Decision-making in the Era of Big Data

Minha Lee
minha.lee@student.uva.nl
University of Amsterdam
Information Science : Human Centered Multimedia

Abstract—As data sources that a decision maker (DM) is faced with increases exponentially, what becomes more significant is the relationship between information sources and the transformation of information as it travels through various phases in and outside of an organization. The value of information is based on data interactions of all inter-dependent organizations and their DMs for an evolving, empowering, and empathetic data environment as context. This paper examines literature on decision-making in an organizational setting to identify how big data will impact decision- making. DMs who regularly face big data are interviewed to devise an evolving information infrastructure to accommodate data-driven decision-making.

I. Introduction

An organization's data environment lies in the emerging inter-reliant interactions of decision makers (DMs) as they continuously search and satisfice [20]. Data analytics for data driven decision-making [10] in organizational settings does not necessarily result in making more effective decisions. This is due to DMs' limited capacity in information processing, divergent goals and contexts in organizations, and information decentralization [17] that bounds decision-making processes. To capitalize on profitability and insight that big data can deliver [10], the critical point is that organizations as "groups are social systems" [11, pp. 428] in which "individual learning in organizations is very much a social, not solitary, phenomenon" [21, pp. 125].

The challenge is that decisions will be made with high levels of uncertainty for all interconnected DMs, which results in "satisficing". No more alternatives are searched for as soon as a choice that matches a DM's level of aspiration is found [20]. Thus, data-driven decision-making must rely on statistical models for data analytics [3] whilst recognizing bounded rationality [21][17]. DMs need an environment that facilitates big data management [10, 13] that simultaneously provide the means to handle human inconsistencies and biases [5] [15] which originates from group interactions within an organization.

As shown in Figure 1, a common approach to data management currently sees data filtering and dissemination as a one-way processing that leads to executable communication. This is only a technical solution for data analysis in order to manage the sheer amount of accumulating data for organizations. Information infrastructures such as Figure 1 disregard the importance of individual DMs as creators and managers of their own data environment. In the contemporary information society, technological solutions that satisfy DMs' information needs are abundantly explored. Yet, incorporating DMs' perspectives to reflect on how to best assist in decision making with big data has to be further developed to revise Figure 1.

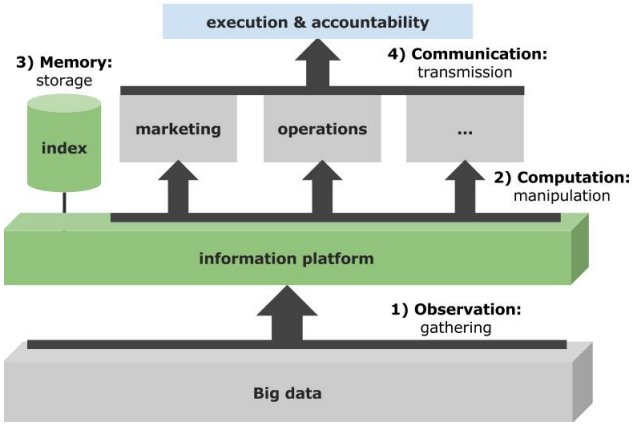

Fig. 1. Information flow

Therefore, the research question that will be addressed is:
what are the conditions for an organizational information infrastructure that support a DM's data-driven decision-making process?
The hypothesis is that the conditions for an organizational information infrastructure that supports a DM's decision-making process is data filtering that is socially enhanced in information communication. The data dimension as an environment dynamically changes based on information interactions of all interdependent organizations and their DMs. The hypothesis was gained from previous works by [10, 13], [20, 21], [11], [18], [17, 16], [6], and [15]. It will be assessed through semi-structured interviews with DMs from the broad field of information technology who regularly interact with big data. The goal is to identify contextual needs of DMs that are unmet by the existing information infrastructure (Figure 1) they operate in.

The paper proceeds first with a literature review to establish a background that the hypothesis stems from. Then, the interview setup and results are discussed. The interview findings will be refined as an evolving model with its implications extrapolated. Final thoughts are shared in the conclusion and future works sections.

II. Literature Review

Relevant works on big data, organization, decision-making, uncertainty, and for decision-making are examined.

A. Big data

Big data adds exponentially increasing alternatives of states that DMs must operate within in relation to increasing amounts of data. To illustrate the magnitude of big data, about 2.5 exabytes of data are created every day and this amount is doubled every forty months as of 2012 [13]. There are countless different formats that data comes in (variety) and the speed of data creation is continuously increasing (velocity), as complemented by the increasing amount of data creation (volume) [13]. Also, quantification of data involves individually complex data types, which means standards and formats for storage and delivery of data can involve many different, complex solutions [12]. Organizations that prioritized data-driven decision-making were five to six percent higher in productivity than originally estimated when considering their investments and use of IT [10]. An important aspect on a data environment is that as of 2013, around 80 percent of data that an organization needs is spread out through the supply-chain cycle and partner organizations [3]. Thus, gathering all relevant information and maintaining shared data across the chain to create an optimal decision-making environment is a challenge that currently remains largely unresolved.

B. An organization

An organization as a group is "an intact social system, complete with boundaries, interdependence for some shared purpose, and differentiated member roles" [11, pp. 429]. Hackman conjectures that the focus should not be on the traditional cause and effect model of organizational action and performance, but rather on the conditions in which groups proceed with their actions [11]. Given the advent of big data integration in businesses, a DM's context is a social system of decentralized processing of numerous information contingencies and decentralized decision-making[17]. This is further complicated since an organization's context is dynamically changing with global distribution of members, tasks, shifting roles, and with DMs often working for more than one organizational context and agenda [11].

