Sci

MDPI

Article

Towards a Democratization of Data in the Context of Industry 4.0

Tobias Harland [1], Christian Hocken [1,*], Tobias Schröer [2,*] and Volker Stich [2]

[1] i4.0MC—Industrie 4.0 Maturity Center GmbH, 52074 Aachen, Germany; tobias.harland@i40mc.de
[2] Institute for Industrial Management at RWTH Aachen University, 52074 Aachen, Germany; st@fir.rwth-aachen.de
*   Correspondence: christian.hocken@i40mc.de (C.H.); sr@fir.rwth-aachen.de (T.S.)

Abstract: Data-driven transparency in end-to-end operations in real-time is seen as a key benefit of the fourth industrial revolution. In the context of a factory, it enables fast and precise diagnoses and corrections of deviations and, thus, contributes to the idea of an agile enterprise. Since a factory is a complex socio-technical system, multiple technical, organizational and cultural capabilities need to be established and aligned. In recent studies, the underlying broad accessibility of data and corresponding analytics tools are called "data democratization". In this study, we examine the status quo of the relevant capabilities for data democratization in the manufacturing industry. (1) and outline the way forward. (2) The insights are based on 259 studies on the digital maturity of factories from multiple industries and regions of the world using the acatech Industrie 4.0 Maturity Index as a framework. For this work, a subset of the data was selected. (3) As a result, the examined factories show a lack of capabilities across all dimensions of the framework (IT systems, resources, organizational structure, culture). (4) Thus, we conclude that the outlined implementation approach needs to comprise the technical backbone for a data pipeline as well as capability building and an organizational transformation.

1. Introduction

The fourth industrial revolution is the only industrial revolution that was announced before it actually happened [1]. However, there were good reasons to believe that industry will go through an unprecedented development back in 2011. A set of new technologies matured and influenced each other in a synergetic way. Amongst others, technologies, such as cloud computing, the (industrial) internet of things, mobile computing and artificial intelligence, made the convergence of the digital and physical world possible and together, enabled the concept of "cyber-physical-systems" [2].

These technologies, indeed, have the potential to not only incrementally improve processes within the current organizational setup but to transform the entire way of working, as well as the business models of whole industries and, thus, should be considered an industrial revolution [3].

The three previous industrial revolutions all were, at least partly, driven by major advances in certain technological fields as well. Those were mechanization, electrification and computerization. Besides technological advances, the three industrial revolutions were accompanied by major changes to the way a company is organized [4]. Manual labor was substituted by machines (first industrial revolution), mass production and the scientific management divided processes in increments and optimized them (second industrial revolution) and automation of repetitive tasks led to the emergence of the knowledge worker (third industrial revolution) [5].

The implications of the abovementioned technologies in the fourth industrial revolution on labor are discussed controversially [6]. However, it is not arguable that there still

will be humans working in industrial companies and that they will have the opportunity to work with much more data and will have access to much more sophisticated tools to gain insights from the data [7]. Currently, a promising concept is being discussed that is supposed to provide the organizational framework to utilize the newly available data and tools: data democratization [8].

At the core of the concept of "Data Democratization" is the idea to provide access to the company's data resources to all employees "given reasonable limitations on legal confidentiality and security" [9] (p. 1). This also implies that data should not only be available to technical experts but also to non-technical staff from the company [10]. Other authors highlight not only the access to data but the company culture of "willingness to share information" [11] (p. 5). The purpose of the concept is described as the "ability of users (. . .) to answer unexpected questions" [12] (pp. 1362–1368). This purpose is also well aligned with the overall purpose of the fourth industrial revolution, because many authors see agility and, thus, the quick and precise response to an unexpected event, as one of the key motivations to transform companies [13].

With this study, we intend to find out how advanced companies in the manufacturing sector operate in terms of a democratized use of their data and how the concept is related to what is commonly understood as Industry 4.0. Referring to a broadly used transformation framework (acatech Industrie 4.0 Maturity Index), the study focuses on a company's capabilities in the four structuring forces: "information systems", "resources", "organizational structure" and "culture".

2. State of the Art

Nowadays, the amount of information that is published is rapidly increasing [14]. Accordingly, the demand for experts who can process this data is high [15]. Data democratization allows not only data experts, but also non-specialists in companies to work with data [8]. Proper use of data can lead to increased revenue and brings the idea of data democratization back to the forefront [16], so that researchers have recently begun to look at data democratization [17]. Comprehensive research results are not available, yet. However, some focus areas must to be mentioned. Descriptive research with a focus on explaining the concepts and normative efforts has been conducted by numerous authors [8,18,19]. Based on those concepts, several case studies have been conducted on establishing the concepts in specific usage scenarios, including offshore drilling [20], banks and B2C businesses [9], real estate [21], medicine [22] and others. Surprisingly, no relevant studies on data democratization relating to manufacturing applications and Industry 4.0 are known to the authors. The processing and use of data play major roles in the context of Industry 4.0. By evaluating data, for example, errors can be detected at an early stage [23]. With 5G, larger volumes of data can be transmitted faster and more securely in the sense of Big Data [24]. The latest technologies, such as blockchain, can provide more data of higher quality and improve the operation of cyber–physical systems [25]. Through the use of smart contracts, processes can be established between IoT devices without intermediaries [26]. The IoT is related to data collection. It is associated with an object that is equipped with sensors to exchange data with other devices [27]. The generation, processing and use of data are, therefore, directly related to Industry 4.0 applications. In principle, the concept of data democratization is easily transferable on Industry 4.0 applications. However, it has not yet been examined from the perspective of manufacturing companies. Therefore, it seems reasonable to study the readiness of those companies for pursuing data democratization efforts.

