6-18-2022

HOW COMMUNITIES OF PRACTICE ENABLE DATA DEMOCRATIZATION INSIDE THE ENTERPRISE

Hippolyte Lefebvre
University of Lausanne, hippolyte.lefebvre@unil.ch

Christine Legner
University of Lausanne, christine.legner@unil.ch

HOW COMMUNITIES OF PRACTICE ENABLE DATA DEMOCRATIZATION INSIDE THE ENTERPRISE

Research Paper

Hippolyte Lefebvre, Faculty of Business and Economics (HEC), University of Lausanne, Switzerland, hippolyte.lefebvre@unil.ch

Christine Legner, Faculty of Business and Economics (HEC), University of Lausanne, Switzerland, christine.legner@unil.ch

Abstract

To exploit the full business potential of their data, enterprises seek to empower more employees to work with data – a phenomenon also known as data democratization. In this way, they establish communities to connect and foster the exchange of practice between experts and a growing network of so-called data citizens. In this paper, we suggest studying data democratization from the perspective of communities of practice (CoP). Based on insights from more than 20 companies, we sketch a multilevel landscape composed of the following CoP: CoP focused on developing skills around tools and methods; CoP focused on a specific data object or data domain; and CoP spreading general data awareness. Our findings advance IS literature on the emerging phenomenon of data democratization and highlight the importance of both generic and situated practices as enablers. For practitioners, we provide actionable insights on how CoP can be structured around key data roles.

Keywords: Data democratization, Community of practice, Data citizen, Situated learning

1 Introduction

Despite considerable investments in their data and analytics capabilities, most companies still struggle to fully utilize the data they collect and generate business value from it (IDC, 2020; Panetta, 2021). To use data to its full business potential, they are seeking to democratize their data so that an increasing number of employees in different business functions use it to work with and make decisions (Zeng and Glaister, 2018; Lefebvre, Legner and Fadler, 2021; Lennerholt, Van Laere and Söderström, 2021). A cornerstone of data democratization is the collaboration and practice exchange between these casual data users often called data citizens and the specialized data and analytics teams (Awasthi and George, 2020; Labadie *et al.*, 2020). While project boards or councils act as formal collaboration mechanisms, data communities emerge as rather informal options that help connect practitioners in the network around shared data practices (Baijens, Huygh and Helms, 2021; Fadler, Lefebvre and Legner, 2021; Hagen and Hess, 2021). Thus, data communities support the development of greater autonomy and confidence in using data. Research has also demonstrated that better decision-making from data and analytics is possible when data democratization also promotes the diversity of opinion and sharing of knowledge and opinions between stakeholders (Hyun, Kamioka and Hosoya, 2020). This raises questions about the communities relevant to data democratization and how they support the development of new data practices.

To date, research has only provided anecdotal evidence about data-related communities and their role in democratizing data in the enterprise context. Despite their relevance, the existing IS body of knowledge neither informs about the typical communities that emerge to foster data democratization nor elucidates their interactions. In other research fields, communities have been briefly studied – for instance, in librarian research as a means of improving data sharing between scholars from the STEM or scientific research academic communities (Tenopir *et al.*, 2011; Springer and Cooper, 2020). To study this phenomenon, we choose to adopt the concept of communities of practice (CoP) by Wenger, McDermott and Snyder (2002). CoP are "a group of people who share an interest in a domain of human endeavour and engage in a process of collective learning that creates bonds between them" (Wenger, 2001, p. 2). This concept allows us to apprehend the phenomenon of democratization as a social transformation process happening between members of a shared data practice. Hence, we formulate the following research questions:

RQ1: Which communities of practice are relevant for data democratization?

RQ2: *How do these communities of practice interact with each other?*

As methodology, we opt for a multiple exploratory case study research design (Yin, 2003), which is appropriate when few research studies have looked into the topic of interest (Benbasat, Goldstein, and Mead, 1987). As part of an expert study, we first conducted semi-structured interviews with data and analytics managers from 17 companies to understand their data and analytics management practices and identify candidate communities. Then, with a narrower scope, we conducted two focus groups (13 and six companies, respectively) to study more closely how companies leverage CoP to enable data democratization. Doing this allowed us to analyze each company case against a common framework (within-case) and then analyze commonalities and differences regarding how companies leverage data communities to democratize data (cross-case). Our results show that companies create three different types of data CoP that foster data democratization: 1) *CoP focused on developing skills around tools and methods*, 2) *CoP focused on a specific data object or data domain*, 3) *CoP spreading general data awareness*. We illustrate each of them with a vignette and show that these three types of CoP encounter each other with different boundary interactions in a multilevel landscape of practice. Besides, we advance IS literature by pointing out both situated and generic practices as enablers for data democratization. Our results are also actionable by practitioners who can now identify and set up relevant CoP around key data roles.

The remainder of the paper is structured as follows. First, we examine and provide a synthesis of the relevant literature and identify the research gap. Then, we detail the methodology, as well as our research process. Next, we present an overview of our findings and discuss in detail each of the communities of practice identified. We then propose an overview of how the identified CoP integrate a landscape of practice. Finally, we discuss our findings, draw conclusions, and provide an outlook on future research.

2 Background

Although researchers have only recently started to study data democratization, they have highlighted the relevance of new structures for collaboration and knowledge sharing between data citizens and data experts (Zeng and Glaister, 2018; Hyun, Kamioka and Hosoya, 2020; Fadler, Lefebvre and Legner, 2021). We argue that data communities, which are the cornerstone of data democratization, should be studied through the lens of communities of practice (CoP) by Wenger (1998).

2.1 Data democratization

Recently, data democratization was defined as an "*enterprise's capability to motivate and empower a wider range of employees—not just data experts—to understand, find, access, use, and share data in a secure and compliant way*" (Lefebvre, Legner and Fadler, 2021). However, these data citizens need to increase their autonomy as they still heavily rely on data and analytics specialists for support and help to carry out more advanced data analysis (Awasthi and George, 2020). To address the skills gap, companies have to establish a shared understanding of basic data concepts (or data literacy) – that is, making sure people can read, work, analyze, and argue with data (D'Ignazio and Bhargava, 2015). Having accelerated the process because of the COVID-19 pandemic and the development of remote work, companies are developing their employees' skill building by leveraging practice exchanges and peer learning (McKinsey, 2021). For instance, they can use their data scientists as teachers for business experts by focusing only on the relevant datasets (Feng, 2017). Research has demonstrated that data and analytics can lead to better decision-making when data democratization also promotes the diversity of opinion and sharing of knowledge and opinions between stakeholders (Hyun, Kamioka, and Hosoya, 2020). Furthermore, the transmission of knowledge from data experts across internal firm boundaries to enable broader data use is positively associated with better value creation from Big Data (Zeng and Glaister, 2018). Hence, companies seek ways to stimulate interactions and knowledge sharing between various groups of users or with specialists from shared practices (Zeng and Glaister, 2018; Fadler, Lefebvre, and Legner, 2021).

