trying to win a game are highly motivated to reach high scores. This is of particular interest for managers, which makes it understandable for HRM to jump on the bandwagon of gamification. There are several definitions of gamification: "the process of game-thinking and game mechanics to engage users and solve problems" (Zichermann & Cunninham 2011: XIV) or "gamification refers to: a process of enhancing a service with affordances for gameful experiences in order to support users' overall value creation" (Huotari & Hamari 2012: 19). The most commonly cited definition reads: "gamification is the use of game design elements in non-game contexts" (Deterding et al. 2011: 1).

It could be interesting to gamify HR development, especially as learning curves are an integral part of any game (Rosser et al. 2007), however, using a gamification system 'off the shelf' will be a source of irritation (Bogost 2014) and will lead to resistance (Deterding 2014). Big data may help to make the system fit organizations, and the gamification system will stay static and finite (Nicholson 2012). Interestingly, video game developers already utilize massive amounts of data to understand their players and adapt their games to their player base. People have different interests and different skills, subsequently, this diverse player base will influence the way they are playing the game. Video game designers act on the knowledge they acquire and design the most fitting experience for these players,

---

so they stay within the game and play it. A game, and massive multiplayer online games (MMOs) depend on a dynamic development of the world to keep their players bound to a particular game.

Big data within video games like 'World of Warcraft' help to individualize the experience of any player and keep the player within the flow (Csikszentmihalyi 2010). The learning curve can especially be individualized. Players learn new elements of the game at their individual speed, so they are not overburdened or bored. The bar is constantly raised (Scholz 2015c). The game conveys a sense of mastery (Nicholson 2012) and enables players to narrate their own story.

Such a concept can be transferred to HR development for big data literacy and the HR department can implement and cultivate such a gamification system within the HR daemon. People will learn about big data literacy in a playful way, and, thus, lower their big data phobia. They will learn at their individual speed, but will learn to become better equipped to deal with big data. Such a system can be designed in a similar way to the tools within the HR daemon and train employees to program their own tools, following the Woz-Principle.

4.2.4 Human Resource Centaur

The HR department reacted to the transformation towards a homeodynamic organization through big data with a new role and created the HR daemon. Both actions require us to deal with big data and the impact of big data on an organization. However, both are still more reactive than proactive. In chapter 2.4 it became evident that the reaction to big data can be polarizing, however, it seems that an augmentation of both worlds would be beneficial for the organization. Big data will not lead to a competitive advantage, but the people augmented by big data will be the source of competitive advantage. Consequently, it will be essential to enable the workforce to exploit big data and augment them by using big data. Until now, big data have changed the role of the HR department and the way it works. In addition, while big data are now everywhere in organizations and immersed completely into them, big data are somewhere in the background and are something that seems to have no direct connection to the employees.

The goal is, therefore, a way to put big data into the hands of the employees. The HR department's task is to design a frontend, in which the employees can interact easily with the HR daemon and the available big data within organizations. The goal is to give the employee a "'cockpit' interface on their computers that they help design" (McDonald 2011). The idea is similar to the concept of augmentation described by Davenport and Kirby (2015) as "starting with what humans do today and figuring out how that work could be deepened rather than diminished by a greater use of machines". Augmenting people with big data depends on the system that is implemented, and this frontend system I will conceptualize under the term *HR centaur*.

Why a centaur? Looking at the evolution of chess, it is well known that Deep Blue beat Kasparov in 1997. Today the best players follow the concept of centaur chess. Human and machine team up and augment each other in an extraordinary way, superior to human and machine alone. "Centaur chess is all about amplifying human performance" (Cassidy 2014). Such a collaboration of human and machine, as observed in chess, has proven to be far superior to playing alone (Ford 2015). Humans can focus on their creative and innovational roles, delegating the grunt work, or at least the operative tasks, to big data tools. Big data can aid and will help "human beings think smarter" (Kelly 2014). Collaboration, "if we handle it wisely, [...] can bring immense benefits" (McCorduck 2015: 51).