C. Decision-making

Decision-making with uncertainty according to the Savage Paradigm is defined as a series of alternatives in 1) states of the world or states that a DM has no control over, with 2) actions or acts that a DM can take, resulting in 3) consequences [17][18]. A DM would choose an act that results in a consequence with a higher expected utility from all possible acts to follow. Taking into account resources needed to follow the Savage Paradigm, Radner enumerates the costly activities (CAs) as 1) observation: information gathering, 2) computation: information manipulation, 3) memory: information storage, and 4) communication: information transmission [16][17] as shown in Figures 1 and 2. These activities are costly because they cause delays in decisions and decrease effectiveness of decisions [17]. This effect is known as the "iron law of delay", for as data to be processed increases, so does minimum delay due to the bounded computational processing power at any given time [16][17]. For an organization, the problem is in the inevitable decentralization of DMs' roles with unclear separation of information ownership and management that adds to the iron law of delay [16]. Furthermore, following the Savage paradigm consistently is beyond the capabilities of DMs, for they operate with bounded rationality, which is defined as "the limits upon the ability of human beings to adapt optimally, or even satisfactorily, to complex environments" [21, pp. 132]. A complex environment can be related to uncertain states of the world in the Savage paradigm for decision-making.

D. Uncertainty

A DM faces "unsureness" when deciding on the act to follow due to the uncertainty in knowing the set of all possible relevant states in the first place [17]. One way to reach a best possible act is to use probabilities of second order with a DM comparing alternatives to arrive at a decision of best possible fit given the circumstances. But, Savage points out that this method is ineffective since when "second order probabilities are introduced, the introduction of an endless hierarchy seems inescapable. Such a hierarchy seems very difficult to interpret..." [18, pp. 58]. Uncertain data environments bring an endless hierarchy of possible actions that DMs cannot effectively interpret in reality.

Thus, DMs face an increasing uncertainty in what information to gather and integrate, but especially DMs are uncertain about the underlying inference of the information they have or search for [17]. This is called "truly bounded rationality" in that DMs cannot effectively infer decisions to follow with information they have, even if relevant information is found [17]. DMs are truly bounded in rationality because they do not know the implications of what they know [17]; big data brings uncertainty in what DMs are already uncertain about.

1) Inconsistency: Psychologists such as Meehl have established that statistically driven systems have fared better than human decision makers in predicting future outcomes [15]. Out of 90 findings from a wide range of fields comparing the predicted outcomes of human DMs and computer systems, it was difficult to find even six studies that showed at least a weak support in favor of the majority of human DMs under evaluation [14]. Meehl's reasoning on this discrepancy is that "there are no strong arguments[...] for believing that human beings can assign optimal weights in equations subjectively or that they apply their own weights consistently" [14, pp. 372]. This approach is furthered by Dawes' work on improper linear models; when the weights are applied in a non-optimal fashion, be it random or based on intuition, they still perform better than solely clinical

intuitions [9]. According to Dawes, DMs excel in knowing what information to look for, but they do not excel in integrating all possibly relevant information with consistency [9].

2) Biases: The benefits of data analytics as a part of an organizational decision-making process has been recognized in business settings [6]. Yet, since more information adds complexity, more information can bring overconfidence that leads to less accurate decision-making. The DMs can inaccurately decipher information due to their pre-existing cognitive biases, as in the case of venture capitalists [25]. Organizations that effectually aimed to decrease biases, such as overconfidence and groupthink mentality, during the decision-making process as aided by data analysis gained a return on investment of up to 6.9 percentage points higher [7][6]. This shows that data analytics for accurate and timely information is not enough in a decision-making process. An effective use of filtered information is dependent on DMs' interpretations of their data environment whilst recognizing established biases.

E. System-supported decision-making

An early example of a system that outperformed its human counterparts was called MYCIN, a decision support system that gave advice on the diagnosis and treatment of infectious diseases [24]. MYCIN made headways for other decision support systems (DSS), yet the problem is that systems mostly target specific problem-solution sets for specific domains.

Contextualization for DMs is difficult since there is not one specific domain that a system can focus on; an organization is often involved in a variety of domains, internally and externally. Under the banner of knowledge management, larger enterprises have focused on search systems that connect involved members with the intention of having employees readily share information with each other.

Enterprise search systems or expert finding systems have been used since the 90's with organizations like Hewlett-Packard, Microsoft, and NASA sharing their explorations on this topic publicly [19] [8]. For example, a self-maintained and web-based directory of experts at HP called Connex was created in 1997, which connected people within HP based on attributes such as skills, interests, expertise, and knowledge as a globally deployed knowledge management system [4]. Systems that support such processes exist and are already in use like relationship capital management tools by BoardEx [1].

Along with enterprise search systems that connect people, an organization's full body of knowledge has been sought to be easily retractable with data integration methods [22]. However, the benefits gained from data integration, enterprise search, and DSS are not being fully exploited. This is due to not only technical challenges, but also because a DM's data environment has not yet evolved to introduce the concept of learning from data instead of the need to control data. Figure 1 exemplifies this challenge with a one-way data stream imposing control over the information flow rather than treating it as a dynamic environment for interactive information discoveries.