3. Material and Methods

In order to answer the research question: "To what extent do industrial companies today have the necessary capabilities to apply the concept of data democratization?", an existing data set on the digital maturity of factories across the globe was analyzed. The data set is the result of 259 assessments of factories across the globe using the framework of the acatech Industrie 4.0 Maturity Index [13]. The framework was developed as a tool

to measure the digital maturity of a factory. It focusses on the business process within the factories and describes the way towards a "learning, agile company". The framework is structured in six maturity levels and four structuring forces. The maturity levels can be summarized as follows:

- Computerization: information technologies are used in the core processes in an isolated way;
- Connectivity: information technology (IT) as well as operational technology (OT) systems offer connectivity and interoperability;
- Visibility: real-time data for end-to-end processes exist;
- Transparency: collected data are aggregated and contextualized, big data applications are deployed in parallel to the business applications;
- Predictability: the ability to simulate future scenarios for the core processes exists and it is possible to anticipate future developments;
- Adaptability: systems autonomously trigger actions and adapt to changing conditions.

The necessary capabilities per maturity level are organized along the structuring forces resources, information systems, organizational structure and culture. A central concept in the framework is that the development in the four structuring forces has to be synchronized to unfold the full potential [13].

The data were collected from 2017 to 2021. The factories originate from various industries with a certain focus on automotive, medical and food and beverage. A detailed distribution of industries is depicted in Figure 1.
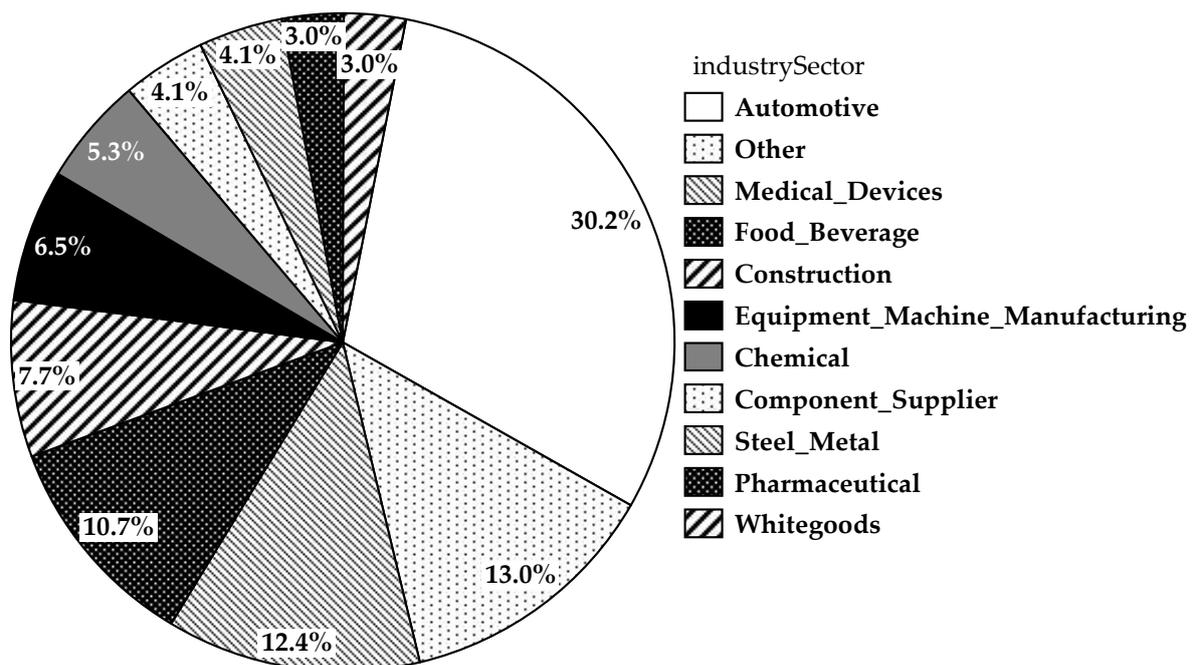
Figure 1. Distribution of industries.

In addition, the assessed factories are located across the globe with Turkey, Germany, the US and the Netherlands making up more than 50% (see Figure 2).

To be able to assess the dimensions of the framework in a standardized and assessor-independent way, a questionnaire was crafted. The questionnaire describes scenarios for each maturity level and each assessed dimension. The procedure for collecting the data was standardized and comprises interviews with experts from each core process in the assessed factory, Gemba walks following the value streams and group discussions to review the results.
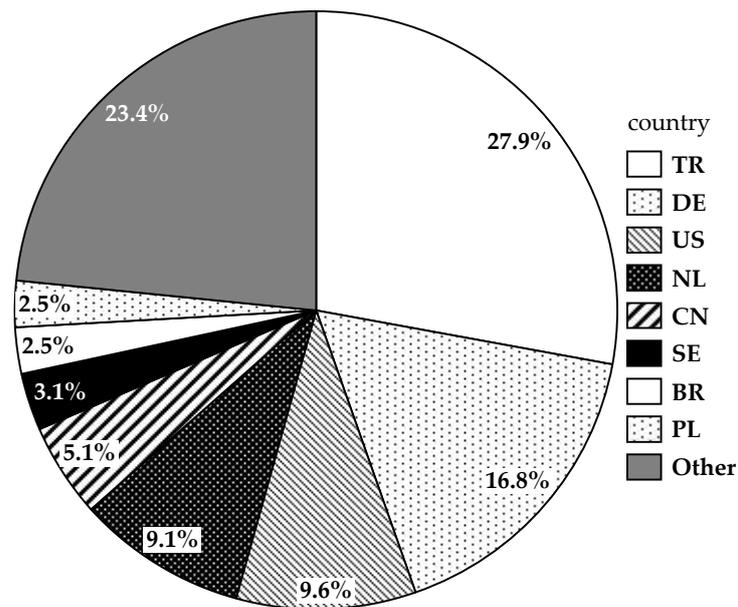
Figure 2. Country distribution.