In a popular song by Daft Punk, the band sings about "work it harder, make it better, do it faster, makes us stronger", and this metaphor is strikingly fitting to the modern world that is enhanced by big data. Big data enable organizations to gain access to an abundance of data and use them for their purposes, however, most organizations drown in the glut of data (Emerson & Kane 2013), and are surrounded by an opaque data fog. It is, therefore, one of the most important tasks of HRM to deal with big data in an efficient way and build a sustainable infrastructure. Gaining a competitive edge or even a competitive advantage out of big data use is a more strategic challenge. People are augmented by big data. They can *work it harder*, as they can specialize in their competencies and use their capabilities efficiently. They can *make it better*: they have a different point of view and so see problems and obstacles that the other would miss. They can *do it faster*: dynamics and velocity are crucial for the success of modern organizations. Working together, division of labor is more precise and, synergies are used in a more fitting way. This *makes us stronger*: such an organization is more capable of tackling a situation in its environment. It can adapt to new challenges and govern the risks surrounding them. Tinkering and performing with virtuosity will lead to the essential competitive edge any organization needs. The HR centaur needs to act as a multipurpose tool kit (Zuboff 2014), to enable people, especially as:

> "[M]achine intelligence does not lower the threshold for human skills – it raises the threshold. Whether it's programmed financial products or military drones, complex systems increase the need for critical reasoning and strategic oversight" (Zuboff 2014).

The HR centaur will augment employees to be able to deal with the increased threshold and give them all the essential tools to exploit big data to a potential competitive advantage. One way to implement such an HR centaur system is to reevaluate the potential of gamification and learn from video games. Big data are, per se, digital, so the link between big data and people is digital as well. Gamification and video games are normally embedded in the digital realm and, therefore, there are many ways to learn from those digital pioneers. I have already defined gamification in a previous chapter. Although gamification can contribute to the HR centaur and there is the potential to increase transparency, individualization, and strategic agility (Stein & Scholz 2016), such a system would be predominantly designed by the HR department. They will act as gamification designers (Raftopoulos 2015), and will

152

constantly update the system to fit the needs of organizations. Apps will be built that will make employees transparent (Buchhorn 2015), however, the HR daemon and the engine behind these will stay shrouded. There are many ways to analyze employees and give them information back about their work. Various components of a game can be translated exactly towards such HR centaur software (Scholz 2013c). For example, a talent tree can show an employee what things are available to learn and what specific programs fit the current job. People can be matched as teams on the basis of their ELO scores (Erhardt 2016), a method to calculate the skill level and rate people on the basis of the score, and so on, but these systems are one-directional from the HR department to the employee. It seems that big data will act as a bridge between video games and the real world; for example, metrics and indicators used in video games are more and more available in the real world due to big data. Hocquet (2016) described this bridge on the case of Football manager video games and the increasing entanglement to the football world and the datafication of the football world. The challenge will be to create a HR centaur system that will be designed by the HR department and the employee, thereby following the Woz Principle in the truest sense.

Radical Gamification

But what would such a system look like? In a conference paper, Stein and Scholz (2016) envisioned a concept of a radical gamification. The following example is derived and adapted to the context of this thesis from the conference paper.

Contrary to most gamification within HRM, which can be characterized as *casual* gamification, a *professional* gamification with proper design, intensive planning and careful coordination could tackle existing problems in a new manner. In order to avoid "gamification is bullshit" critics (Bogost 2014) and the reproach of engaging in pure 'gamewashing' (in analogy to "greenwashing" in the corporate social responsibility debate (Dahl 2010)), I will present a short example of a proposed HR centaur system.

Radical gamification of HRM would best be possible in an organization with a non-existent or underdeveloped HRM function (although that does not mean no HR department) and a workforce open to change. They are also able to gather big data and have a rudimentary understanding of the HR daemon. An organization of that kind would be a start-up, mainly consisting of "digital natives" (Prensky 2001) with basic programming competences, and in a field that allows them to gather data digitally: an IT startup. Cultural elements of gamer culture (Shaw 2010) and hacker culture (Levy 2001) would be beneficial. Only under those conditions would employees be intrinsically motivated to participate and to increase the *gamification rate*. They have also the ability to utilize big data in an efficient way and, therefore, the organization will have an interest in using big data in such a way.