[1]http://corp.boardex.com/

F. Summary

The literature review reveals coming changes in an information environment for DMs. Big data in organizational decision-making adds greater uncertainties that DMs cannot effectively construe due to DMs' bounded rationality alongside factors such as human inconsistencies and biases. DMs' perspectives have to be integrated in foreseeing the direction that the information infrastructure will mature in. Thus, ways to improve the current model of information flow in Figure 1 has to be identified by interviewing different types of DMs who regularly engage with big data.

III. Interview setup

A semi-structured interview is defined as an interview in which an interviewer starts with a series of questions that are general for all interviews, but the sequence of questions can be varied [2]. This was chosen as an approach for participants' viewpoints to openly develop. Seven interviewees from the broad field of information technology participated. Three test interviews were conducted beforehand with those in the managerial positions at Hewlett-Packard of the Netherlands. All participants were males between the ages of 40 to 60. Four participants were based in the Netherlands, two participants were in the U.S. and one participant was in Ireland. All interviews were in English.

The participants were contacted via snowball sampling through the CTO of HP Netherlands. They are divided into three categories of innovation strategy, government, and account management. These three categories represent selective groups of DMs who interact with big data daily in varied ways due to their wide-ranging roles and organizational associations. The participants were interviewed in following order: three pursued innovation strategy in HP (I1, I2, I3), two from the Dutch government with one as an IT representative of the judiciary (G1) and the other as a CIO of a municipality (G2), and two were account CTOs of HP (A1, A2). Three interviews with I1, I2, and I3 were conducted via audio chat and four interviews with G1, G2, A1, and A2 were conducted face-to-face at their respective workplaces. All interviews were between 40 to 55 minutes long and were transcribed verbatim.

Questions related to Radner's CAs for decision-making [17][16] were asked. Topics were introduced in following order: 1) decision-making processes, 2) observation: gathering information and sufficiency of information, 3) computation and memory: collaboration and automation of processes, 4) communication: privacy issues, format and content of information, search, and 5) guidelines and future outlook.

IV. Interview results

The section illustrates findings on how decision-making can best benefit from increasing amounts of data. The

participants' current data interactions and viewpoints are elaborated on to arrive at key aspects that will be used to form a user-driven information environment. Aforementioned Radner's axioms for CAs in decision-making are followed for analyzing interviews.

A. Decision-making

The conclusive finding is that decision-making processes are in effect for all participants in different degrees of formality, from those in the government adhering to more formal procedures to those working in innovation strategy with more fluid decision-making processes. The account CTOs worked more independently since they focused on their assigned accounts as representatives of their organizations.

Pursuing innovation involves higher risk, thereby making agility crucial (I3). The decision-making steps taken can be broadly understood as first generating ideas by following leads, then brainstorming with team members, and later arriving at outcomes that are different than imagined at the start (I1, I2). These steps are organic since "each new idea comes with its own requirements that do not fit the mold" (I2). Ideas that are further investigated for their innovativeness and applicability become crafted as invention disclosures. Starting with invention disclosures, the process becomes more formal due to the procedural nature of the patent office (I3). Thus, the decision-making process is only formalized at a later stage, when the internal organization prepares for the delivery of the invention to an external organization.

For both G1 and G2, decision-making is formalized based on respective government organizations' operational approach and the existing hierarchies. They are dependent on the board of directors or the council of the judiciary to set parameters and to gain approval for projects. For G2, a triangulation process was in place in his municipality with the three legs being the state, where the IT board belongs to, the management, and new developments. This was seen as an advantage, for the IT board has a chance to have a voice in the decision-making process. The judiciary supports eleven courts and four supreme courts of the Netherlands and has structured decision-making processes, usually referred to as IT governance (G1). A subject to be pursued is placed on the agenda e.g., portal strategy. After many iterations on what the strategy should be, the first draft is constructed, then presentations on the strategy are given with various representatives from internal groups such as civil law or information management participating. Afterwards, an agreed upon strategy is established for the next four years (G1).

For A1 and A2, the organization's hierarchy is less influential. There were no real formalized workplace decision-making processes and DMs fulfilled various roles in consulting, business development, or account management (A2). With these different roles, DMs were "working as a group of individuals" (A2). In decision-making, the prominent influence was in the relationships with clients of the assigned accounts (A1, A2). Understanding the workings of a client's organizational hierarchy, while functioning with the strategic direction of the organization they belonged to, was where their individual decision-making struck a balance. Reaching a consensus with involved parties on a subject matter at hand is a part of a never-ending and evolving decision-making process in adding value to a client's organization and one's own (A1).

B. Observation

The observation stage in Radner's CAs deals with information gathering in a decision-making environment. DMs gather information continuously and for each decision-making context, the sufficiency of gathered information is considered.

1) Information gathering: No distinction was made between internal and external information sources (I1). Since the focus is on new inventions for innovation, a wide net is cast for various sources (I1, I2, I3). All possible information is gathered initially to gain leads and when an idea becomes mature, information is gathered on the potential practical value of an idea from selected sources, such as interested customers with non-disclosure agreements. And lastly, the idea is shared publicly to establish leadership in organization's contribution in innovation (I3).

Account management also requires diverse information sources, but for different purposes. Research is geared toward a customer's industry developments and organizational hierarchy (A1, A2). Information gathering serves relationship building purposes for more than half of the time (A2). This is done to identify and gain contact with business influencers. Linking all information together was reported to be the most time consuming and ongoing task (A2). Representing an organization means that information from various parties within the company has to align as one vision for a customer (A1). This emphasizes the need for validating information with all involved contacts (A1). Gathering information is seen as a balancing act that is never at a full equilibrium between wanting to be an open organization and a secure organization (G1).

