DOI: 10.1561/9781638282372.ch3

Chapter 3

Using Causal Diagrams to Understand and Deal with Hindering Patterns in the Uptake and Embedding of Big Data Technology

By Anne Marie Weggelaar-Jansen, Sandra Sülz and Rik Wehrens

3.1 Introduction

In this chapter, we explain the interdependencies between actors and factors that influence the uptake of big data technology and provide more insights into the adoption and spread of big data technologies [1]. The systematic literature review by Günther *et al.* revealed that to advance our understanding of big data technology, [2] research should move beyond BigMedilytics (BML) study levels and examine how work practices, organizational models, and stakeholder interests interact with big data technology practices. In the BML project, we had a unique opportunity to review 12 study projects using different big data technologies aimed at different goals in several European countries. The studies:

> *cover three themes with the greatest impact on the sector. Population Health & Chronic Disease Management and Oncology comprise the 78% of deaths [in non-communicable] diseases. The third theme represents operations and equipment cost, covering the 33% of the expenditure in the sector.*[i]

i. BigMedilytics on: https://www.bigmedilytics.eu/ Accessed on August 29th, 2022.

Where possible, we captured interactions during the development of concrete BML studies, the organizations in which they took place, and the healthcare systems to which the organizations belong.

Our study contributes in two ways to gain more insights in the uptake of big data technology. First, we outline what policymakers should consider when developing public policy for big data in healthcare. Second, we give directions to management and healthcare professionals aiming to use big data technologies for the benefit of their patients and efficient processes.

The study presented here involved three single, yet aligned, multidisciplinary studies. The overarching aim was to examine how stakeholders in the 12 BML study projects worked on the performance, embedding, legitimation, and value creation of their big data application [3]. We include three broad categories: normative barriers (including cultural and ethical norms), market failures, and technocratic barriers (related to technological issues and government processes and regulations) [4].

First, we studied governance approaches, regulatory challenges, ethical dilemmas, and societal debates about big data technology. We conducted 145 semi-structured interviews in eight European countries: Austria, France, Germany, Ireland, the Netherlands, Spain, Sweden, and the United Kingdom. Respondents were identified via desk research and via our partners in BML. Respondents were: (1) healthcare professionals and management involved in big data studies; (2) ethical and legal experts knowledgeable about the key discussions in their country; (3) technology developers and data scientists; (4) representatives of patient and professional associations; (5) visible actors in the public debate to capture public perspectives on big data; and (6) policymakers and additional policy experts. Interview data were triangulated with policy documents, news articles, scientific papers, presentations, and gray literature provided by the respondents and a supplementary analysis of online documents. All interview transcripts and documents were analyzed abductively by qualitatively (open, thematic, and axial) coding [5].

Second, we monitored the performance of big data technology value over time with BML study-specific key performance indicators. For each BML study, workshops were organized to select relevant indicators and tailor these to the specific patient cohort, big data technology, and the aim of the BML studies. In these workshops, study team members and researchers developed the indicators based on an adopted version of the Balance Business Score Card to reflect the multidimensionality of performance: patient satisfaction [6], process outcomes, patient outcomes, and financial outcomes. Next, for most of the studies, a baseline measurement taken in the period before the implementation of the big data technology was followed by 6-monthly measurements during and after the implementation of the big data technology. The data collected were displayed on a digital dashboard available to all project team members.

Third, we studied via interviews on the dynamic processes involved in embedding big data technology in the daily work practices of healthcare professionals, organizations, and sometimes even societies. We developed insights into the underlying mechanisms, including how big data applications do or do not become embedded in organizational routines. Based on insights from normalization process theory [7], we include the different actor dimensions: (1) sense-making work: interpretations of what technology can add to work processes; (2) relational work: efforts in building a community of practice around the application; (3) operational work: the work involved in establishing new task divisions; and (4) appraisal work: formal and informal assessments conducted to assess the value [8].

During the BML project, we collected data on three levels: macro (as described before, e.g., ethical and legal experts, representatives of patient and professional associations, policymakers, policy experts, and public opinion makers), meso (organizational), and micro (professional interactions) levels. We included 'hard' data on structures, strategies, and procedures, and 'soft' data regarding stories, conflicts, and values as these point toward underlying patterns about 'the way things are'. The latter were collected through in-depth interviews with project members and stakeholders on both national and European levels as well as regular feedback moments with key actors in the BML project. Additionally, we conducted observations during general assemblies and study project meetings, and collected relevant policy and information documents.