Figure 3 shows that factories with more than 1500 employees, factories with 500–1500 employees and factories with 100–500 employees are evenly distributed. Only small factories with less than 100 employees are underrepresented.
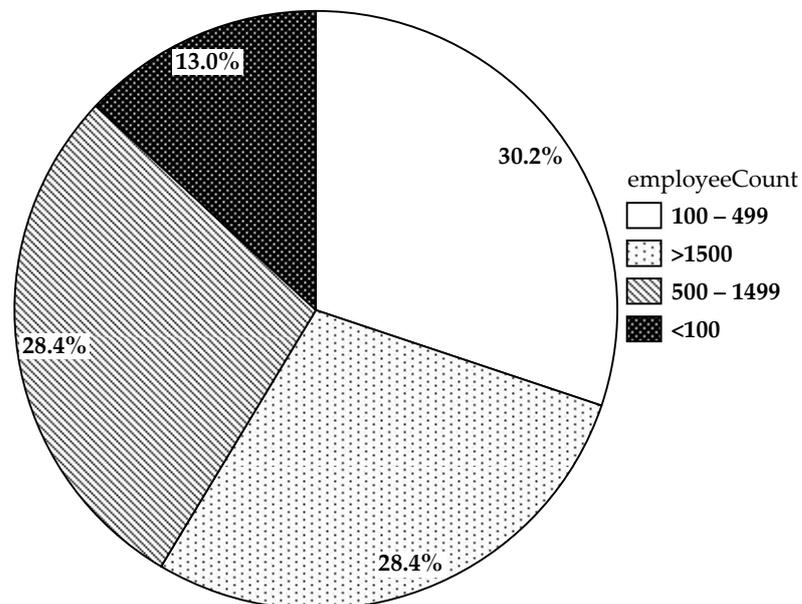
Figure 3. Number of employees.

The data set provides insights to many more capabilities than used for this study. In total, depending on the process to be analyzed, around 40 capabilities were assessed for each core process in the factories. However, we selected 10 of these capabilities for this study that provide insights to validate or falsify our hypotheses on the current state of the factories.

4. Results

The acatech Industrie 4.0 Maturity Index proposes that the digital maturity of a factory has to develop in its four structuring forces ("IT systems", "resources", "organizational structure" and "culture") in a synchronized way in order to unfold its full potential [13]

(p. 21). Following this idea, we formulated a hypothesis building on the abovementioned definitions of "data democratization" for each of the structuring forces accordingly. For the structuring force "IT systems" two hypotheses were examined.

4.1. IT Systems

Hypothesis 1. *The necessary data to provide a sufficiently detailed view on the current and past condition of processes and objects in the factory are available in a digital format and of an accurate quality.*

The existence of digital, accurate data for the area of the company under consideration is obviously an essential basis for the idea of data democratization. If there are no data to be shared to draw conclusions from, there is no sense in democratizing them.

To investigate this hypothesis, we examined the two capabilities "data quality" and "decision support (IT)". In this context, we understand data quality as the property of the data to be complete, accurate, current and consistent [28]. The maturity of this capability was characterized by the following statements (see Table 1). The percentage of answers per described scenario is depicted as well in the following table.

Table 1. Results for the capability "data quality".

Level	Scenario	Answers
1	"Data quality is not sufficient to further process the data."	22.92%
2	"The data quality is sufficient for further data processing, but there is a partly redundant collection of and storage of the data."	52.37%
3	"Redundant collection and storage of data with reference systems is avoided. The data quality allows a non-automatic further processing of the data."	19.35%
4	"The completeness, accuracy, currency and consistency of data sets are ensured. An automated data cleansing process is in place. High data quality ensures automatic further processing of the data ("fit for use")."	4.74%
5	"Regular profiling practices are used to identify errors in data sets and to ensure high data quality in the long term."	0.62%
6	"Systems for the self-healing of data sets as well as automated consistency checks and adjustments are in place."	0%

In order to comply with the concept of data democratization, the minimum target level is 4. However, most of the examined factories are currently at a lower level. More than 75% of the surveyed factories do not have the appropriate level of data quality to enable an automated use of data.

A second capability that is related to the first hypothesis is the decision support by IT systems. Decision support systems have been discussed in the literature for decades and are, for instance, defined as "interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems" [29] (pp. 1–26). Other authors confirm that these systems support decision-making in "semistructured tasks" [30] and emphasize that they "enhance the traditional information access and retrieval functions with support for model building and model-based reasoning" [31] (p. 6). The capability to create meaningful data-based insights to improve decision-making is at the center of the concept of data democratization and, thus, needs to be considered in this study.