Security of information was a concern with the level of classification for internal information always being checked (A1). There are difficulties in dealing with highly sensitive information due to the reality of constant vulnerability to cyber attacks (G1). Also, high level of trust has to be gained with customers, for having access to the mode of operation of a client's organization comes with separate classification needs (G1, A1). To complicate the matter further, organizations often work with competing companies, like Oracle or Microsoft concurrently, and the companies bring their respective requirements for confidentiality and security (G1). To mitigate this problem, idea developments are socialized with partners or customers through non-disclosure agreements (A1, I3). For the IT group of the judiciary, the position that is strived for is to be independent of parties while providing IT service to all dependent parties (G1). When the viewpoints of involved parties are in agreement, decisions that are mutually beneficial are considered (A1).

Data mining is regularly done at a low level in a municipality (G2). Data such as social security numbers are gathered and stored for when a citizen wants to apply for products such as social benefit (G2). The application is digitized and is automatically put in legacy systems. Data on citizens is gathered by the municipality and by non-government organizations like the banks to decide on whether or not an individual receives social benefit (G2). Data analysis on locational statics is regularly performed such as who lives near which locations like gas stations or airports to assess possible risks for events such as emergency evacuations and housing developments (G2). The rules on data collection are federally regulated while the actual collection of data is done locally by the municipality. Sometimes this results in an unequal outlook on how to serve citizens, for the regulations can be outdated or overlook the current information needs of the municipality that is in direct contact with the citizens (G2). Information gathering for maneuvering relationship networks to connect an organization's capabilities directly to a client's business needs or a citizen's welfare needs is essential (A2, G2).

2) Sufficient amount of information: Since innovation is a creative process, investing time and becoming more involved as an idea develops is an evolving process (I2). There is recognition of instinctive awareness of having sufficient amount of information, but this is a part of a continuum (I3). Instinct based on personal experience was important in trying to understand whether or not there is enough commitment from representatives to actualize a project (G1). G1 and G2 were dependent on the workflow system of the government's hierarchical decision-making processes.

Actions were described to be "instinctive based on the information I have...[with] more information better decisions I can make", and the reliance on instinct arises with time constraints (A1). If a suggestion to a client was an assumption, valid reasons for supporting the assumption were given (A1). Searching for a satisficing decision mainly comes with gaining clients' trust in understanding their issues to gain more insight, for instance in how a client's budget might behave, since larger plans are not always shared initially (A2). The aim is to arrive at a decision that might influence the clients' future directions (A2).

Information gathering slows down when no more new sources are found (I3). This is an indication that the search process should come to a stop since nothing novel is being found anymore (I3). Being able to accommodate all information to a sufficient level is difficult with the federal law since all rules are still for physical, paper-based documents (G2). Identifying the sufficiency of information for the digital age whilst working within the paper-based era is a challenge (G1,G2).

C. Computation, memory, and communication

The interview findings indicated that computation, memory, and communication as Radner's CAs will merge as direct and indirect collaborations and will eventually become indistinguishable. DMs' interactions as communication will be computed and stored repeatedly. Therefore, certain formalized processes can be automated. Information content and format preferences will emerge in communication and information search for DMs. The issue of privacy is intrinsic to how DMs conceptualize their information flow.

1) Collaboration: At the core, information technology is all about collaboration (A1). This is seen to be emerging in practice with the culture of any organization also needing to evolve (A2). Collaboration is essential for underdeveloped ideas, for there needs to be an ongoing dialogue from multiple sources to get a variety of information and opinions (I3, I2). Yet, collaboration was seen as a distracting factor for tasks that are repeatable since the goal then is to make a process faster instead of pursuing creative solutions (I3). The most supportive tools are those that assist with editing and brainstorming over files for working digitally (I2). Video chat systems and functions such as screen sharing were often used, but real-time collaboration over files was seen as too immediate without providing ample opportunities for reflection over various viewpoints (I2). The need for a collaborative environment that adapts to a DM's dynamic situations and roles was apparent (I1).

Historically, collaboration tools and DSS were not dynamically aware because they expected static user roles and predefined processes (I1). Contextualized collaboration is seen as an evolutionary step away from providing information that is relevant enough to a large number of DMs to providing information that is relevant for one DM. Contextualization should continuously consider who a DM is with, what a DM is searching for, when it is, and what is happening in the surrounding (I1). There should be no restrictions in what tools or devices to use as it is dependent on user preferences (I1, G1, A2).

The sense of trust regardless of the usefulness of a system has to organically emerge (I1, I2), for secrecy still exists within collaboration environments (G1). Collaboration environments are not as transparent as they can potentially be; communities have to be fostered for effective collaboration (G1). Collaboration is naturally shaped by DMs' existing organizational culture, thus the facilitators have to respect the existing culture (G1). The notion of digital collaboration is still taken to be sending information over emails by many, which is time consuming for all who are involved (A2). A collaboration environment that allows information to travel quicker is needed (A2).

Collaboration tools are taken to be more effective for distributed teams (A1). One reason could be that distributed teams naturally define collaboration dynamics by the environments they gravitate toward since digital collaboration is from the start, the norm. On the other hand, teams with centralized locations that may often meet face-to-face do not necessarily equate collaboration with digitally fostered environments (A1). Which collaboration environments best foster communities while adapting to the changing roles of DMs has to be explored.