After two and a half years, based on a thorough understanding of our data, we distilled a list of relevant factors and actors that influenced each other. Next, we drafted diagrams showing the interdependencies and patterns between actors and factors. Using arrows and loops in the causal models, we visualized patterns and identified underlying dynamics. These initial visualizations represent causal models (see Section 3.2) that were validated and improved in five workshops involving key actor groups: clinicians, technicians and data scientists, vendors, managers, policymakers, and funders. The draft models were adjusted and refined based on feedback and insights gathered in the workshops.

3.2 Findings

We present the three most important mechanisms for the uptake and embedding of big data technologies derived from our studies. We used causal modeling to synthesize our findings on the interpretations of our respondents on the uptake of big data technologies. This is especially relevant as causal models not only depict the actors and factors reinforcing patterns, but also focus on identifying leverage points where intervention is possible.

Causal models[ii] derive from a tradition of systems thinking in organizational studies, which focuses on examining the interdependencies between parts to understand the whole dynamic, interconnected system. Causal models are a powerful tool to deal with organizational change issues characterized by content complexity (the multidimensional and ambiguous character of organizational problems) and process complexity (the large number of people involved in the organizational problem, all with different viewpoints and interests) [3]. A typical characteristic of causal models is discerning feedback mechanisms (both positive and negative). These important mechanisms help to explain why some organizational issues tend to persist, despite many efforts to address them. These mechanisms are often invisible, as causes can be subtle and far removed from the consequences, often producing delayed effects [9].

In the following paragraphs, we present three causal models – The Information Road, The Golden Mountain, and The Swamp of Rules – which encapsulate the result of our interviews and engagement with the BML study projects. In each case, we provide a description of the particular causal model and summarize a set of recommendations for relevant stakeholders.

3.2.1 The Information Road

Measuring the impact of big data innovations is equivalent to identifying and quantifying the causal effect. We want to be able to quantify how much the quality of care (or other performance dimensions) would differ if a big data innovation had not been present compared with if it were present. This requires a clear understanding of the mechanisms of cause and effect. Without that understanding, we can still measure various indicators for monitoring or learning purposes, but we cannot attribute a causal meaning to the measurements.

How can big data impact healthcare? The central claim is that deploying big data does not improve healthcare directly. But big data innovations can help contribute by changing the information upon which decisions are made. This translates into a sequence of cause-and-effect relations, outlined in the conceptual model in Figure 3.1. We subsequently zoom in to the various cause and effect relations depicted in the model: we face a dynamic environment in which data are generated continuously. Mobile apps record our physical activities, smart devices keep track of our lifestyle, and eHealth technology increasingly monitors clinical alarms. Facilitated by technology, this continuous data-generating process is coupled with a growing belief that combining and analyzing that data can improve decision-making. For

ii. See for an explanation of this methodology: https://www.youtube.com/watch?v=cH4ybsGN2lA&t=375s.

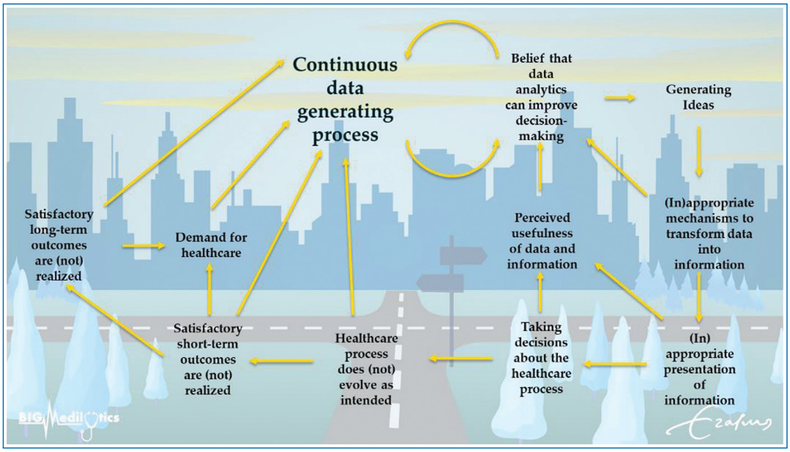

Figure 3.1. Causal model of how to measure the impact of big data technology.