While some companies establish formal alignment and collaboration mechanisms (e.g., boards in the context of projects), others create more informal settings where employees can exchange and grow their expertise. In this way, members of shared practice – often geographically separated – collectively learn from other members in a community setting and apply this knowledge in their own working area. By gathering practitioners from diverse teams around a shared domain of interest, communities stimulate data use, fight tribal knowledge, and foster practice-based learning (Lefebvre, Legner, and Fadler, 2021). They support the exchange of practices between business departments that might not be aware of each other's initiatives and would otherwise lead to redundancy, unrealistic expectations, and wrong communication. Several communities may also interact with each other, usually to establish a link between data provision and data use (Swanson, 2021). For instance, business managers and data expert communities may establish a shared understanding (e.g., requirements definition, technical constraints), which is critical for the success of a Big Data & Analytics project (Hagen and Hess, 2021). Other studies have investigated data communities with a narrow scope or in another context – for instance, with an emphasis on data sharing between scientists (Tenopir *et al.*, 2011) or between STEM scholars or scientific research academics working on the same data types (Springer and Cooper, 2020).

Overall, data democratization is still an emerging phenomenon in IS. Its implementation relies heavily on the development of the autonomy and trust of non-specialists who need to apply new practices regarding data in their working area (Zeng and Glaister, 2018; Labadie *et al.*, 2020; Lefebvre, Legner and Fadler, 2021). Anectodical evidence shows that communities as informal structures are an important means to connect across practices (e.g., data-related practices and business-related practices). However, we know little about how and why data communities emerge or how they are organized. Furthermore, the existing IS body of knowledge neither informs about the typical communities of practice relevant for democratizing data nor elucidates their interactions.

2.2 Communities of Practice

To study data communities as a phenomenon relying on practices, we adopt Wenger's (1998) concept of communities of practice (CoP). CoP revolve around the idea that learning is part of a social transformation process fostering collective empowerment around shared interests rather than a unilateral transfer of tacit knowledge by experts to neophytes (Wenger, 1998). Three essential characteristics identify and characterize communities of practice (Wenger, McDermott and Snyder, 2002). First, members should share a common *domain* of interest with concrete expected outcomes. Second, they benefit from mutually engaging, regularly interreacting, sharing, and learning together in a *community* setting (simplified as *Members* in the following). Third, members work on developing a shared repertoire of resources, or *practice*, that they will be able to implement in their working area. Thus, the CoP are sustained as long as members share common goals for the domain of interest and improve their own practice. The CoP can be distinguished from other group types such as regular work teams, where practice is defined by requirements and tasks assigned – hence, from project teams too. While CoP can be viewed as networks because they connect their members, networks involve more passive participation and focus on sharing rather than collectively developing a shared practice (Brown and Duguid, 2001). Nowadays, CoP rely heavily on ICT and have increasingly been held virtually. Thus, a deeper understanding of communities' characteristics may be required to account for this new way of working that engages geographically dispersed people, among others (Dube, Bourhis, and Jacob, 2006).

Furthermore, as organizations become more complex, members may need to belong to several CoP that encounter each other, with boundary objects delimitating members' inclusion (Wenger, 1998). As a result, CoP can be observed from a landscape perspective showing their interconnections and boundary interactions (Carlile, 2004). In such a system, members might explore the applicability of their practice in a "totality of local communities," that is, other CoP, requiring them to cross their own initial boundaries (Wenger-Trayner *et al.*, 2014). In this way, they also become knowledgeable about other practices and identify as a member of a larger body of knowledge not limited to their own local practices. Landscapes of practice (LoP), thus, represent the same body of knowledge (Wenger, 1998) made out of interconnected practices with clear identities, well-defined boundaries, and knowledgeable members (Pyrko, Dörfler, and Eden, 2019). LoP ignore organizational structures by focusing on practice only while weaving both boundaries and peripheries on the different CoP belonging to it. Such a view is likely to reflect the reality of work and learning, as practitioners must be knowledgeable beyond their local practice to perform their own tasks well. This means that communities belonging to the landscape are "*accountable to one another in terms of their respective practice-based knowing*" (Pyrko, Dörfler and Eden, 2019, p. 483) to enable situated learning (Gherardi, 2000). Local practices are regularly renegotiated based on practices observed within other communities in the landscape and formatted by more general practices applicable to the different landscapes, i.e., applicable to various bodies of knowledge in the organization (Wenger, 1998). Overall, a difference should be noted between situated practices (i.e., practices fostered within members' own working area) and generic practices that influence both the whole community landscape, as well as local practices (e.g., presenting, teaching, analyzing) (Pyrko, Dörfler and Eden, 2019).

3 Methodology

We opted for a multiple exploratory case research design (Yin, 2003). Such a method is particularly relevant to capture rich insights about the phenomenon of interest in its natural setting and to capture as much knowledge as possible from practitioners (Benbasat, Goldstein, and Mead, 1987; Yin, 2003). In the following, we lay out the research process, data collection, and case analysis.

3.1 Research process and data collection

Our research process is divided into two main research phases: first, an expert study to get a broader understanding of data and analytics management practices in enterprises, including the formal and informal mechanisms for alignment and collaboration between the data organization, business, and IT departments; and second, two focus groups with a narrower scope on communities of practice that foster data democratization (see Table 1).