2) Process automation: The value of automating processes was seen to be dependent on an individual's context, the pre-existing environment, and the usefulness of automated processes. The overarching point for all participants was that a DM's autonomy needed to be respected. Providing guidance instead of automation was taken more positively since mandated step-by-step procedures give DMs no room for interpretation (I1, G1). The provided processes are likely to be more accepted if DMs can trust the quality of a system's service and if processes make DMs' lives easier by providing efficient ways to consume and share data (I2). Thus, DMs should define their own processes, with automation as a possible option (G1, I1).

There is a thin line between what a system can do for a DM and what a DM needs to do independently, but a system's ability to link information together should be maximized (A2). Many processes for decision-making can eventually be automated, especially simpler processes such as cross-validation of the reported and the actual financial state of a citizen who applies for social benefit (G2). If a copy of the bank statement is not received, the automated process will get the information on the financial state of an applicant from the bank (G2). Other example cases of repeatable processes are shared below.

1) A tool for linking viewpoints of involved DMs together to proceed in decision-making exists for alignment optimization and allows for virtual brainstorming (A1) [2]. This is especially useful since it provides an alignment process for complex decision-making for distributed teams (A1).
2) A process for changing organizational operations to better align IT for business practices has been constructed as BATOG diagnostics [3] (A2). This type of approach is in use to support a customer organization's executable long-term transformation (A2). Each step towards transformation can also come with a set of short-term processes, such as cost estimation, that can be automated if consistently followed.
3) In innovation, processes are defined after the research and development phase of an idea. Specifically, collaborative discussions over risk assessment and mitigation to define a process that will work for operational purposes are necessary (I3). Collaboration with an operations team still occurs in case any bugs need to be fixed while deploying a process. Then, a process can become automated if it will regularly occur for other projects (I3). The quality of the process is evaluated at regular intervals by collecting surveys and feedback (I3).
4) In the judiciary, two million cases are reported to be handled every year and of those, one million are small cases with a DM often dealing with hundreds of cases in a short span of time, which can be greatly helped with automation (G1). Yet, the more complex a case is, the less likely that a DM at an upper level of a hierarchy would want any automation (G1).

The process of constructing a process, whether or not for automation, can be thought of as collaboratively procedural. One solution is to look at process automation from a HR perspective. By putting highly educated DMs in the beginning of the process, for example DMs who directly serve citizens on a daily basis, the citizens' questions can be answered quickly and an organization's system can learn from updated answers when they are processed (G2). The system gets a chance to better learn immediately by example. Currently, the municipality's system does not get a chance to learn because the decisions are made at the end and are solely archived (G2).

3) The content and format: The relationship between content and format was overall seen to be subjectively related. Due to the dynamic situations and roles of a DM, items should be seen as "logical artifacts that have a life cycle that spans different phases" (I2). Though the content may be similar, the purpose it serves may result in changing formats as an idea morphs through a development cycle, with a changing audience depending on the phase of an idea. The difficulty in understanding an individual's context accurately was noted in that the resulting correlation between content and format can be weak (I3).

Knowing the previous incarnation of information content as a trajectory of formats during content search or interaction would be useful (I2). For certain information interaction, a template that is fitting to the phase of the content progress with appropriate sections and subsections to be considered would increase efficiency (I2). Also, the format preferences of DMs should be respected. In the judiciary, lawyers as users are textually oriented while engineers as developers are visually oriented (G1). Presenting content in different formats was seen as translation in itself. The solution that was reached is that there is "always a story with a picture, a picture with a story" (G1).

For every conversation, adaptors that translate the discussed material to the level that DMs prefer, such as a summary which can be searchable and identified tasks with deadlines, would effective (I2). Open format for everything was suggested so that information can be exchangeable across all devices (A2). The content should also be open to make information sharing easier, yet considerations on the intellectual property of an organization for secure communication should still hold (A1, A2). The information content should reside in the cloud and not be reliant on data centers, which means that instead of allocating budget for hardware, organizations can focus on functionalities (A2).

4) Search: The format and content connection becomes constructive when considering what a DM is looking for. Information search can be made easier by having context in the beginning of the process (G2) and multi-perspectives should be shown when a DM enters a search (A1). For instance, if a user is searching for information on a local bridge, a picture of the bridge should come before the text description since a user who lives locally is likely to be

[2] https://www.schellingpoint.com/

[3] Jasser, Ulf. *BATOG Diagnostics – Overview*, Presented in Hamburg, Germany for *Electronic Data Systems*. March, 2006

more familiar with how the bridge looks (G2). In retail, two dimensions of mobility and industry domain can be synchronized to bring greater value to a customer's organizational workflow (A1). If the content is on new products that are arriving in stores, the format can be presentations on new products on mobile devices for retail associates, which empowers them to have information ready whenever and wherever to communicate with their customers (A1). This can potentially cultivate customer to customer communication (A1).

When a DM performs a search on a domain e.g., cloud storage, the industries that are supported by that domain within an organization such as retail and banking should be summated, possibly with services, industries, the divisions of interlinked organizations provided as multiple perspectives, as well as information on whom to contact (A1). A query dashboard that incorporates those perspectives gives a DM a meta-view (A2). In the judiciary, a meta-view can contain items of tasks, cases, files, and comments that can be made by other users (G1). Those same items will have to represent multiple views, for the views of a judge and a judicial employee already represent two different views (G1). User-specific presentations providing interaction facilities are in need, including touch interfaces that can be used in courts when digital presentation of documents can be accepted as a part of the legal process (G1).