Table 2 contains the possible scenarios for this capability. The minimum level for this capability that is necessary to implement the approach of data democratization is level 3. According to the described scenario, factories at level 3 are able to visualize operational data. At levels 1 and 2, data are not yet visually processed or decision-making is not supported with data. The data set shows that more than 68% of the assessed factories are not at that

level. However, the fact that 47.82% of factories at least collect data on the current machine status indicates that some basics for databased decisions on their operations are present.

Table 2. Results for the capability "decision support (IT)".

Level	Scenario	Answers
1	"Decision-making is not supported by systems."	20.84%
2	"Machine status and condition data of the machine park are available in the leading system, but not visually processed and therefore not directly available for decisions."	47.82%
3	"Visualizations are created with dashboards (e.g., target / actual quantity, downtimes)."	24.59%
4	"Users are informed about standstills and limit exceedances by means of alerting and escalation processes."	6.23%
5	"In addition to level 4, effects are shown in advance."	0.51%
6	"Where possible, decisions are made automatically and presented to the user in a comprehensible manner."	0%

Comparing these results with other studies on data quality and databased decision support in manufacturing, the results even indicate a larger gap in these capabilities compared to the findings of other studies. For instance, Spath et al. found that 58.9% of surveyed companies are not able to detect relevant events in their production automatically and, thus, based on data, 43.7% of participants see the missing actuality of production data as the root cause of manual interventions to production planning [32]. Lanza et al. found, in their study on digital shopfloor management, that today, in more than 50% of the examined cases, operational KPIs were calculated manually; in more than 66%, the visualization of these KPIs done in an analog way [33].

In conclusion, the hypothesis cannot be considered validated. Thinking of this as a prerequisite for decisions based on democratized data, one central field of action for most companies is the systematic improvement in the quality of their operational data in combination with systems that make them available for decisions.

Hypothesis 2. *The data are organized in a way that it is accessible for ad-hoc analyses.*

"Data Democratization" and "Industrie 4.0" place the stakeholders at the center. They are expected to optimize their area of responsibility through the use of data, e.g., by making decisions themselves on the basis of their own data analyses. These analyses often arise spontaneously (ad hoc) due to unexpected events stakeholders need to cope with but can also be caused by the pursuit of general performance or quality improvements.

In many cases, analyses involve several data sources, e.g., order data, material flow data or quality data. Data mostly originate from IT systems that are used for controlling or monitoring processes. However, other data sources might also be used, e.g., feedback from operators on process or product quality. In hypothesis 1, we showed that these data must be of sufficient quality and that the type and scope of the data must be sufficient so that they can be used for data-driven decision-making. However, further requirements must be met in the area of data organization so that stakeholders can perform analyses on their own. Data from different sources must be associated with one another, and data must be accessible to the user. The organization of data must meet these principles so that data can be used for ad-hoc analysis.

To investigate this hypothesis based on the available data set, we focus on two capabilities. "Data model" describes the conceptual approach to structure data. It includes information on data objects, references between objects and it defines valid operations [34]. In this study, we focus primarily on the scope of the data model, which can encompass individual processes up to an entire production network. The maturity of the capability "data model" was characterized by the following statements. The percentage of answers per described scenario is depicted as well in the following table (see Table 3).

Table 3. Results for the capability "data model".

Level	Scenario	Answers
1	"There is no data model for the process and relevant data points are unknown."	18.27%
2	"There is only an isolated data model for the process."	59.27%
3	"There is an integrated cross-process data model. Example for production: orders, production parameters can be combined with quality and maintenance data."	20.78%
4	"An internal data model exists across all sites."	1.53%
5	"There is a cross-site data model in which suppliers and customers are integrated."	0.14%
6	No scenario defined	

The results show that in more than 18% of the performed assessments, no data model was available and data points to monitor or control the process were unknown. Companies that meet this maturity level are not yet able to offer their stakeholders the basis for their own data analyses. In almost 60% of cases, data models exist that refer to individual processes and, thus, enable analyses at process level. Practice shows that in production processes, Scada/Level 2 (ISA 95) data are often used (time-series data that can include machine states, sensor values or information on product quality). However, the analysis options are limited depending on the question. In cases where interfaces to other processes are to be examined or for the determination of performance indicators, data from higher layers (ISA 95), such as from Manufacturing Execution Systems and Enterprise Resource Planning Systems, are missing. This is made possible by cross-process data models, which are currently available to almost 21% of the plants examined. It can be observed that data models are limited to the plant itself in almost 98% of the cases surveyed. Only 1.5% state that they have integrated data models for all sites. The integration of suppliers and customers into company-wide data models has not yet played a role according to our data set.

The second capability we focus on is what we call "Collaboration in IT". This expresses whether social software or enterprise collaboration systems are being used for stakeholder communication and collaboration. A digital collaboration environment with a uniform user interface is important for collaboration between users, including working together on documents and other use cases [35]. The use of such solutions has two advantages. On the one hand, data are made available in a central location and can usually be accessed and consumed by users with the appropriate rights. On the other hand, they promote the formalization of tacit knowledge [36]. Table 4 contains the possible scenarios for this capability.

Table 4. Results for the capability "Collaboration in IT".