Research activities	**1- Expert study on data and analytics management practices**	**2- Focus groups on data communities**	
Period	09/2020–10/2021	10/2021–11/2021	
Objective	• Understand data and analytics management practices, including the alignment and collaboration between the data organization, business, and IT • Identify communities in the different organizations	• Collect data communities that connect and foster exchange between different roles or groups • Identify most relevant communities of practice for data democratization	
Data collection	90-minute semi-structured interviews with 31 experts from 17 companies	Focus group with 30 experts from 13 companies	Focus group with 16 experts from seven companies
Data collected	47 data communities provided by 17 companies	18 data communities provided by seven companies	
Qualifying as CoP	40 CoP from 14 companies	10 CoP from five companies	

Table 1. Data collection

From September 2020 to October 2021, we conducted semi-structured interviews with 31 data and analytics management experts from 17 companies to gain an overview of their data and analytics practices. The interviews were conducted by two researchers via videoconference using MS Teams and were scheduled for 90 minutes (actual range of duration: 66–90 min). The interview guideline was structured around five topics: *Business drivers and data strategy*, *Relevant scope of data and analytics*, *Data and analytics organization*, *Data and analytics roles*, and *Alignment and collaboration*. To ensure the interviewees had an overview of both global and local practices relating to data and analytics management, we selected managers with data-related leadership and oversights responsibilities (Table 2) and at least three years of experience in the company. All interviews were recorded and documented. We also completed the documentation of this primary data by reviewing existing documentation and searching for relevant public sources (e.g., keynotes, press articles, company website). Thus, we could triangulate the information documented during the interview and ensure validity (Yin, 2003). We eventually sent the documentation back to the interviews for review, to confirm our understanding, and to address the remaining open questions.

We then offered companies a follow-up discussion around CoP relevant to data democratization to get additional details about the main CoP observed, as well as interesting new cases of communities that we had identified in the meantime. Focus group 1 happened in September 2021 as part of a workshop with 30 practitioners from 13 companies from various industries and that manifested interest in the topic. Using a Miro board, participants were invited to describe examples of communities of practice that foster data democratization in their company against the three criteria introduced by Wenger *et al.* (2002). We also asked the experts about a set of structuring characteristics from Dube *et al.* (2006) as we expected most of these CoP to happen more and more in virtual settings due to the COVID-19 context. We focused on the five following essential structuring characteristics: *Size, Community leadership, Life span, Creation process, Degree of formalism*. We purposely ignored characteristics related to maturity and lifecycle as we focus on drafting a landscape of communities that are only emerging in most companies. Thus, we also did not include characteristics with regard to the members' enrolment process. In addition, we inquired about further facts such as meeting frequency or communication channels. For companies that did not participate in the semi-structured interviews and to ensure additional validity, we followed up with them to collect further documentation of their described CoP. For instance, we could obtain documents such as community procedures or examples of community meetings documentation. Necessary clarifications could also be made during focus group 2, which happened in October 2021 in the form of a Web session with 16 practitioners from seven companies. The research team presented the preliminary results of this study, and a majority of participants confirmed they were relevant in the context of data democratization as they support broader data use and learning.

In Table 1, we present the output from our two research activities (i.e., the counts of candidate communities collected). One of the challenges we faced relates to practitioners' understanding of communities versus teams or informal networks requiring further clarification with certain companies. As a result, some of the candidate communities could not qualify for the study and were removed. Examples collected from companies H and Q did not qualify as CoP. Eventually, we were able to identify 40 data and analytics CoP in research activity 1 and 10 in research activity 2. Owing to the participation of several companies in both research activities, five CoP were mentioned twice, leading to a total of 45 CoP identified during the whole research process.

3.2 Within and cross-case analysis

We performed the within-case analysis by coding each community documentation against the community of practices criteria by Wenger *et al.* (2002) i.e., *Domain, Members, Practice*. We completed it by adding the available structuring characteristics collected during the focus group. In addition to *Creation process, Degree of formalism, Size, Community leadership, Life span,* our data from the expert study allowed us to characterize *Geographical dispersion and Orientation* (Dube, Bourhis, and Jacob, 2006) as supplementary characteristics mentioned as relevant by the participants of the focus group. The *Creation Process* of CoP may either be deliberate (e.g., established by management) or emerge among a group of employees with a shared domain of interest and seeking to exchange around shared practices.

Company	Industry	Revenue/employees	Key informants	Research activities	Examples of CoP as mentioned by companies
A	Public transportation	$1B–$50B/~35,000	Product owner data strategy; Enterprise Architect for Data & Analytics	1	Data science community; Analytics capability community
B	Manufacturing, chemicals	$1B–$50B/~5,000	Head of Corporate Data Management	1, 2	Master data material community; Master Data Lunch
C	Packaging, food processing	$1B–$50B/~25,000	Director of Global Master Data Strategy	1, 2	BI community; MDM community
D	Manufacturing, automotive	$1B–$50B/~90,000	VP Data & Analytics Governance	1, 2	Data domain manager round table; Global Data science and AI community; SAP analytics CoP; Data quality circle; Enterprise Architecture community
E	Consumer goods	$50B–$100B/~350,000	Master Data Lead; Group Manager Data and Analytics Products & Services	1, 2	Master Data community; Analytics communities for specific tools (e.g., PowerBI)
F	Manufacturing, automotive	$1B–$50B/ ~150,000	Head of Master Data Management; Head of Advanced and Self-Service Analytics	1, 2	Master Data Management community; Data Science and Analytics Experts groups
G	Pharmaceutical	$1B–$50B/ ~70,000	Global Data Lead; Enterprise Solutions Architect Analytics Lead	1	Monthly global communication of data practice
H	Consumer goods, retail	$1B–$50B/ ~30,000	Vice-President: Data & Analytics	1	Data domain working groups
I	Consumer goods, retail	$100B–$150B /~450,000	Head of Enterprise Data Management	1, 2	Data sharing community; GS1 community
J	Chemicals	$50B–$100B/~120,000	Product Manager Data Governance & Stewardship	1	Reporting & Analytics community; Data steward community
K	Fashion and retail	$1B–50B/~60,000	Head of Data Quality; Data Catalog Community Governor	1, 2	Data Quality community; Data catalog community
L	Pharmaceutical, chemicals	$1B–$50B/~100,000	Head of Enterprise Master Data	1, 2	Master Data Management community
M	Pharmaceutical devices	$1B–$50B/ ~65,000	Senior Manager Business Analytics	1, 2	Master Data Management community
N	Adhesive & beauty products manufacturing	$1B–$50B/~20,000	Director Master Data & Product	1	Data Expert community linking regional hubs
O	Outdoor power products manufacturing	$1B–$50B/~20,000	Senior Director Business Transformation Data Management	1	Data governance community
P	Technology & networks	$100B–$150B /~200,000	Head of Corporate Data Management	1, 2	Data modeling community; Data quality community
Q	Pharmaceutical	$1B–$50B/~70,000	Enterprise Data and Analytics Operations Cluster Chair; Finance Data Director	1, 2	Supply Chain Master Data Team; Customer data team
R	Software development	$1B–$50B/~100,000	Solution Advisor Expert	2	Material master data community; Customer master data community; Governance learning series
T	Network & telecoms	$1B–$50B/~100,000	Head Information Operations Management	2	Information and Data Architecture community
U	Logistics operations	$1B–$50B/~70,000	Head of Global Master Data Management	2	Data mobilization