Instead of always searching and gathering all information, DMs will automatically bring their own information (G1). Information will be commoditized whether or not it will be privatized for commercialization. The inevitable digitization of all information means that an organization should design a framework to adapt to it, instead of rigid rules that does not encompass all possible situations (I1, G1). Now, difficulties arise even in searching for an organization's self-generated information due to the lack of an overarching meta-view. The goal is "to have content in place in the process we work" (A1) along with evolutionary changes of any content (I2).

5) Privacy: Privacy is a central concern in making use of data interactions. Yet, the predominant perspective is that there is no difference between professional and personal lives. The reliance on the adequacy of regulations is instead moving towards individual accountability (I1). Individuals have to be accountable for personally identifiable information, but organizations also have to transform their models of information accountability (I1, A2). A data dimension that does not differentiate between professional and personal information brings the need for accountability with an understanding that it is closely related to transparency (I1). Transparency means that DMs have to acknowledge that data interaction accountability will result in a dialogue in a network, intentional or not (I1). Certain benefits possibly arise e.g., knowing if a customer already had a meeting with a competitor beforehand (A2), accessing and storing information with a single authentication method (A2), automatic classification and placement of confidential and public information (G2, A2), and lowering taxes for citizens by commoditizing data that a municipality generates (G2). Adequate privacy laws are open to interpretation and if the provided value of transparent information flow is high, the DMs will see the risk of always having to be individually accountable as an asset (I2).

D. Guidelines

The interviewees were lastly asked to expound on potential guidelines or rules for decision-making with big data. The predominant conclusion is that there is a paradigm shift away from assumed rules as systematic expectations. The interviewed DMs strongly leaned toward learning-based system support for decision-making that is individually tailored instead of rule-based system support.

Appropriate guidelines are not yet clarified for anyone (I1). For many, the situation is that "we have lots of information that's useful for lots of purposes. We even know the value of it and we don't know how to use it. And in coming ten years, that will be changing" (G2). This resonates with a DM's truly bounded rationality [17] as discussed in section II.D. A framework must be able to incorporate various viewpoints so that the best mode of working for individual DMs emerges (G1). Many trial and errors will be made as we progress along defining a framework (G2), which has to be evolvable because the roles and contexts of DMs are themselves dynamic. To gain all possible benefits of the digital world, what has to be realized is that some rules can only be made when errors occur (G2).

The challenge is that many organizations are not yet ready to exploit data interactions that DMs have since moving toward the digital era is still an ongoing process (I1,G1, G2, A1). The laws that the judiciary abides by have not updated to encompass the intricacies of the digital era (G1). The materials that are used in courts have to be permissible by law, which has not updated yet since 1996 (G1). Currently, all organizations have their own independent IT, and one common environment should be in place for digital processes to work (G1). Yet, establishing a common semantic model that remains largely unbiased towards a specific organization or clusters of users is difficult to establish (A1, G1). Finding a new taxonomy that incorporates multiple perspectives of all relevant DMs is critical (A1, A2).

A better decision-making process has to involve all of those in the hierarchy, from the lowest to the highest (G2, A2). This is crucial in understanding how interlinked organizations and factions within organization can work together by being aware of their differences (A2). The traditional hierarchical information structure works best only with highly structured situations (I1). A DM can neither be summated to a single role nor a single organization, so hierarchies are becoming more difficult to incorporate (I1). Thus, an enforced environment with a fixed view of the user that is imposed on DMs cannot hold. More external points of view must be gained to avoid the danger of being too closed in, for larger organizations have the tendency to want to do everything internally without being exposed to a larger network (A2).

Decision-making with big data is about individual empowerment, for proliferation of information as it gets generated and consumed will come to resemble vastly interconnected networks (I1). Organizations have to recognize that the speed of how a framework evolves will occur at a faster pace (A1). The importance of being agile is not only in anticipating ways for a system or an organization to be adaptable, but being able to adapt without anticipating guaranteed first-time success. Thus, an evolvable framework has to be envisioned rather than specific rules, for "there are many unknown unknowns ahead of us" (I3).

E. Summary and findings

All interviewees had a spectrum of formal and informal decision-making processes. Various sources are considered for information gathering to get novel ideas, build relationships, and to provide services for customers and clients while being mindful of information security for involved parties. Computation, memory and communication of information are collaboratively cyclical with DMs' contextualization emerging organically. The dominant findings are enumerated below.

1) Certain decision-making processes are or can be automated as an information environment learns of DMs' succession of repeated tasks.
2) A system's data linking capabilities should be exploited for DMs.
3) A system has to be dynamically aware of DMs' changing surroundings and roles.
4) In a system, DMs' ways of working must organically emerge for individual context creation and merge as collective context creation in an organization.
5) Considerations on the content and format of information is subjectively dependent on the information's life cycle as an item for continuous interaction.
6) Differing format for the same content is seen as a translation that can be contextualized during a DM's information search.
7) Multiple viewpoints within and outside of an organization have to be accommodated.
8) DMs need overarching meta-view for an evolving content within the decision-making process.

From above eight points, it can be concluded that an evolvable system has to be pursued that goes beyond the static treatment of DMs in Figure 1. DMs and their organizations make decisions dynamically, especially under increasing uncertainty due to big data. In crystallization, there are neither concrete guidelines nor wholly hierarchical decision-making in the non-structured information flow with big data.