The radical gamification of HRM starts with the basic idea that all the HRM functionalities that are needed by employees or by management are to be developed bottom-up. Everybody is entitled to write add-ons such as apps and to modify those emerging functional worlds (Sotamaa 2010) as needed. The *look and feel* of that gamified HRM, then, imitates the design of a sandbox game like 'Minecraft', where the players can do whatever they imagine: a holiday scheduling system, performance measurement, monitoring presence, multi-project management, a team task assignment support tool – the possibilities are endless. Incentives can be coupled with gamification contributions. No longer being a traditional HRM department, a HR gamification designer will be given the task of supervising the gamification system and simply acting as a corrective. Everything else will follow the market principle and the logic of self-organization. An employee in need of a specific functionality simply *buys* an existing tool from the market or programs it autonomously. The lack of a distinct functionality can be understood as a strong indicator that there is simply no need for it in the organization.

Such radical gamification is scalable and develops concomitantly with an organization's growth. The integration of the employees who need to acquire competences in utilizing and developing *their own* HR centaur system is crucial. Issues such as relative fairness will be tackled so that people cannot cheat or exploit the system. It interferes, but never with self-organized teamwork or competition. Nor does it cancel them out.

In a system of that kind, employees shape their range of HRM and at the same time live a gamification culture to its fullest. The HR gamification designer merely acts in the background supervising the people-related *engine* or the HR daemon of the organization. Stenros (2015) talks about second-order design backed by Salen and Zimmermann: "As a game designer, you can never directly design play. You can only design the rules that give rise to it" (2004: 168). Leveraging transparency, individualization and strategic agility will benefit employees and the organization – both mutually increasing their ability to make homeodynamics work.

4.2.5 Big Data Membrane

At this point the homeodynamic organization is fully implemented due to the integration of big data within the organization. The HR department has reacted with new roles, created the HR daemon, and proactively augmented its employees through the HR centaur. But these changes also lead to a high transparency concerning big data and potential critical information about the organization. If everybody has access to most of the data within the organization, keeping the data within the organization will be complicated. I postulated that big data plus people will generate a competitive advantage, but a competitive advantage which is only possible if this knowledge is kept a secret. The problem is that big data are truly everywhere. Big data seem to be unbounded and free floating. It is, therefore, essential to have

ways to protect personal data from the outside world, such as personal data about employees, data about the organizational signature, and data that describe the competitive edge of an organization.

In nature, there is a semipermeable membrane that selectively allows an exchange between the outside and the inside. Such a big data membrane would be capable of deciding what data are shareable and what data are critical and, therefore, will not be shared under any circumstances. Some data can be exchanged freely; others will be kept within the organization. In the context of an organization, this is comparable to open innovation. Chesbrough (2006) explicitly points out that the only innovations which are shared are those not critical for the competitive advantage of an organization.

One way to deal with data sharing is to focus on the membrane and improve the selection of such a membrane, however that implies a critical reflection of what is valuable and what is not. Big data are known for being vulnerable (Newman 2015b) and it may be beneficial to keep critical data in one place rather than outsource them to the cloud (e.g. Kraska 2013). Big data, thus, need to be encrypted and people need to be trained to follow the encryption rules. Both elements can be achieved through the HR department, however, every encryption is breakable. In a recent paper, Zyskind et al. (2015) linked the protection of big data to the concept of a block chain.

> "A block chain is a type of database that takes a number of records and puts them in a block (rather like collating them on to a single sheet of paper). Each block is then 'chained' to the next block, using a cryptographic signature. This allows block chains to be used like a ledger, which can be shared and corroborated by anyone with the appropriate permissions" (Government Office for Science 2016: 17).

In addition to achieving a certain transparency and traceability of big data, which is beneficial for big data within organizations as well, the data are protected in a relatively strong way (Swan 2015). Organizations will be able to encrypt their organizational signatures and will have a ledger of all changes. The ledger decreases the risk of manipulation from inside and outside. Employees know that their personal data are encrypted and their personnel file is completely transparent to those people that are allowed to see it. A block chain can be used to secure it from the outside and make it less susceptible to manipulation. It is, furthermore, a way to identify changes within the big data. Additionally, it enables organizations to have a form of time machine and retract changes (interesting for the data farm) as well as reconstruct after manipulations. A block chain would not be able to prevent corporate espionage, however, it would just make it more difficult for anybody to steal information. Although the concept may sound a bit futuristic at the moment, block chains will influence organizations and radically transform them. Tapscott and Tapscott are using the words "agility, openness, and consensus" (2016: 90) as well as "decentralization" (2016: 91). They are, furthermore, talking about the importance of the code of the block chain system that seems comparable to the organizational signature.