Table 2. List of companies involved in the research process

Degree of formalism relates to different levels of formal recognition and integration in the enterprise structure: from unrecognized (i.e., invisible to most employees in the organization), bootlegged (i.e., visible only to a specific group), legitimized (i.e., officially sanctioned as valuable entity), supported (i.e., receiving direct resources) to institutionalized (i.e., official status and functions) (Wenger, McDermott and Snyder, 2002, p. 28). *Size* is an important characteristic in the context of data democratization as it informs of the number of people involved in the CoP, hence is a proxy for how data citizens are engaged. While smaller communities are more likely to facilitate shared understanding and practice exchange, larger communities with hundreds to thousands of members may lead to a dilution of members' contribution. Hence, larger communities may inherently face the difficulty to sustain members' interest. *Community leadership* might be assigned through formal structures and responsibilities in a governance model for instance, which is particularly suitable for larger communities (Dube, Bourhis, and Jacob, 2006). However, depending on the needs or expertise, community leaders might also emerge naturally. Communities' *Life span* might vary greatly. It can be tied to a specific initiative or project (i.e., temporary), or may support a larger and long-term perspective (i.e., considered as permanent). *Orientation* informs about the general purpose of the community i.e., whether it rather supports a strategic use case or rather focus on operational practices. *Geographical dispersion* describes how physically dispersed community members are. Hence, widespread communities might face several challenges such as the participation of members belonging to various business departments or from different premises, or even time zones.

The two researchers present at each of the interviews performed this analysis to ensure a common understanding of each sampled community. As a basis for discussion, we could leverage the contextual understanding collected during the expert study research activity concerning companies' strategies, their current data initiatives, or their organizational structure for data. Moreover, we could collect information with regard to roles and responsibilities as well as headcounts which helped us to better apprehend within-case coding. Next, we conducted the cross-case analysis to understand commonalities and differences across the sample. By using pattern-matching (Yin, 2003), we were able to identify 1) a generalizable set of typical communities of practice relevant to the context of data democratization and 2) patterns in the way companies form their CoP across domains and practices, and with various audiences.

4 Results

4.1 The three types of CoP that support data democratization

Most participating companies (except Q and H) operated one or several communities to align and collaborate with an increasing number of employees in different business functions and different locations. Based on the concept of CoP, we can identify three types of typical CoP that foster data democratization: 1) *CoP focused on developing skills around tools and methods*, 2) *CoP focused on a specific data object or data domain,* 3) *CoP spreading general data awareness*. We find that these CoP are observed at different levels of practice and cover seven domains of interest. Type 1 CoP typically address *Data quality and management*, *Reporting and analytics*, and *Data modelling and architecture*. Type 2 CoP typically address *Data-driven innovation* or focus on *Data object/product*. Type 3 CoP mainly foster *Global data awareness*. While typical members for these CoP can be identified, we notice that several roles may belong to more than just one community. CoP are partly established and managed by data organizations, but we also observe communities that emerge directly from practitioners of shared practices who acknowledge the value of such exchanges. CoP for data democratization are mainly considered permanent as members expect long-term benefits from them. However, they can also be created for specific purposes or needs (e.g., an innovative use case) and thus be rather temporary . We provide an overview of the three types of CoP and their characteristics in Table 3.

Overall, these CoP foster data democratization by connecting people from diverse data expertise around shared domains of interest. Besides creating general awareness about data in the company, they enable people to develop their skills in collaboration with domain or technical experts and share data. By sharing their experience, the participants enable situated practice; in other words, the members will apply these newly learned practices in their own working area. In the following sections, we detail the three identified CoP types and illustrate them with vignettes from the cases.

Community type	CoP focused on developing skills around tools and methods			CoP focused on a specific data object or data domain		CoP spreading general data awareness
Boundary interactions	Boundary practices			Boundary encounters		Periphery
Case companies	C, D, F, I, J, K, L, M, N, O, P	A, C, D, E, F, J, K	D, P, T	A, D, F	B, D, E, I, K, R, T	G, U
Domain of interest	Data quality and management	Reporting and analytics	Data modelling and architecture	Data-driven innovation	Data object / product	Global data awareness
Members	Data stewards, data (governance) managers, data owners	BI experts, data analysts, data scientists	Data/Enterprise architects, data modelers	Chief Data Officer, process owners, data architects	Data stewards, data architects, data domain owners	All interested with a core group of data influencers
Shared practice	DQ tools, KPIs, Definition and standards	Analytics tools and techniques	Data modelling, Database design, architecture	Analytics use cases, innovation, and creativity techniques	Data lifecycle, DQ measurement, and improvement	Updates, feedbacks, requests, facts and stories
Orientation*	Governance	Operational	Governance	Strategic	Operational	Operational
Community leadership	Clearly assigned	Negotiated based on expertise	Negotiated based on expertise	Negotiated based on expertise	Clearly assigned	Clearly assigned
Geographic dispersion	High	Medium to High	Medium	Low	Medium	High
Creation process	Spontaneous or intentional	Spontaneous or intentional	Spontaneous or intentional	Intentional	Spontaneous or Intentional	Intentional
Life span	Permanent	Permanent	Permanent	Temporary	Temporary/ permanent	Permanent
Degree of formalism	Legitimized	Legitimized	Bootlegged to Legitimized	Bootlegged	Legitimized	Institutionalized
*We extend the work of Dube, Bourhis, and Jacob (2006) by integrating governance as a third orientation between operational and strategic as introduced by Fadler and Legner (2021)						