instance, data-driven prediction models can alert to the potential deterioration in a patient's condition. These prediction models are expected to help clinicians to decide whether and when to intervene. Notably, the data-generating process and the belief in data analytics go hand in hand and reinforce each other. An increase in data availability leads to a stronger belief that there is something useful in the amount of data. And at the same time, the more the belief is verbalized, communicated, and debated, the more new data are generated. With an ever-increasing amount of data and technology, data scientists and healthcare practitioners generate new ideas on how data technology can help to improve decision-making. For instance, one of the BML teams was wondering whether the integration of real-time weather data can actually improve the accuracy of the prediction models. Data scientists and healthcare practitioners are jointly exploring how these ideas can be developed further and implemented in routine care.

The ideas that are generated affect the transformation process from data to information. To transform data into information, mechanisms and algorithms are in place to extract and pool relevant data. Inherent here is the idea that adequate mechanisms and algorithms are deployed that extract the right data from the right place at the right point in time and are provided to the right person at the right time in the right format/visualization.

This requires effort and a lot of unseen work such as 'cleaning' and correctly annotating data that can be used to develop algorithms. Importantly, the data transformation phase is not free from errors since integrating data can introduce bias if any data are incomplete or inaccurate. For instance, sometimes synthetic data are generated for testing a model or demonstration purposes, and this needs to be filtered out during the analytics. Such efforts affect the belief in big data technologies and can either reinforce or diminish that belief.

After the transformation phase, information is presented to the decision-makers in a customized fashion, frequently supported by visualization tools. Inherent here is the idea that the right information is presented in the right format to the right person at the right time. This is where things can go wrong, too. It might require effort to obtain the information because it is not automatically integrated in the workflow.

Too much detail can lead to information overload. Mismatches between the information context and the decision-maker's information literacy can occur, and decision-makers may fail to adequately understand what is presented. This affects how information is interpreted and what meaning is attributed to different options. This could cause frustration and dissatisfaction. Data and information can be perceived as less useful, which can negatively affect the belief in how far big data and data analytics can improve decision-making.

The way information is interpreted affects the decision-making process and which decisions are taken. Prediction models, for instance, have fuzzy decision points. Patients are more or less likely to respond to chemotherapy, and for some patients, this is not a clear-cut decision. Clinicians and patients might respond differently to the uncertainty in the information, which might cause differing decisions about the patient's health trajectory.

Once decisions are taken, health processes may not go as intended. The patient's condition might deteriorate unexpectedly during treatment, making ad hoc adjustments necessary. Or other exogenous challenges like the COVID-19 pandemic can cause disruptions and force rearrangements in health service processes. For instance, during the pandemic, in-house consultations had to be postponed or replaced by remote consultations, which affected the cost-of-service delivery. But was this change in cost attributable to the big data technologies or rather a consequence of the COVID-19 pandemic?

The way the health trajectory evolves affects how far we can achieve satisfactory short-term outcomes such as a reduction in hospitalizations. It also affects the data that are gathered at this stage. And it affects whether new demand for healthcare is generated if, for instance, patients are readmitted to the hospital. Therefore, whether satisfactory long-term outcomes are achieved and what type of new healthcare demand and data are generated all depend on a sequence of decisions and exogenous factors. Mortality, for instance, is affected not only by how well an algorithm supports clinical decision-making but also by the patient's underlying health condition. The extent to which changes in long-term outcomes can causally be attributed to big data, therefore, depends on how rigorously we can establish the counterfactual scenario of what might have happened if the big data innovation had not been developed and implemented. Also, the rigor and relevance of the data provided to the data scientists are influencing this.

In sum, the most important lesson learned:

- Data scientists and healthcare professionals need to define together feasible, acceptable, measurable, and informative indicators.

If data scientists and healthcare professionals intend to measure patient satisfaction, professional satisfaction, costs, and population health, the time period needs to be sufficiently long to track these in a proper way.[iii]

3.2.2 The Golden Mountain

This causal model explains how the innovative nature of big data fosters several processes that undermine the uptake of big data.