Another issue involves the information that is sent through the membrane to the outside environment. In the common practice of competitive intelligence (Kahaner 1997), for example, it becomes increasingly less difficult for organizations to gather

all the information available about their competitors. This information makes organizations more transparent from the outside, though organizations can proactively work against it. Big data enable organizations to gather external information as well and can reconstruct the picture competitors have of them. An organization can then play proactively and spread either correct information or false information, altering and distorting the picture the competitor has. It is, thus, possible to improve the protection of certain knowledge that is linked to competitive advantage, by flooding big data into the outside world in order to mislead the competitors.

Both acts are enabled through technology and big data, but it is essential that they are not driven by big data. Big data are shackled by the computational logic and, therefore, susceptible to being de-encrypted. This computational logic of big data needs to be transformed into an irrational protection system and the people within organizations must be capable of supplying irrationality. These critical data can be easily translated into a gamified system. People can be sent on missions to hack the system from within (in a secure environment) or sent outside to spread misinformation. Due to the possibility of obtaining a *precise* picture of organizations from the outside, employees will see the effect of their work and be intrinsically motivated to keep certain elements a secret and to spread other information.

Apple Car

There is a hypothetical case about an Apple car by Shen et al. (2011), where they discuss a potential extension into the automobile sector. Today, there is still no official information regarding an Apple car, but many people (e.g. Harris 2015, Hotten 2015, Jones 2015) seem to know that Apple is working on an automobile. We are, therefore, talking about a hypothetical case, but it reveals the existence of a big data membrane around Apple.

There are many data on the Apple car and many rumors on the car, however, it seems that Apple is, in a way, directing the information. They are using this leaked news to get something out of it. Take another example: it seemed that everybody knew that Apple was working on a television. Apple never talked about the project, but it got a clear picture of the chances of such a device on the market. Without ever openly talking about the project it realized that there is potentially no profitable market for such a product. In the case of the Apple car, CEO Cook made it clear that they would not comment on the Apple Car, but at the same time he teases people about it:

> "Yeah, I'm probably not going to do that. The great thing about being here is we're curious people. We explore technologies, and we explore products.

> And we're always thinking about ways that Apple can make great products that people love, that help them in some way. And we don't go into very many categories, as you know. We edit very much. We talk about a lot of things and do fewer. We debate many things and do a lot fewer" (Cook in Lashinsky, 2016).

In times of big data, information is ubiquitous and organizations cannot control all information concerning their organization. Information is leaked and rumors are spread, however, those rumors can be steered and governed. Media companies have focused on Apple concerning the Apple Car project and harvested a great deal of big data to gather new information. Apple is not capable of producing a car on its own, and therefore have an army of subcontractors and suppliers. What technologies hide behind Apple's current products (e.g. batteries)? Apple monitors information about the Apple car and sees where the information originates. The media is currently monitoring Apple's recruitment efforts, leading to a list of potential project members (Kahn 2015).

This is all especially important, as Apple is always mysterious about its new products. Although it is still unclear whether there is a car in development, Apple can focus on developing a car and everybody else can speculate about the chances of this car on the market. This is relatively cheap market analysis. They get to know what customers want and what they dislike. Even if they put the car in mothballs, they have probably improved the battery technology of their laptops, tablets, and smartphones.

Big data increase the risk of losing a potential competitive advantage and make organizations more transparent than ever before, although it seems that organizations can steer the data stream to a certain degree. If organizations invest resources in the big data membrane, they will be capable of exploiting this apparent weakness. They can convert it into a strength by utilizing the abilities of big data and people combined to the fullest, so in a certain sense, we are talking, in analogy to centaur chess, of *centaur intelligence*.