Table 3. Overview of the identified types of CoP and their characteristics

4.1.1 Type 1: CoP focused on developing skills around tools and methods

The first type of CoP uncovered focuses on developing specialized skills around tools and methods. They bring together members interested in developing specific areas of expertise in either data quality or data management, reporting and analytics, or data modeling and architecture. Members seek to share experience and grow their technical expertise by learning from their peers. These members are often data specialists (e.g., data architects, data managers) who exchange about relevant practices (e.g., data models). In that case, groups might be self-organized, not known by others (bootlegged), and community leadership is then typically negotiated based on expertise or seniority. However, other data roles in business might also exchange about specific methods on how to use data (e.g., Data quality KPIs), requiring the clear assignment of a community leader. Because of their nature, Type 1 CoP are rather permanent and seldom have a pre-defined life expectancy. They are usually legitimized (i.e., officially recognized as valuable entities). As their reach across the network can be quite high depending on the domain of interest, preferred communicated channels are e-mails, conference calls, or collaboration platforms such as Yammer. For topics that are relevant for data professionals in central teams (e.g., data modeling and architecture), members' geographical dispersion is lower than for topics that are relevant to every business function and location/site in a multinational enterprise.

Vignette 1: Data Quality community at Company P

Company P is a global leader in technologies, network, and telecommunications solutions and has more than 200,000 employees worldwide and in more than 170 countries. To scale up data use for the entire organization and foster a data-driven culture, the company has set up a Chief Data Office with about 50 FTEs reporting to the CIO (board level). The Chief Data Office's responsibilities include deploying the governance framework, rolling out procedures, and maintaining the information architecture across 100+ business objects. Its oversight spans over 10 function areas, for instance, finance and logistics, that have their own data management and data quality teams. Depending on the functions, such teams might sometimes be bigger than the central team. Thus, data is widely democratized in functions.

To monitor data quality across functions, a corporate data quality index across 22 data domains (built along functions and divisions) is measured twice a year, signed off by the CFO and CIO, and presented to the board members. A minimum score of 60% is currently set as the threshold for requirements completion. Thus, the assessment is performed against 1200 data objects provided with requirements and ownership on both the data and its metadata. To enable this global effort for better data quality, a large community has been established by the Chief Data Office as essential to foster cross-domain alignment about data quality. This community is institutionalized (i.e., officially recognized as highly valuable) and comprises more than 1,000 data stewards, data experts, and business analysts. The members are geographically dispersed all over the world but gather monthly to exchange and learn how data quality can be improved in source systems. They also represent their local practice in each of the data domains and seek to increase their domain quality index. By learning about practices in the most successful domains and presenting their challenges to others, members expect to improve their own domain data quality. Community members meet every month, but further exchanges happen through a group chat and a dedicated wiki supported by a dashboard to monitor quality improvements.

4.1.2 Type 2: CoP focused on a specific data object or data domain

The second type of CoP uncovered focuses on specific data domains or data objects. These kinds of CoP gather practitioners from different expertise area contributing to a high-level domain of interest. For instance, several experts from different Type 1 communities might exchange with domain experts about a data domain (e.g., marketing or sales) or data object (e.g., product master data). Experts with specific skills (e.g., from Type 1 CoP) can then be seen as orthogonal to data domains by bringing their expertise in broader topics of interest. This leads to a "grid view" of data domains versus a panel of technical experts from Type 1 communities involved in various ways. While strategic alignment usually happens through formal mechanisms (e.g., boards and committees), we also observed temporary CoP popping up at the strategic level when needed. For instance, business, IT, and data stakeholders might spontaneously create a CoP to discuss and align on the development of a new strategic data and analytics capability in business or related to a specific data product. Their leadership is not really negotiated based on expertise but rather guided by team leaders or hierarchy. Overall, Type 2 CoP might be geographically quite dispersed to allow for sharing practices across locations, but they can also be associated with a specific site.

Vignette 2: Data Catalog community at Company K

Company K is a leading fashion and retail company with more than 55,000 employees worldwide and is implemented in 160 countries. The company releases more than 80,000 articles every year. Each article has more than 400 data points fed by more than 2,000 data creators and generated from 73 products systems. For instance, each sports article reference might be referred to by further attributes (e.g., color, size). In early 2021, the new company strategy released highlights of its ambitious e-commerce goals. Currently, the business model is mainly wholesale and will progressively shift to consumer business. The company was strongly impacted by COVID-19, leading to a surge in its e-commerce sales. This change of business model requires more data, enhanced data quality, and data management. In addition, the wholesale channel also requires further data quality as wholesale partners need it for their e-commerce too (e.g., accuracy of description). Eventually, from an operational perspective, the new strategy targets that 90% of the products should be sustainable. This leads to the collection of new data objects to capture the overall sustainable footprint (e.g., water and electricity used in factories).

To enable data-driven decision-making from their huge amount of data, Company K seeks to establish an enterprise data culture of awareness, credibility, and trust, combined with a strong data quality improvement initiative. A growing central data team (120 FTEs) handles data management, data quality, data platform, BI, and analytics. In business, dedicated data owners manage master data (and few transactional data) in data domains. A decentralized analytics team in the sales department focuses more on the fast-moving analytics products (e.g., product recommendation on e-commerce). The company seeks to foster data sharing and collaboration between the different data and analytics teams by offering a 360° view of data. In short, data democratization at company K means establishing a sustainable link between data creators and data consumers in the data mesh. These two groups then form a temporary community of practice around specific data products. Their exchange is organized through the data catalog (provided by Collibra) community and facilitated by a dedicated formal role in the central data management team: a Data Catalog community governor, for whom "*data is not only for geeks.*" More precisely, the latter orchestrates the onboarding of the required data onto the data catalog. In this way, the company can address data siloes generated by product systems and drive data quality necessary for its ambitious analytics use cases (e.g., in-season product forecasting).

4.1.3 Type 3: CoP spreading general data awareness

The third type of CoP uncovered is special in the sense that it involves less practice exchange but aims to disseminate updates, best practices, stories, and training (i.e., general awareness about data). Type 3 CoP might be temporarily created in the context of large corporate events or be legitimized as a regular practice exchange. They are set up to support a larger context or initiative with regard to data in the enterprise (e.g., data democratization). They are piloted by a core team usually located in the central data or analytics organization and seek to reach any employee interested in data to create a data culture down to the operational level. Hence, one of the challenges of these types of CoP is to sustain engagement of a large (typically 100+) and geographically dispersed member base with limited accountability on their own data practice. As lurkers or guests, most of the participants remain in the CoP's periphery and might never become full members. However, if engagement is sustained, members can integrate these mostly generic practices into their local practices. For instance, Type 3 communities might be highly relevant to foster shared language and vocabulary concerning data. As a result, the core team invests a lot of its time and resources, and dedicated FTEs might be assigned for continuous community entertainment.