Due to the innovative nature of big data technology in the healthcare sector, healthcare reimbursement systems do not cover the entire data chain. This chain starts with data collection, combining various data sets, goes on to data storage, analysis, and developing an algorithm all the way through until the algorithm is eventually used for the benefit of the patient, professional, or organization. Given that there are no financial systems in place for big data technology, if you want to develop or apply an algorithm, you would need a grant to fund your work on, for example, developing a machine-learning-based decision-support system in precision medicine or preventive healthcare or a deep-learning-based algorithm for care prediction or real-time alerts. Alternatively, you might start a research project without additional funding.

Healthcare professionals and data scientists need to work together to write the grant application or start the research. To receive a grant, the proposal must contain specific, measurable goals expressed in tangible, appealing deliverables. The tendency is to have ambitiously high hopes for what will be developed, and this might lead to overpromising. This applies especially to big data technologies: what will be developed and how they will affect healthcare are promises, which can be seen as the "golden mountain" we all strive toward. However, it is hard to reach this golden mountain for four reasons.

First, the whole process of grant application is time-consuming. For instance, the reviewing committee takes time to decide, legal arrangements must be made, and partners need to hire staff. Meanwhile, knowledge of big data technologies increases, and before you know it, the ideas expressed in the grant might make less sense.

iii. See explanation of the Information Road causal model: https://www.youtube.com/watch?v=RrYvtg0_508&t=2s

A second reason is that daily practice can change even before the project starts. For example, a healthcare organization could restructure or develop a new care pathway that changes daily operations. The COVID-19 pandemic saw many changes that heavily influenced the set-up of big data projects. How can we collect valid, reliable data when the whole system is unstable and the results could be dubious? Can we use machine-learning algorithms based on data from old processes in the same way as from new processes? And sometimes daily operations change during a project. For example, the problem one BML study was designed to address vanished entirely after data collection. Understanding the problem showed that there was no need to develop an algorithm. Simply rearranging the processes was enough to solve the issue.

The third reason concerns the amount of work that needs to be done. This is often underestimated in grant proposals. Think, for example, about the work required to collect and clean the data, build the algorithm, and implement it in the daily practice of healthcare professionals. We noticed many BML projects did not consider how much additional work was needed before anything concrete could be shared.

The fourth aspect, common in innovation, is also worth mentioning. As explained above, you need to adapt your ideas to tailor them to a new situation, and this requires flexibility. However, the promises made in the grant proposal regarding the aims and methodology cannot easily be changed. By having to stick to the agreements made, you do not have the flexibility to meet the requirements of committed deliverables, and, as a result, means and ends are decoupled. It is tempting to proceed with the methodology agreed upon, but in the end, these projects get stuck in the middle of nowhere as they will not produce the results practice needs, and stakeholders – especially healthcare professionals – will be disappointed. One might argue that this inflexibility should be changed. However, this is usually not possible in the arrangements made with the funder. Changing things without the funder's agreement could lead to credibility issues, which will decrease the chance of future funding.

The slippery slope to the top of the golden mountain has many unexpected turns. If you take the other road and conduct a research project without additional funding, the same problems will occur. Researchers cannot simply change their methodology, as doing so will compromise the validity and reliability of their study. Additionally, no short-term results can be expected, as people cannot devote much time to the big data technology project. In some BML studies, big data research was a kind of hobby for healthcare professionals, next to patient care and organizational tasks. Again, the lack of short-term results is a disappointment for all involved, but especially for the data scientists.