Level	Scenario	Answers
1	"There is no collaborative platform for employees."	7.98%
2	"Collaboration is supported through static IT systems, such as email accounts, file servers, and folder structures."	61.36%
3	"A collaborative platform is in place which facilitates the management of tasks and projects"	28.32%
4	"The collaborative platform is in addition used for the communication between employees. All internal communication takes place via the platform, so that email correspondence is reduced to a minimum."	2.26%
5	No scenario defined	0.08% *
6	No scenario defined	

* Due to false statement in one assessment.

The minimum level for this capability that is necessary to implement the approach of data democratization is level 4. The data set shows that, so far, only 2.6% of the companies

surveyed have such a solution in place. However, just under a third (28.3%) are on their way there and have at least a solution in place that facilitates the management of tasks and projects. However, a large proportion of the companies surveyed still rely on classic file sharing via a network drive (61.36%). Just under 8% forgo the use of collaboration solutions altogether.

The research shows that we are still in the early stages of organizing data for use in ad-hoc analyses. Although data are available, they can usually only be correlated at the individual process level. Overarching models that encompass the entire value-creation process of a factory or production network are scarce. A similar picture can be drawn for the joint creation and use of data. Although data are shared in many cases, classic approaches, such as network drives, are used for this purpose. Our data set shows that comprehensive collaboration platforms have not been used much so far. However, due to the long duration of the study, it is possible that collaboration platforms are underrepresented in our results. Other studies seem to indicate that many companies adopted such platforms during the COVID-19 pandemic [37].

4.2. Organizational Structure

Hypothesis 3. *The organizational structure in the factory empowers employees to proactively improve their routines and initiate and implement improvements on their own.*

The effective use of data democratization requires that employees are involved in improving their working conditions, tasks and routines. It even goes one step further and relies on employees themselves identifying potential for improvement and implementing measures. The organization should not only tolerate this behavior but should actively promote it. It has long been recognized that improvements brought in by employees contribute to the company's success. Continuous improvement approaches are, for example, an important part of management and work organization methods, such as Lean [38] and Kaizen [39–41]. Our hypothesis is that the same mechanisms could be applied to the ideas of democratization of data. It is important here that stakeholders accompany both the analysis and the implementation of measures.

To examine our hypothesis, we look at two capabilities in the data set. First, we look at whether companies have implemented an "innovation process" and how far reaching it is. This capability provides a good understanding of whether and how employees are involved in improvement processes. The maturity level of the "innovation process" capability was characterized by the following state characteristics. The percentage of responses per described scenario is also shown in the following table (see Table 5).

In order to comply with the concept of data democratization, the minimum target level is 4. About 23% of the examined factories currently are, at least, at this level. We observe that almost all of the factories studied have implemented an innovation process, with only around 8% not doing so at all. In just under 5% of the plants surveyed, idea generation and idea development are based on the open innovation principle. Employees at all levels contribute to idea generation and evaluation. Measures and their effectiveness are systematically monitored, documented and, if necessary, adjusted in accordance with the PDCA cycle. As such, 33% have at least one process for collecting suggestions for improvement (level 2), and another 35% circulate information about ideas submitted, evaluation and implementation (level 3). Thus, in more than half of the companies, there is an active innovation process that leads to implemented improvements. Only the participation of employees in the implementation of measures is not yet sufficiently well developed. Overall, however, these are good conditions for data democratization.

"Decision power and responsibility" is the second capability we examine. This is about how and by whom decisions are made in the organization and how decision-making is supported. We are particularly interested in the hierarchical level at which decisions are

made for operational activities. The defined maturity levels and the associated scenarios are listed in Table 6.

Table 5. Results for the capability "innovation process".

Level	Scenario	Answers
1	"Improvements and innovations are not considered responsibilities of operative employees."	8.16%
2	"A process exists through which employee ideas can be collected and evaluated (e.g., suggestion scheme, CIP)."	33.07%
3	"In addition to level 2, there is visibility about the ideas submitted, their evaluation and their implementation."	35.56%
4	"There is a process in place to evaluate and subsequently implement ideas. Employees are involved in the implementation of their ideas."	17.60%
5	"Idea generation and idea development take place according to the open innovation principle. Employees at all levels contribute to the generation and assessment of ideas. Measures and their effectiveness are systematically monitored, documented, and, possibly, modified in accordance with the PDCA cycle."	4.61%
6	"Employees are given sufficient leeway to develop new ideas and concepts and test them."	0.99%

Table 6. Results for the capability "Decision power and responsibility".

Level	Scenario	Answers
1	"Operative employees do not have the power to make decisions. All decisions are made at the management level. Clear hierarchies with inflexible reporting lines have been established."	15.19%
2	"Decisions on operations activities are made by operative employees (within the group). There is no support in the decisions process by IT systems. Uncertainties are clarified and conflicts are resolved by (first-line) managers or supervisors."	35.09%
3	"Decisions on operative activities are made by operative employees (within the group). Information to support the decision-making process are provided by IT systems. Uncertainties are clarified and conflicts are resolved by (first-line) managers or supervisors."	45.79%
4	"Decisions, including those on operational activities, are made by the employee with the most expertise. This is not necessarily a supervisor or manager. A holocratic approach is taken.	3.92%
5	No scenario defined	
6	No scenario defined	

At least level 3 is required to use data democratization. Decisions concerning operational activities must be able to be made by operators. In addition, however, there must also be support from IT systems that hold data relevant to decision-making. We see that just under half of the factories studied meet these requirements (49.71%). Among them, 3.92% even choose holocratic approaches for decision-making. In around 35% of the factories surveyed, decisions are made as a team (level 2). However, this group lacks the necessary IT support. Only slightly more than 15% still rely on distinct hierarchical decision-making structures.