Vignette 3: Data mobilization at Company U

Company U is a leading provider of global logistics solutions and has more than 70,000 employees across 2,100 locations worldwide that seek to achieve financial excellence by 2024. In the company's business strategic plan, three specific goals are targeted: maximize cost and performance transparency; enable instant data availability; and drive digitalization across all finance operations. Harmonization of master data is a key initiative launched in 2021 to support this journey toward excellence. For the CFO who kicked off the global data mobilization, "*Master data management (MDM) is not a head office project. It involves all of us.*" There has to be a culture of completeness and correctness (i.e., data quality), especially at the operational level, to enable the digitalization of the finance function. To support this vision, the global data management teams, which consist of eight FTEs and are already well-connected to their regional counterparts, developed the data mobilization to engage with the whole company about master data management. These sessions, which take place twice a month, welcome any data citizen interested in learning more about master data (and particularly business partner data). The global team then provides awareness sessions, project updates, guests presentations, and newly available training sessions. Exchanges are fostered through quizzes and prize winners, Q&A sessions, and feedback/requests collection. As the leader of the community, the global MDM team focuses on sustaining interest by creating excitement about the topic and not by just addressing operational matters. "*You have to be attractive and keep them entertained,*" says the Head of MDM. For instance, themes are associated with each session (e.g., Halloween sessions discussed *horror* topics, including cyber challenges). Between sessions, continuous engagement is organized through the company collaboration platform where the global team collects members' new topics of interest that could inform upcoming sessions. In less than a year, attendance grew from about 100 attendees at the first session to more than 400 in October 2021, with 70% of participants coming from non-MDM functions.

4.2 A landscape of practice for data democratization

Together, and due to their members' interaction, the identified CoP form a landscape of practice for data democratization. It is then composed of several CoP (see Table 3) that encounter each other with different boundary interactions. For instance, several key roles are involved in different types of communities (data steward, data owner, data architect) casting the relevance of the landscape of practice for data democratization. We display such a landscape in Figure 1.

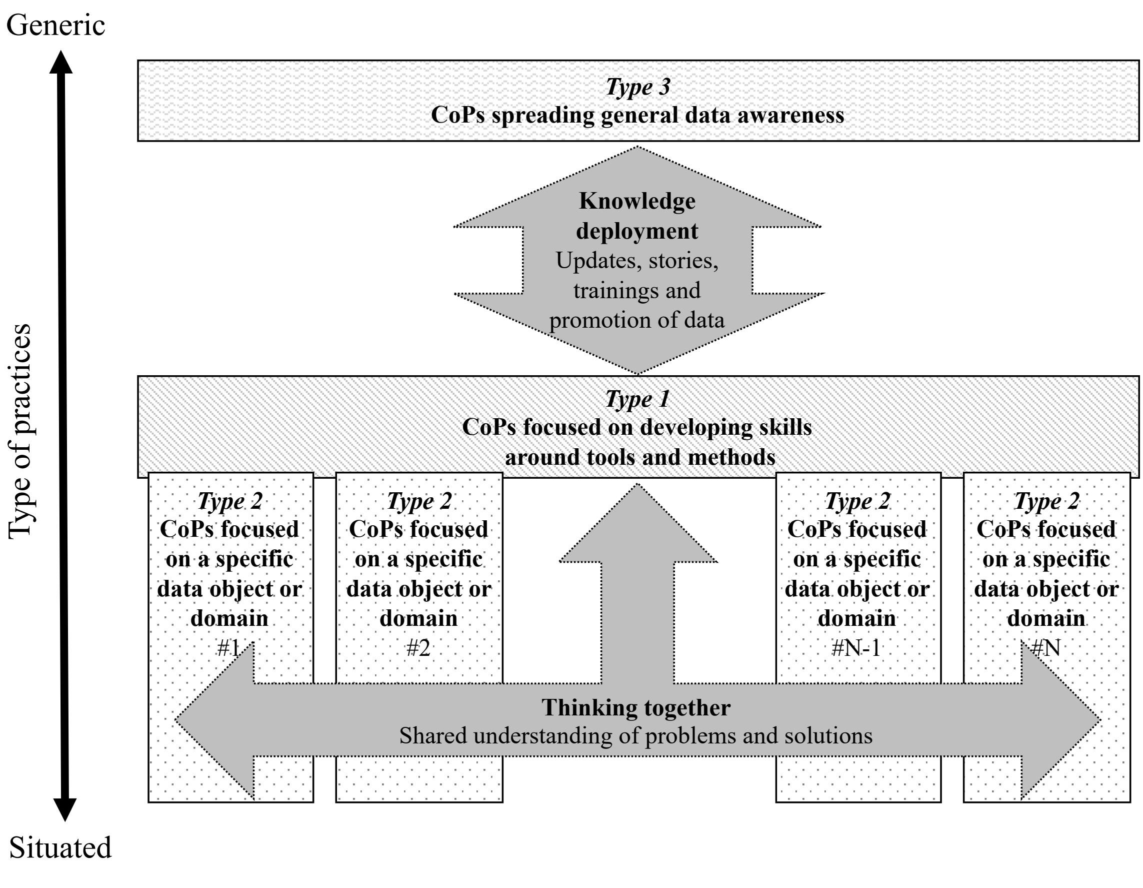

Figure 1. Landscape of practice for data democratization

Overall, our results show that the CoP landscape relies on epistemic boundaries of different "flexibility." Thus, we observe practitioners crossing boundaries and becoming knowledgeable about other practices. We can then provide each CoP type with an interpretation of their boundary interactions (Wenger, 1998). Through the vignettes, we could notably extract concrete examples of boundary objects defining membership or not into the CoP. As they focus on developing skills and expertise for practices such as tools or methods, boundary interactions for Type 1 CoP can be considered *boundary practices*. In the context of data democratization, boundary objects are mostly practices related to artifact design and use. For instance, analytics tools (e.g., BI) are typical boundary objects defining the identity of the *Reporting and analytics* CoP. Type 2 CoP interact with Type 1 communities by bringing together practitioners from both business and data teams around specific data objects or a data domain, thus relying on *boundary encounters*. Here, boundary objects are rather linked to shared processes and data lifecycle. Owing to the specific set of activities conducted in Type 3 CoP, most of its participants are guests who do not exchange about their practice as much as in the other two types. Hence, the interactions of Type 3 CoP happen through *peripheries* and enable legitimate peripheral participation (Lave and Wenger, 1991), mainly with discourses as boundary objects.