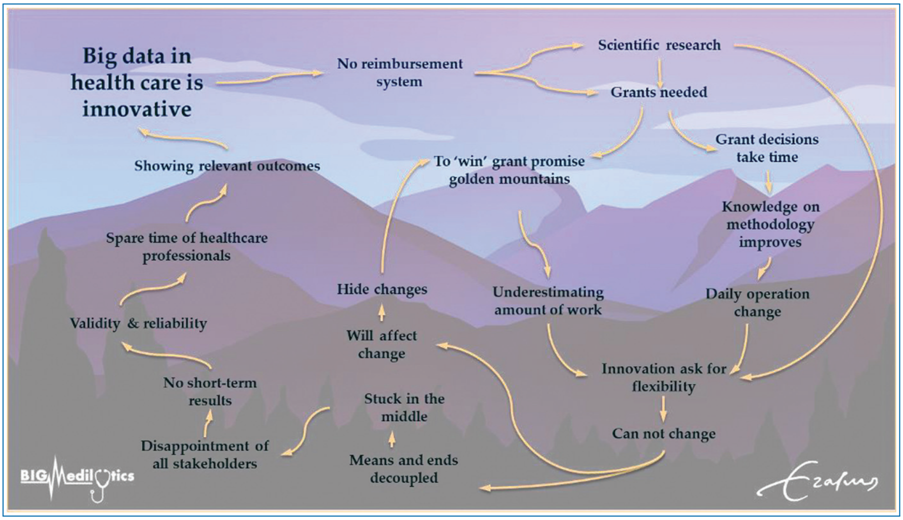

Figure 3.2. Causal model of how to measure the impact of big data technology.

In BML, we learned that to keep things on track, it is important to show the relevant short-term outcomes that prove that big data technologies are innovative and can be useful for daily practice (Figure 3.2). Even modest outcomes align healthcare professionals with data scientists. Discussing the methodology, data set, and outcomes leads to a shared understanding that helps to get and keep people engaged.

In sum, the most important lessons learned are:

- Data scientists and health professionals need to reflect on the promises they make in grant proposals.
- Grant proposals should include a description of all the work (not just the deliverables) that needs to be done to develop and employ big data technologies.
- Grant-funded projects need to have room to adjust the aims and adapt the plan to develop and employ big data technologies.[iv]

3.2.3 The Swamp of Rules

This causal model explains the interdependencies between actors possessing different expertise and their aligned rules and regulations. It explains how the innovative nature of big data slows down its uptake.

iv. See explanation of the Golden Mountain causal model: https://www.youtube.com/watch?v=aSmQTueUt0o&t=5s

We learned that policymakers, healthcare managers, and the public find big data technology in the healthcare sector risky for two reasons. The first is the media attention given to data breaches and privacy issues. Second, many stakeholders lack knowledge of what big data entails. The common reaction of people is to avoid risk or contain unwanted outcomes by asking for regulations, laws, norms, guidelines, and policies. This calls for written rules that provide guidance on what is allowed and how big data technology should be used. These rules should apply to the entire data chain (see Section 3.2.2) and focus on various topics, such as privacy, security, safety, ethics, and legal aspects. These diverse topics require different experts, such as lawyers in the field of big data, privacy and security officers, medical ethical advisers, data protection officers, and cyber security experts. Consequently, new jobs emerge with their own language and perspectives on the rules that need to be set for big data in healthcare.

As big data is a fuzzy concept, experts have their own opinions as to what it involves. For example, in the BML projects, some teams used an eHealth application or a medical device for data collection. Experts argued that eHealth and medical devices were big data, and thus the team had to comply with specific big data rules, which resulted in new approval procedures. The wide variety of new experts, who all focus on subsets of rules for big data, leads to project teams being confronted with a range of individual opinions on what is allowed or not allowed. This is because experts working in different fields are often not in close contact with one another. Thus, their expertise gets lost in knowledge silos.

As each knowledge silo makes their own rules, big data study teams can get bogged down in a misalignment between the rules set by the different experts, something we call "the Swamp of Rules." This misalignment creates uncertainty about which rules should be met before the big data project can start. Rule misalignment is especially troublesome when data are shared beyond the borders of organizations or even countries and whenever diverse experts in different organizations have differing opinions on which rule matters or should be prioritized.

We noticed in projects that were sharing data across nations that differences popped up in the interpretation of rules, despite legal attempts at harmonization, such as the General Data Protection Regulation (GDPR). Moreover, in addition to European legislation, every country has its own rules and specific derogations. Not only must big data projects comply with different rules, but the rules are also still developing rapidly. For instance, in one project, new national rules put the whole project on hold until the new privacy application was approved. In another project, the vendor of the data collection device was sold to a company in another country, which heavily influenced the progress of their project.

We noticed that, in general, the people involved in the BML project felt stuck in a cumbersome swamp of rules. Healthcare professionals and data scientists needed advice on how to comply with the different rules, especially when they appeared contradictory or open to different interpretations. Yet, in some projects, the role of the experts was only in checking compliance with the rules of the big data practice. In this context, unintentional mistakes can happen. Such unintended mistakes gain much media attention, which influences public opinion and, in turn, confirms the idea that big data in healthcare is risky.