V. An evolving system

Descriptions on an evolving system for a DM's contextualized decision-making are provided in this section. A direction for an evolvable information infrastructure emerges, as shown in Figure 2 from the gathered insights summated

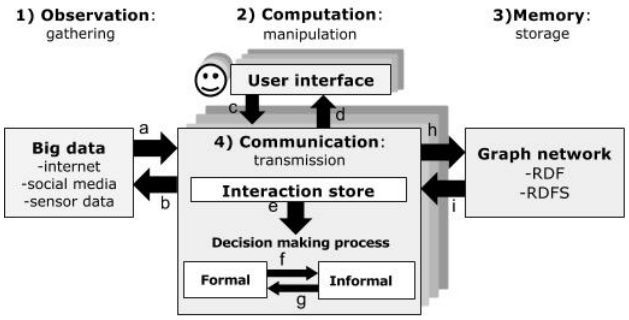

Fig. 2. Revised information flow

to the above eight points. Figure 2 stands as a skeletal structure with DMs of an organization developing contextualized versions themselves. The structure is based on the interviewed DMs' collective views that are evaluated with Radner's costly activities as labeled in Figure 2. Observation starts with structured and unstructured data. The observation stage continues via a user interface, such as a web browser or an information dashboard for gathering (a)(d) and accessing (b)(c) information. Computation deals with information manipulation as a continuous process. Data from the internet, social media, and sensors, such as locational information from mobile devices, surround all DMs for observation. Without distinguishing between internal and external sources, information is gathered for idea development, data analysis, and relationship building. Data can come in many types such as documents, presentations, blogs, images, videos, audio, news updates, wikis, social media updates, mobile locational awareness and other types of DMs' data interactions. Data interactions as query paths, item interaction paths, and item creation and edit paths are traced, which materialize as an organizational interaction store. The key aspect is that computation is communication; interactions that manipulate information are equal to communication amongst DMs. This flows to support decision-making processes (e) and generates a graph network (h) that will be further described in this section.

For all DMs, there are decision-making processes that are highly repeatable such as the patent application procedure for innovation strategists and small claims procedure for judicial workers. Then there are complex processes such as intricate trial procedures for judges and building growing relationships with customers for account management. Some start as formal, repeatable processes that become informal with data transferring into that process (f), such as certain formal social benefit procedures that require home visitations by a DM from a municipality to assess the state of an applicant's health, which can be an informal, unstructured process. Others start as informal procedures that aid in the formal process (g), such as an idea that morphs into a patent application for innovation strategy. More examples are aforementioned in section IV.C.2. All decision-making processes are dynamic and they become reusable via data processing.

The main difference is that repeatable processes have a cyclical start and finish periods, such as quarterly or yearly deadlines, but complex processes do not necessarily have a finite completion time. While judges have to decide on a ruling for complex cases at a certain point, relationship building with customers can be an ongoing and evolving process for account management. Thus, computation occurs within DMs' data interactions and decision-making processes in (e), (f), and (g) as communication. Computation as communication moves into memory, as demonstrated by information storage through (h) and (i) with a graph network as dynamic storage.

A graph network is composed via Resource Description Framework (RDF), a common data model for information representation on the web [4]. RDF is built through triples (entity-attribute-value) as statements. The connections between statements, resources (objects, e.g. authors, places), and properties (descriptive resources, e.g. "owned by") results in a graph-based model that links all data [1]. RDF is domain independent, so a RDF schema (RDFS) [5] is constructed by DMs in how a system should interpret their terminologies. This method provides flexibility that is crucial to encompass DMs' information needs in an evolving system, be it in the cloud, data warehouse, or anything in between.

Thus, a DM's itemized data paths, such as query paths taken or files edited and redirected within repositories, are nodes that are connected through DMs' interactions. Metadata input can be contained in the graph network as tags or descriptions that a DM might give to items of interaction, which can also be automatically generated based on item content. Formalized processes can become automated with DMs' repeated procedures that a system learns. The results of informal or organic processes that are not repeatable can be categorized for future retrieval and reference through RDFS.

The computation phase occurs again in re-processing information in the graph database, synchronously as information storage, with expected minimum delay in accordance with the iron law of delay [17]. This feeds back into a DM's data interactions (i) in a cyclical manner for interaction and also for observation as the information gathering phase through (b). Communication as information transmission is a continuously occurring phase for a DM with subjective demonstrations of information format, content, and search. This is contextualization as process in a user interface and in data interactions as communication that involves all DMs who share the system.

The iron law of delay [17] still poses a challenge with the model in Figure 2, yet the model is mindful of decentralization of decision-making in an organization. A different type of delay will occur with computation as communication. Efficiency of data processing is reliant on how efficiently DMs' communication materializes in the interaction store that continuously flows for re-processing and gathering.

[4]http://www.w3.org/RDF/

[5]http://www.w3.org/TR/rdf-schema/

A. Empowerment and empathy

An evolving system empowers and empathizes with its users. All participants shared the need for a constantly evolving environment that empowers individuals through a cyclical, rather than hierarchical, information flow. Empathy in this case does not equate to systematic responses to categorization of remarks or emotions such as ELIZA [23]. A system empathizes by empowering DMs as individuals, not roles. An empathetic environment channels big data into manageable connections, be it data connections that are not evidently obvious to DMs or social connections that data cannot be understood without. An evolving information contextualization stems from an empathetic environment for DMs' empowerment. When "there's no difference between personal and professional life" or what one says on Twitter or in physical presence (G2), the notion of information solely within an organizational setting is not empowering for a DM. An empathetic system first assumes that all decisions are made for countless reasons in countless ways and the walls of an organization come second. This is beneficial for an organization, for a DM expressed that "I want to understand where the world is going to so that I can adapt to that and you don't learn that from an internal course" (A2). While human and computational rationalities are bounded, organizations and DMs cannot be assumed to be bounded.