To complement the landscape, we differentiate between situated practices, i.e., practices fostered within members' own working area, and generic practices that influence the whole community landscape (Pyrko, Dörfler and Eden, 2019). This distinction is of particular relevance as companies should not only develop employees' data and analytics skills, but develop channels for the promotion of data value (Lefebvre, Legner and Fadler, 2021). We further label knowledge exchanges, which serve to generate a shared understanding of the problem and solutions between Type 1 and Type 2 CoP, as *thinking together*. This contrasts with generic practices like less intensive exchanges, which "*can take the form of exchanging facts or stories at the various layers of CoP periphery*" (Pyrko, Dörfler and Eden, 2019, p. 489). Overall, we find that Type 3 CoP focus on deploying knowledge on how to work with data, while Types 1 and 2 rather support practice exchange that support stakeholders to learn from others, share their expertise, or find solutions to their daily work challenges.

5 Discussions and conclusion

We observe that communities of practice (CoP) play a critical role in democratizing data. They connect remote employees working with data in various reporting lines, across teams, and around a shared practice. Our results uncover three types of CoP foster data democratization interacts: Type 1: CoP focused on developing skills around tools and methods, Type 2: CoP focused on a specific data object or data domain, and Type 3: CoP spreading general data awareness. Together, they form a landscape of practice displaying multilevel practices where practitioners might be members of several CoP and, thus, collaborate with various roles in the network. We also shed light on the type of practice promoted by each type of CoP. Type 1 and Type 2 CoP bring data roles in business and data experts to interact across a range of data products or data domains. Hence, their practices are rather situated. Type 3 CoP deploy more general practices and aim to spread awareness to a large audience but in a less specialized manner. Generic practices might then resonate with diverse cultural fields in the enterprise. This was highlighted in Vignette 3, where a majority of community members came from a cultural field other than data. The main outcomes expected from such rather generic practices are the creation of a data democratization culture (Hyun, Kamioka, and Hosoya, 2020) and a shared vocabulary about data (Sternkopf and Mueller, 2018). We further observed a clear core group in Type 3 CoP, hence adding nuance to existing research stating that "*generic practices do not have a clear core group*" (Pyrko, Dörfler and Eden, 2019, p. 489). However, based on the structuring characteristics of the three types of CoP, we acknowledge that CoP need to be legitimized as the network grows. Although CoP might start as spontaneous, their growth could lead to a need for further coordination and support from a management authority.

Overall, our results advance the scarce body of knowledge on data democratization by building on well-established research concepts. Observing data democratization via the concept of communities of practice provides a pragmatic understanding of these informal structures that complement more formal structures and allow people to align and collaborate. We also highlight the importance of certain key data roles for data democratization (e.g., data steward, data owner, data architect) by elucidating their relevance in establishing a link between governance and operational orientations. We then argue that data governance cannot be ignored when investigating data democratization as the former supports the definition of roles and responsibilities, and the development of formal mechanisms for collaboration and practice exchange. As data organizations become pervasive, they require a dynamic reshuffling of their informal structures to enable the growing network of data users (Peppard, 2018). In that context, CoP related to data management are critical to define, coordinate and roll out the standards and definitions required to operate data lifecycle. Furthermore, our results show that although CoP might emerge spontaneously, they usually end up being recognized at different levels as they progressively need to establish proper communications channels or dedicated platforms (e.g., Yammer). This aligns with research on CoP that mentioned the struggle faced by growing communities to remains self-managed (McDermott and O'Dell, 2001; Swan, Scarbrough and Robertson, 2002; Barrett *et al.*, 2004).

Moreover, we argue that CoP might emerge to democratize data around specific data use cases driven by a data strategy, hence be rather temporary as shown for Type 2 CoP. Also, while described as permanent, Type 3 CoP observed at Companies G and U, which are among the least advanced companies in terms of data and analytics management practices, might eventually be temporary if we consider a longer-term perspective and could rather be tied to the existence and duration of the associated corporate initiative. Overall, the interplay between data strategy, data governance and data democratization provide interesting opportunities for further research. Companies also face industry limitations or regulations preventing them to provide universal access to data hence hindering value creation from it (Awasthi and George, 2020; Lefebvre, Legner and Fadler, 2021). Thus, they should understand their optimal portfolio of formal and informal mechanisms to best democratize their data. This provides exciting research opportunities, using control theory for instance (Wiener *et al.*, 2016).

For practitioners, we provide valuable insights on which data communities could be relevant to support data democratization in their own company. In particular, our vignettes illustrate real-life scenarios on which they can rely. Firms may also use our results to understand how they can set up or organize communities when certain roles have been defined and assigned.

Our study has limitations. Since we engaged only with multinational firms with large headcounts and global operations, our analysis cannot be generalized to smaller structures. In fact, small and medium-sized enterprises might have more straightforward ways of exchanging, especially if they are all located on the same premises. Second, our data collection was done by engaging with more data management experts than analytics experts. However, our results do not seem to show excessive bias toward data management communities, especially when observed from the landscape of practice perspective. Third, our study is based on a snapshot of CoP in the case companies and, therefore, does not provide an extensive review of their evolution or lifecycle. This could provide interesting opportunities for future research (e.g., through a longitudinal case study). We also observed during our research that due to the COVID-19 context, companies have reorganized or reframed their CoP as virtual communities when relevant. This offers new research avenues to study these organizational changes and their impact more closely.

References

Awasthi, P. and George, J.J. (2020) 'A case for Data Democratization', in *Proceedings of the 26th Americas Conference on Information Systems (AMCIS)*. Virtual conference.

Baijens, J., Huygh, T. and Helms, R. (2021) 'Establishing and theorising data analytics governance: a descriptive framework and a VSM-based view', *Journal of Business Analytics*, pp. 1–22.