Both capabilities studied show that we are at a turning point. It is apparent that organizations want to involve employees more and are increasingly granting them freedom to do so. Based on the available data, it can be seen that both capabilities can still be expanded

in the majority of the factories surveyed. The most important factors include employee participation in the implementation of measures and the creation of organizational conditions for the use of data in decision-making processes. The potential of employee participation in terms of data democracy still relies on the maturity of the data dimension, though.

4.3. Culture

Hypothesis 4. *The culture among the associates in factories encourages data-driven decision-making and an open sharing of knowledge and insights.*

Besides formal processes, organizational structure and assigned responsibilities, the culture among the associates has a paramount influence on how the concept of data democratization is embraced in a factory. In general, we see culture as the commonly accepted set of values within the organization that guides the actions of employees [17,42]. Specifically focusing on the concept of data democratization, these values need to comprise the "willingness to share information" [11] (p. 5). For that reason, the capability "knowledge and knowledge responsibility" is selected from the data set. In addition, the capability "data-based decision processes" was analyzed to understand to what extent the companies base their decisions on data [43].

In Table 7, the possible scenarios for the capability "Knowledge and knowledge responsibility" are listed. With every level, the importance to formalize and to share knowledge increases. Thus, this capability describes to what extent "tacit knowledge" or "explicit knowledge" guide the course of actions of the employees. "Tacit knowledge" solely exists in the minds of individuals and, thus, is not shareable and available to the rest of the organization [44]. For the idea of data democratization, the willingness to share knowledge is crucial. Due to that, the minimum level of this capability is level 3. At this level, a company collects and shares knowledge in a structured way. In nearly 30% of the assessed companies, such a culture can be observed. However, about two-thirds of the companies have a culture that does not actively encourage the formalization and sharing of knowledge.

Table 7. Results for the capability "Knowledge and knowledge responsibility".

Level	Scenario	Answers
1	"Knowledge is not shared, as it helps the individual to become indispensable. Thus, employees typically rely on their own experience."	4.75%
2	"Employees network within their respective areas and from expert groups. Employees are not overly willing to formalize their knowledge and make it available to the entire company."	63.87%
3	"Experience and practical knowledge are systematically collected and transformed into explicit knowledge. Employees are willing to share their knowledge in a formalized way with the company."	29.80%
4	"Knowledge guides the actions of all employees. Employees are willing to participate in an intensive exchange of information and voluntarily seek to understand relationships and generate new knowledge, even outside of work. The provision of knowledge is supported by appropriate IT solutions, such as knowledge management systems and collaborative platforms."	1.58%
5	No scenario defined	
6	No scenario defined	

The results in Table 8 for the capability "data-based decision processes" confirm the impression of the previous capability: About 40% of the companies already use data as

the basis for their decision-making. Still, around 60% rely on intuition and individual knowledge or only partly consider data in their decision-making processes.

Table 8. Results for the capability "Data-based decision processes".

Level	Scenario	Answers
1	"Decisions are made on the basis of the knowledge of the intuition of individual employees."	14.17%
2	"Data from IT systems is taken into account in the decision-making process, but decisions are adapted case by case on the basis of personal knowledge or the intuition of individual employees."	45.08%
3	"Current and historical data are used to support the decision-making process."	27.05%
4	"Data analyses are made to support the decision-making process."	11.86%
5	"The decision-making process is supported by simulations and scenarios."	1.83%
6	"Decision-making processes are largely automated; employees supervise the decisions and intervene in exceptional cases only."	0%

Based on the analysis of the two capabilities, the hypothesis cannot be confirmed. Even if around two-thirds of the companies have already adopted a culture of data-based decision-making and of an initiative-taking sharing of insights, still, most of the evaluated companies are lacking such a culture. Other studies partly confirm these findings. For instance, Lanza et al. found that in five out of nine analyzed case studies on digital shopfloor management, there is an exchange of knowledge, but it is an analog exchange without the support of digital tools [33]. A recent study of YouGov, with more than 3500 surveyed decision-makers, found that 56% of the companies consider themselves as data driven, which is a significantly higher share than in the analyzed data set [45]. Another study with more than 1000 participants, conducted by Capgemini in 2020, found that 50% consider their decision-making as data driven [16]. However, the results in the present analysis are based on external, independent observations, whereas the cited studies rely on a self-evaluation of the surveyed participants.

As a conclusion, it is a central field of action for these companies to create such a culture on their way towards a democratization of data, to not only have the technology ready, but also create a "pull" for data in their workforce.

4.4. Resources

Hypothesis 5. *The employees have the appropriate capabilities to work with data and interpret them in the context of their domain.*

As Belli et al. state in their definition of data democratization, it is the "ability of users to access all data using well-defined and easily used analytic patterns to answer unexpected questions" [12] (pp. 1362–1368). The definition implies, besides organizational aspects, such as the access to data, and technological aspects, such as the easily used analytic patterns, that the users themselves need to be capable. To examine the necessary skills of the employees, the two capabilities "IT competencies" and "Interdisciplinary skills" were analyzed.