4.3 Homeodynamic Goldilocks Zone

The complete implementation of big data within any organization through the new roles of the HR department, HR daemon, and HR centaur, will enable the homeodynamic organization to be more dynamic and consequently, capable of gravitating around a certain form of balance. Big data are always about a strategical decision between polarities, as shown in the core dimensionalities mentioned earlier, and the positioning between those polarities. Organizations constructed in a dynamic way will be able to correct their course, and though as they are complex systems, small changes may have big impacts and oversteering due to time-lag is always a possibility (Liu et al. 2011, Diesner 2015). This is true especially as big data within an organization will make the organization potentially faster, but real-time remains an illusion (Buhl et al. 2013). A homeodynamic organization will probably not be able to achieve perfect homeodynamic balance, but will keep organizations close, especially as it is not necessary to balance everything out exactly.

Organizations need to be in the right zone, that is the so-called 'Goldilocks zone'. The term is derived from the story of 'Goldilocks and the Three Bears', in which

a girl searches for things that are "just right" (Spier 2011: 148). The term emerged and gained popularity for describing the zone of solar systems in which planets are potentially habitable (Kasting et al. 1993). The concept proposes that planets need to range in a certain zone of variables to be habitable that may depend on distance from the sun, luminosity of the sun, size of the planet, certain elements (e.g. helium) available in certain amounts and so on (Lineweaver et al. 2004). A homeodynamic organization will also be stable within a certain *homeodynamic Goldilocks zone* as shown in Table 15.

Table 15: Positioning of the Homeodynamic Goldilocks Zone

Polarities of the Core Dimensionalities in a Homeodynamic Organization		
Data Linearity		Data Monodology
Data Rigor		Data Swiftness
Data Island	Homeodynamic Goldilocks Zone	Data Assemblage
Social Constructivism		Data Constructivism
Data Risk Avoidance		Data Risk Seeking
Social Shadow		Data Shadow
Self-Determined		Data-Determined
Data Reliance		Data Bias

Organizations need to find a way to deal with changes and evaluate influences on their position within the zone. With the HR daemon, the HR department, and the HR centaur, organizations are capable of keeping themselves 'just right', however, there are several constraints that will be incorporated into the calculations. These are (1) the organizational signature, (2) the trust climate, and (3) the rate of dynamization and the complexity parameter. The organizational signature is the core DNA of any organization, therefore it will not be changed all the time. Consequently, the organizational signature seems to be a fixed influence on homeodynamic organization. The trust climate is critical and influences the reaction time of an organization. Without trust in the HR department or in big data use, there will be resistance and distrust. Changes will not be implemented and organizations will drift into a lock-in situation, and depending on the current situation, potentially move outside the homeodynamic Goldilocks zone and "fall apart completely" (Spier 2011: 148).

The next constraint is the degree of dynamization and complexity. People tend to prefer a static, orderly, observable, and linear environment, but reality resembles the opposite (Maguire et al. 2011). In order to categorize the facets of dynamization, Stein (2015: 3–4) presented the following:

- More dynamic in the strategy-related sense of 'more differentiated'
- More dynamic in the mechanics-related sense of 'faster'
- More dynamic in the organics-related sense of 'more versatile'
- More dynamic in the culture-related sense of 'more strategically agile'
- More dynamic in the intelligence-related sense of 'more methodologically competent'
- More dynamic in the virtuality-related sense of 'more flexible'

Although it makes sense to improve the dynamization, there is tradeoff in the sense of complexity. Homeodynamic organizations are complex systems and big data increase the complexity furthermore. The elements of unpredictability, non-equilibrium and non-linearity in particular (Maguire et al. 2011) lead to unexpected threats, opaque and secondary effects, and uncertainties within a complex system (Dörner 1989). Organizations will deal with those potential risks.

In a nutshell, it is possible to keep organizations within the homeodynamic Goldilocks zone, however, it is a complex task that is supported and disturbed by big data. Depending on big data use, and, therefore, depending on the HR department, organizations gain the ability to remain within the zone. It sounds like a difficult task for any organization and a costly project to transform an organization into a data-augmented homeodynamic organization, but it will increase the survivability of any organization drastically. People will become *the* competitive advantage of organizations and they will transform big data into something more than just the standardized tools many organizations are currently using: although it is expensive, utilizing big data in this extensive way allows the management to have a precise view of their organization. People are no longer an opaque cost pool, but their contribution can be accounted for. At the very least, utilizing big data will be beneficial for all employees, as it allows everybody to focus on their strategic and innovational input, and delegate the operational and automatable tasks to the HR daemon.