We observed people using three ‘steppingstones’ to escape from or avoid the Swamp of Rules. The first steppingstone involved inviting all the experts in the project team to share their ideas on how to solve any rule misalignment or any other practical issues related to the set rules. The second steppingstone involved ‘workarounds’ and ‘elephant paths’ that project teams used as a form of knowledge brokering between the silos. For instance, patients in one study were asked retrospectively to give consent for using their data. Physicians made home visits to explain the reason. The third steppingstone placed the project in a context that permitted ‘learning by doing’, naturally with the consent of all the stakeholders involved. As a result, the innovative nature of big data was no longer seen as a risk, and the whole project team gained the opportunity to learn how to avoid problems caused by the set rules.

Figure 3.3 ties all the pieces together to reveal the opportunities to change aspects that sometimes hinder big data uptake.

In sum, the most important lessons learned are:

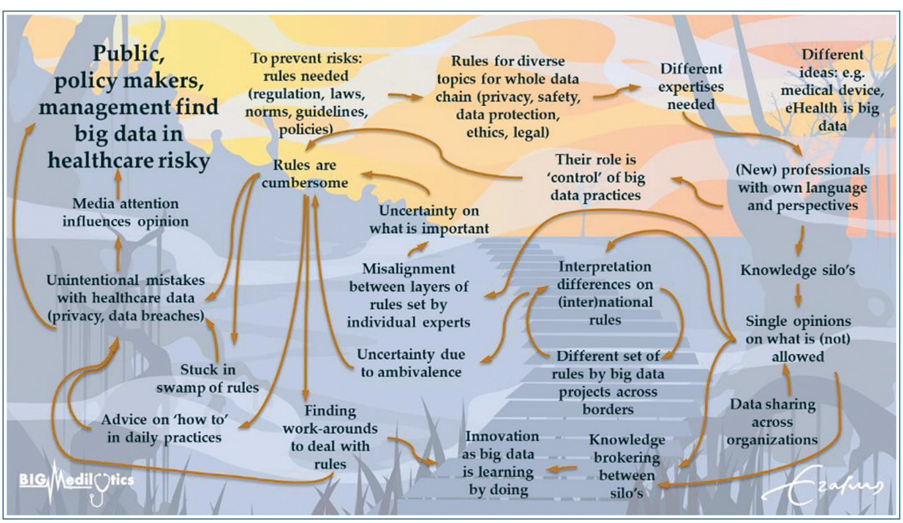

Figure 3.3. Causal model of influences that slow down uptake of big data innovations.

- Experts in the fields of privacy, security, safety, ethics, and law should work together and align their (expertise) work in big data technology projects.
- The experts in privacy, security, safety, ethics, and law on the team should advise healthcare professionals and data scientists on how they can comply with different rules.
- The experts in privacy, security, safety, and ethics on the team should inform the healthcare professionals and data scientists how to use 'workarounds' and 'elephant paths' that will enable them to follow the rules.[v]

3.3 Reflection on the Use of Causal Models

We have shown in this chapter that causal models are helpful to understand how to break self-enforcing patterns and how doing this can bring about change. Tying the pieces together in a causal diagram visualizes the obstructive and supportive actors and factors, and this in turn is helpful to understand when and how to intervene to overcome the problems revealed. In his work, Vermaak distinguished three main approaches to causal loop diagrams [9]. In the rationality-oriented approach, the emphasis is on making a solid causal loop diagram that represents 'reality' as accurately as possible. The aim is to produce a diagram that is as precise, objective, and valid as possible [10]. Second, the commitment-oriented approach focuses on building support to facilitate change. Such diagrams function as tools to bring diverging opinions closer together. Rather than accuracy and objectivity, the focus is on recognition and support. Finally, the development-oriented approach prioritizes learning and exploring.

Causal loop diagrams are generated collectively to share and exchange observations, points of view, and mental models [11]. The goal is neither complete accuracy nor unanimous consensus. Instead, the diagrams serve as input for dialogue and awareness-raising. Therefore, enhancing learning is a core criterion.