As depicted in Table 9, the surveyed scenarios range from a rudimentary utilization of existing system functionalities with manual bypasses to the system to a full utilization of the features and an independent covering of information needs. In only 3.69% of the companies, the employees can cover their information needs themselves, which would be the required level for the concept of data democratization. The vast majority are only capable of handling basic system features. Still, in around one-quarter of companies, it can

be observed that the associates are able to understand the logics and dependencies of the systems they are using. This leads to the conclusion that, at least in this group of companies, a certain awareness for the multiple data in their IT systems and its value exists.

Table 9. Results for the capability "IT competencies of employees".

Level	Scenario	Answers
1	"Employees can only partly handle the functions that are necessary for their operational activities. This leads to an avoidance of systems or using Excel-solutions although application systems exist."	9.02%
2	"Employees can only handle functions that are necessary for their operational activities."	59.36%
3	"Employees are aware of the logics and dependencies of the used system functions."	27.94%
4	"Employees are able to independently operate systems beyond their standard functionalities in order to satisfy their information needs, e.g., analytical applications required for their daily work."	3.69%
5	No scenario defined	
6	No scenario defined	

Table 10 contains the scenarios for the capability "interdisciplinary skills of employees". The scenarios range from employees having a very limited scope for their own considerations to employees who understand the dependencies with and their impact on other processes and activities in the company. This capability is important for a successful implementation of a data democratization approach, as the whole idea is based on accessing the relevant data from the whole organization and not only from their own limited domain. In order to make sense of these data, it is inevitable that the context the data is sourced from is understood.

Table 10. Results for the capability "Interdisciplinary skills of employees".

Level	Scenario	Answers
1	"Employees focus on subject- or domain-specific questions and problems."	24.59%
2	No scenario defined	3.04% *
3	"Employees know about neighboring process steps/activities, include them in their considerations and exchange information."	61.67%
4	"Employees are aware of the impact of their activities on neighboring process steps/activities, include them in their considerations and proactively address interdisciplinary problems and questions on their own."	10.70%
5	No scenario defined	
6	No scenario defined	

* Due to revision questionnaire.

The results show that the associates of only around 25% of the companies strictly focus on their own domain. A large group either is aware and considers neighboring processes or even proactively involves them in their activities.

Looking at these two capabilities, a slightly higher readiness can be observed compared to other examined capabilities in this study. In addition, it is not visible in the data whether the capability "IT competencies of employees" would be on an even higher level if the companies were to provide their employees with more capable systems. After all, employees were not able to demonstrate the capabilities required for the higher levels, even if they had had them. It is obvious that the competencies required to work with data effectively rely on the data itself being available and structured, which cannot be considered true, as already mentioned earlier.

The comparatively high maturity level in the capability "Interdisciplinary skills of employees" might be related to an increasingly high maturity level in lean methodologies [46]. The concepts of interdisciplinary collaboration and process-oriented thinking are core ideas of lean management [47]. This, once again, supports the importance of considering Industry 4.0 and Lean Management or comparable approaches in an integrated way [48].

5. Discussion

The presented study was conducted to answer the question "To what extent do industrial companies today have the necessary capabilities to apply the concept of data democratization?" We translated the overall concept of data democratization in the first step to a set of hypotheses on the existence of the relevant capabilities in manufacturing companies. Theses hypotheses were evaluated by utilizing an existing data set containing detailed data on the current digital maturity of 259 factories from multiple industries and regions of the world. The results of this study were related and compared with existing studies (see description of results per examined capability). In most of the cases, other studies could confirm the findings. However, in general, the findings of other studies were slightly more optimistic regarding the existence of certain capabilities in the industry. We relate this difference to the fact that many other studies rely on a self-evaluation of the surveyed person. We believe that the used data set for this study provides a much more realistic view on the status of manufacturers, since all data were collected from an external team of assessors.

Still, this study has its limitations. The used data set was collected over the period of four years and, thus, does not, for all evaluated companies, reflect the latest development stage. In addition, the evaluation process requires a certain investment of time and money from each company. Due to that, companies who feel the need themselves to improve in terms of digital maturity would typically go through the process in order to identify opportunities to improve. Companies that already perform at a high level tend to not go through the evaluation. Due to that, the data set might be biased. Furthermore, there might be capabilities that are relevant for the concept of data democratization but that are not part of the used data set.

Thus, further research can focus on finding and filling these potential existing gaps and, with this, complement the results of this study. In addition, the results of this study can be used as a starting point for the development of a more detailed concept of "Data Democratization for Manufacturing" that is closely aligned with the current state of relevant capabilities in the industry and, thus, is tailored to the needs of practitioners.

Lastly, the results can be used directly in manufacturing companies to benchmark their status quo and to start shaping their own roadmap towards a democratization of data in their own factories.

6. Conclusions

The conducted analysis clearly observed a strong interdependency between the maturity of a company in terms of Industry 4.0 and the ability to utilize data in a democratized manner. Considering this fact, it becomes an obvious ambiguity to think of data democracy as an enabler for monetizing data in terms of business cases, on the one hand, and having many aspects of Industry 4.0 as a prerequisite, on the other hand. In short terms: structured, consistent and available data are necessary to start a data democracy—an initiative of which the business value can hardly be estimated ex ante. Therefore, we suggest a set of essential capabilities based on our studies (Figure 4).

We suggest starting the process from the dimension "culture", in order to first create a "pull" from the organization instead of first developing the technology with the risk of missing essential user requirements or overengineering. Based on the findings, most companies still base many decisions on individual experience and intuition and not on data. Thus, procedures to obligatorily base decisions on data need to be established. To

initiate this cultural change, an option might be to let the executives serve as role models for this kind of decision-making and let a data-driven culture grow from this nucleus [49].