VI. Conclusion

An information environment has to reflect the social dynamics of DMs' data interactions [12]. It also must best adapt to organizations as social systems in which decision-making processes at all levels emerge as conditions instead of attempting to structuralize the cause and effect of organizational performances [11]. The hypothesis from section 1 stands to be simplified to three points based on the interview findings: *a DM's information infrastructure for decision-making has to be evolving, empowering, and empathetic.*

The end purpose of data analytics is to evolve, empower, and empathize with DMs as dynamic individuals, not static roles. The research findings conclude that transformation of information flow from Figure 1 to Figure 2 centers on dynamic DMs. Their interactions result in computation as communication that will contextualize the information flow in Figure 2 for an organization. Big data as quantifiable mass input cannot be fully controlled and big data as a phrase merely signals a change in how to conceptualize our relationship with information for discovering how to best satisfice. The framework has to be refined as unknown unknowns emerge within our bounded rationality of progressively boundless data dimension.

VII. Future works

Further advancement is necessary in two ways. First, more participants are essential in order to collect diverse viewpoints, for the current set of participants were all broadly involved in IT and were simillar in terms of age

range and gender. Second, how organizations and their DMs develop the skeletal model of information flow in Figure 2 for contextualization has to be investigated for numerous industry sectors and organizations. Observations on how an information environment grows and evolves and how it can best support decision-making with big data stands to be scrutinized.

VIII. Acknowledgements

The author thanks Frank Nack of University of Amsterdam and Peter Beijer of Hewlett-Packard for valuable comments and assistance. Furthermore, a special thanks to HP of the Netherlands for hosting this research.

References

[1] F. van Harmelen R. Hoekstra A. Grigoris, P. Groth. *A Semantic Web Primer, third edition*. The MIT Press, April 2012.

[2] A. Bryman E. Bell. *Business Research Methods, Second edition*. Oxford University Press, Oxford, England, 2007.

[3] D. Biederman. It's not all in the numbers. *Journal of Commerce*, 3PL Market Report: Services and Technology, special report.

[4] T. Carrozza. From hyperlinks to human links at hewlett-packard. *Knowledge Management Review*, 3:28–33, 2000.

[5] D. Lovallo D. Kahneman. Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, 39:17–31, 1993.

[6] O. Sibony D. Kahneman, D. Lovallo. Before you make that big decision... *Harvard Business Review*, June:51–60, 2011.

[7] O. Sibony D. Lovallo. The case for behavioral strategy. *McKinsey Quarterly*, March, 2010.

[8] T. Davenport. Ten principles of knowledge management and four case studies. *Knowledge and Process Management*, 4(3):187–208, 1997.

[9] Robyn M. Dawes. The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7):571–582, 1979.

[10] Kim. E. Brynjolfsson, L. Hitt H. Strength in numbers: How does data-driven decisionmaking affect firm performance? *Social Science Research Network*, April 22, 2011. Available at : http://ssrn.com/abstract=1819486 or http://dx.doi.org/10.2139/ssrn.1819486.

[11] J. Richard Hackman. From causes to conditions in group research. *Journal of Organizational Behavior*, 33(3):428–444, 2012.

[12] N. Gall A. Lapkin V. Sribar M. Beyer, D. Feinberg. 'big data' is only the beginning of extreme information management. *Gartner Report*, April 7, 2011. Available at : http://www.gartner.com/id=1622715.

[13] E. Brynjolfsson A. McAfee. Big data: The management revolution. *Harvard Business Review*, 2012. Available at : http://hbr.org/2012/10/big-data-the-management-revolution/ar/1.

[14] P. E. Meehl. Causes and effects of my disturbing little book. *Journal of Personality Assessment*, 50(3).

[15] P. E. Meehl. *Clinical versus statistical prediction; a theoretical analysis and a review of the evidence*. University of Minnesota Press, Minneapolis, Minnesota, 1954.

[16] Roy Radner. Bounded rationality, indeterminacy, and the theory of the firm. *Economic Journal*, 106(438):1360–73, 1996.

[17] Roy Radner. Costly and bounded rationality in individual and team decision-making. *Industrial and Corporate Change*, 9(4):623–58, 2000.

[18] L. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

[19] Pavel Serdyukov, Henning Rode, and Djoerd Hiemstra. Modeling multi-step relevance propagation for expert finding. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 1133–1142. ACM, 2008.

[20] H. Simon. Rational decision making in business organizations. *The American Economic Review*, 69(4):493–513, 1979.

[21] H. Simon. Bounded rationality and organizational learning. *Organization Science*, 2(1):125–134, 1991.

[22] Petros Venetis, Hector Gonzalez, Christian Jensen, and Alon Halevy. Hyper-local, directions-based ranking of places. In *37th International Conference on Very Large Data Bases (VLDB)*, pages 290–301. Stanford InfoLab, February 2011.

[23] J. Weizenbaum. Elizaa computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45, January 1966.

[24] L.M. Fagan S.M. Wraith W.J. Clancey A.C. Scott J. Hannigan R.L. Blum B.G. Buchanan S.N. Cohen Yu, V.L. Antimicrobial selection by a computer: A blinded evaluation by infectious disease experts. *Journal of the American Medical Association*, 242(12):1279–1282, 1979.

[25] Andrew L Zacharakis and Dean A Shepherd. The nature of information and overconfidence on venture capitalists' decision making. *Journal of Business Venturing*, 16(4):311 – 332, 2001.