Barrett, M. *et al.* (2004) 'Learning in Knowledge Communities: Managing Technology and Context', *European Management Journal*, 22(1), pp. 1–11.

Benbasat, I., Goldstein, D.K. and Mead, M. (1987) 'The Case Research Strategy in Studies of Information Systems', *MIS Quarterly*, 11(3), pp. 369–386.

Brown, J.S. and Duguid, P. (2001) ‘Knowledge and Organization: A Social-Practice Perspective’, *Organization Science*, 12(2), pp. 198–213.

Carlile, P.R. (2004) ‘Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries’, *Organization Science*, 15(5), pp. 555–568.

D’Ignazio, C. and Bhargava, R. (2015) ‘Approaches to Building Big Data Literacy’, in *Proceedings of the Bloomberg Data for Good Exchange Conference*. New York, p. 6.

Dube, L., Bourhis, A. and Jacob, R. (2006) ‘Towards a Typology of Virtual Communities of Practice’, *Interdisciplinary Journal of Information, Knowledge, and Management*, 1, pp. 069–093.

Fadler, M., Lefebvre, H. and Legner, C. (2021) ‘Data governance: from master data quality to data monetization’, in. *Proceedings of the 29th European Conference on Information Systems (ECIS)*, An Online AIS Conference.

Fadler, M. and Legner, C. (2021) ‘Toward big data and analytics governance: redefining structural governance mechanisms’, in *Proceedings of the 54th Hawaii International Conference on System Sciences*. Hawaii, USA.

Feng, J. (2017) *How Airbnb Democratizes Data Science With Data University*, *Medium*. Available at: https://medium.com/airbnb-engineering/how-airbnb-democratizes-data-science-with-data-university-3eccc71e073a (Accessed: 6 April 2021).

Gherardi, S. (2000) ‘Practice-Based Theorizing on Learning and Knowing in Organizations’, *Organization*, 7(2), pp. 211–223.

Hagen, J. and Hess, T. (2021) ‘Collaboration for Big Data Analytics: Investigating the (Troubled) Relationship between Data Science Experts and Functional Managers’, in *Proceedings of the 54th Hawaii International Conference on System Sciences*. Hawaii, USA.

Hyun, Y., Kamioka, T. and Hosoya, R. (2020) ‘Improving Agility Using Big Data Analytics: The Role of Democratization Culture’, *Pacific Asia Journal of the Association for Information Systems*, 12(2), pp. 35–63.

IDC (2020) *Rethink data*. Consultancy Report. IDC.

Labadie, C. *et al.* (2020) ‘FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs’, in *Proceedings of the 22nd IEEE Conference on Business Informatics (CBI)*. Antwerp, Belgium, pp. 201–210.

Lave, J. and Wenger, E. (1991) *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lefebvre, H., Legner, C. and Fadler, M. (2021) ‘Data democratization: toward a deeper understanding’, in *Proceedings of the 42nd International Conference on Information Systems (ICIS 2021)*. Austin, Texas, USA.

Lennerholt, C., Van Laere, J. and Söderström, E. (2021) ‘User-Related Challenges of Self-Service Business Intelligence’, *Information Systems Management*, 38(4), pp. 309–323.

McDermott, R. and O’Dell, C. (2001) ‘Overcoming cultural barriers to sharing knowledge’, *Journal of Knowledge Management*, 5(1), pp. 76–85.

McKinsey (2021) *Skill building at scale during the pandemic*. Available at: https://www.mckinsey.com/business-functions/people-and-organizational-performance/our-insights/building-workforce-skills-at-scale-to-thrive-during-and-after-the-covid-19-crisis?cid=other-eml-alt-mip-mck&hdpid=977c5153-e998-46e5-ba26-5e97352982cc&hctky=12385363&hlkid=3c5aa2277dee45b39e61a550e9278470 (Accessed: 30 October 2021).

Panetta, K. (2021) *A Data Literacy Guide For D&A Leaders*, *Gartner*. Available at: https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guide-to-data-literacy (Accessed: 12 January 2022).

Peppard, J. (2018) ‘Rethinking the concept of the IS organization: Rethinking the concept of the IS organization’, *Information Systems Journal*, 28(1), pp. 76–103.

Pyrko, I., Dörfler, V. and Eden, C. (2019) ‘Communities of practice in landscapes of practice’, *Management Learning*, 50(4), pp. 482–499.

Springer, R. and Cooper, D. (2020) ‘Data Communities: Empowering Researcher-Driven Data Sharing in the Sciences’, *International Journal of Digital Curation*, 15(1), p. 7.

Sternkopf, H. and Mueller, R.M. (2018) ‘Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations’, in *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*. Big Island, Hawaii.

Swan, J., Scarbrough, H. and Robertson, M. (2002) ‘The Construction of `Communities of Practice’ in the Management of Innovation’, *Management Learning*, 33(4), pp. 477–496.

Swanson, E.B. (2021) ‘When Data Becomes Infrastructure and Our Lives Depend On It’, in. *Proceedings of the 29th European Conference on Information Systems (ECIS)*, An Online AIS Conference, p. 16.

Tenopir, C. *et al.* (2011) ‘Data Sharing by Scientists: Practices and Perceptions’, *PLOS ONE*, 6(6).

Wenger, E. (1998) *Communities of Practice: Learning, Meaning, and Identity (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge: Cambridge University Press.

Wenger, E. (2001) ‘Supporting communities of practice: a survey of community-oriented technologies.’, *Selfpublished at www.ewenger.com/tech.* [Preprint].

Wenger, E., McDermott, R.A. and Snyder, W. (2002) *Cultivating communities of practice: a guide to managing knowledge*. Boston, Mass: Harvard Business School Press.

Wenger-Trayner, E. *et al.* (2014) *Learning in Landscapes of Practice: Boundaries, identity, and knowledgeability in practice-based learning (1st ed.)*. Routledge.

Wiener, M. *et al.* (2016) ‘Control Configuration and Control Enactment in Information Systems Projects: Review and Expanded Theoretical Framework’, *MIS Quarterly*, 40(3), pp. 741–774.

Yin, R. (2003) *Case Study Research: Design and Methods, Third Edition, Applied Social Research Methods Series, Vol 5*. London, UK: Sage Publications, Inc.

Zeng, J. and Glaister, K.W. (2018) ‘Value creation from big data: Looking inside the black box’, *Strategic Organization*, 16(2), pp. 105–140.