Furthermore, Vermaak discussed the balance that must be maintained in developing causal models [12]. While they benefit from intelligent simplification, they should not be too superficial, as they also seek to unravel and clarify underlying processes. One pitfall is not addressing the complexity of content, which happens when causal models are used as a discussion aid, but analytical rigor is discarded. Then, diagrams are drawn as a 'fuzzy visualization tool for intuitive insights' [9, p. 232]. Another pitfall is not addressing the complexity of processes. This happens when experts operate from inside their ivory tower, locking themselves away

v. See explanation of the Swamp of Rules causal model: https://www.youtube.com/watch?v=B3NCbg__4_4&t=28s.

to achieve research rigor. Causal loop diagrams therefore need to fulfill various criteria. "They need to be rich enough to capture underlying mechanisms, precise enough to spot leverage but also simple enough so that most important dynamics clearly stand out" [9, p. 233].

Our study applied the development-oriented approach to creating causal diagrams, as our aim was to support the study teams' activities and facilitate their efforts to embed big data technology in broader organizational routines. However, this approach may have biased our findings, especially because the data we used was derived mainly from BML project members and other actors involved in the project on national and European levels. On the other hand, we included many actors involved in the uptake and embedding of big data technology in organizations. Additionally, we realize that making and testing causal loop diagrams are interventions on their own. The awareness developed through these diagrams can empower early adopters, shift power balances, and so forth. Thus, our causal diagrams contributed in various ways to the uptake of big data technologies and hence the overall results of the BML project.

References

[1] Sweeney LB, Meadows D. The systems thinking playbook: exercises to stretch and build learning and systems thinking capabilities. Chelsea Green Publishing. 2010.

[2] Günther WA, Mehrizi MHR, Huysman M, Feldberg F. Debating big data: a literature review on realizing value from big data. The Journal of Strategic Information Systems. 2017. 26(3), 191–209. DOI: 10.1016/j.jsis.2017.07.003.

[3] Wehrens R, Sihag V, Sülz S, Van Elten H, van Raaij E, de Bont A, Weggelaar-Jansen AM. Understanding the uptake of big data in health care: protocol for a multinational mixed-methods study. JMIR Research Protocols. 2020. 9(10), e16779. DOI: 10.2196/16779.

[4] Heitmueller A, Henderson S, Warburton W, Elmagarmid A, Pentland AS, Darzi A. Developing public policy to advance the use of big data in health care. Health Affairs. 2014. 33(9), 1523–1530. DOI: 10.1377/hlthaff.2014.0771.

[5] Tavory I, Timmermans S. Abductive analysis: theorizing qualitative research. University of Chicago Press. 2014.

[6] Kaplan RS, Norton DP. The balanced scorecard: measures that drive performance. Harvard Business Review. 2005. 83(7), 172. URL: https://hbr.org/2005/07/the-balanced-scorecard-measures-that-drive-performance.

[7] May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, *et al.* Development of a theory of implementation and integration: normalization process theory. Implementation Science. 2009. 4(1), 1–9. DOI: 10.1186/1748-5908-4-29.

[8] Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, *et al.* Normalisation process theory: a framework for developing, evaluating, and implementing complex interventions. BMC Medicine. 2010. 8(1), 1–11. DOI: 10.1186/1741-7015-8-63.

[9] Vermaak H. Using causal loop diagrams to deal with complex issues: mastering an instrument for systemic and interactive change. Consultation for Organizational Change Revisited: Research in Management Consulting. 2016. 23, 231–254.

[10] Richardson GP. Problems in causal loop diagrams revisited. System Dynamics Review: The Journal of the System Dynamics Society. 1997. 13(3), 247–252. DOI: 10.1002/(SICI)1099-1727(199723)13:3<247::AID-SDR128>3.0.CO;2-9.

[11] Kim DH. Gui delines for drawing causal loop diagrams. The Systems Thinker. 1992. 3(1), 5–6. URL: https://www.cs.toronto.edu/~sme/SystemsThinking/GuidelinesforDrawingCausalLoopDiagrams.pdf.

[12] Vermaak H. Working interactively with causal loop diagrams: intervention choices and paradoxes in practical applications. Intervening and Changing: looking for meaning in interactions. John Wiley & Sons. 2007. 175–194.