Make data-based decision-making mandatory & let executives serve as role models

Incentivize the sharing of data, insights and knowledge

Build capabilities to develop "citizen analysts"

Leverage existing cross-functional team structures as a starting point

Improve data quality of operational data

Make the collected data accessible in an automated way

Develop a unified data model that comprise all domains of the factory

Push collaboration platforms that enable a seamless and direct communication

Build on existing approaches on involving employees in improvements

Foster decentralized decision-making by leveraging existing structures and combining them with real-time insights available for everyone

Data Democratization for Manufacturing

Culture   Resources   Organization   IT Systems

Figure 4. Essential fields of Data Democratization for Manufacturing.

Another finding of the study is that the motivation to formalize and share knowledge can become a roadblock for the democratization of data. Therefore, a field of action is to set specific incentives to foster a mindset of knowledge sharing. For instance, existing practices to incentivize the participation in suggestion programs or in lessons-learned programs can be used as a reference [50].

As a second step, we recommend considering the dimension "resources". Here, the study shows that in most companies, the employees are able to use the existing IT systems to perform their day-to-day work but lack the capabilities to cover their individual information needs. Thus, capability building to so-called "citizen analysts" might be a field of action to enable the associates. These are employees who are qualified in self-service analytics tools to cover their data needs own their own [10]. In addition, the results indicate that the existing awareness of the importance of cross-functional problem solving can be capitalized on. It is, for instance, part of many problem-solving techniques to involve experts from multiple domains to contribute to the process. Such groups may provide fertile ground for promoting the concept of data democratization.

After having a first set of associates selected and qualified, the capabilities of the IT systems have to be expanded. The examination of the related capabilities points out that the relevant data on the factory's operations are not of sufficiently good quality, nor can they be easily accessed, nor are data from different domains logically linked. That leads, on the one hand, to the conclusion that companies need to launch initiatives to improve the data quality. These initiatives might include organizational measures in the field of data governance [51] but also measures to collect data in an automated way, for instance, directly from the machine PLCs [52]. Furthermore, the accessibility of the data in the systems seems to be another roadblock. Typical measures in this field comprise the implementation of a middleware or an enterprise service bus [53]. In order to logically link data from multiple sources in an automated way, it is necessary to have a unified data model across all domains that defines the database schema for the data platform and builds the basis to map the data from various sources.

Besides the mentioned fields of actions related to data quality and data access, the analyses of the dimension IT systems show that the roll-out of social software and collaboration platforms is another field of action. They are supposed to catalyze the distribution of and communication about insights in the organization.

In order to utilize the insights from the "citizen analysts" the examined capabilities in the dimension "organization" can serve as a basis. The results indicate that many companies already have procedures in place to involve employees in the continuous improvement process. The insights gained through the analysis of data are another source to identify opportunities to improve the mentioned existing procedures to channel and route these

opportunities and build a solid foundation. The field of action here would be to enable the integration of both initiatives, the continuous improvement programs and the data democratization program.

Another examined capability in the dimension organization is the decentralization of decision-making. The rationale here is that if associates are able to gain insights on their own, they also need to be able to act on these insights on their own. The results indicate that many companies are organized in a way that enables these decentralized decisions. Thus, this is another capability that acts synergistically on a successful data democratization program.

In total, as Figure 4 shows, an organization will iterate multiple times through these fields of action. An iterative approach that starts in a limited area of the organization and from there, is scaled to other domains, helps to avoid risks and generates the first tangible results earlier.

References

1. Bauernhansl, T.; ten Hompel, M.; Vogel-Heuser, B. *Industrie 4.0 in Produktion, Automatisierung und Logistik*; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2014; ISBN 978-3-658-04681-1.
2. Pishdad-Bozorgi, P.; Gao, X.; Shelden, D.R. Introduction to cyber-physical systems in the built environment. In *Construction 4.0*; Sawhney, A., Riley, M., Irizarry, J., Riley, M., Eds.; Routledge: London, UK, 2020; pp. 23–41. ISBN 9780429398100.
3. Schwab, K. The Fourth Industrial Revolution. *J. Int. Aff.* **2016**, *72*, 17–22.
4. Veza, I.; Mladineo, M.; Peko, I. *Analysis of the Current State of Croatian Manufacturing Industry with Regard to Industr 4.0*; Croatian Association of Production Engineering: Vodice, Croatia, 2015.
5. Kagermann, W.H.; National Academy of Science; Deutsche Post AG; German Research Center for Artificial Engineering. *Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Final report of the Industrie 4.0*; Working Group, Acatech: Frankfurt, Germany, 2013.
6. Sima, V.; Gheorghe, I.G.; Subić, J.; Nancu, D. Influences of the Industry 4.0 Revolution on the Human Capital Development and Consumer Behavior: A Systematic Review. *Sustainability* **2020**, *12*, 4035. [CrossRef]
7. Anderson, J.; Rainie, L. Artificial Intelligence and the Future of Humans. *Pew Res. Cent.* **2018**, *10*, 1–9.
8. Lefebvre, H.; Legner, C.; Fadler, M. Data democratization: Toward a deeper understanding. In Proceedings of the International Conference on Information Systems (ICIS), Austin, TX, USA, 12 December 2021.
9. Awasthi, P.; George, J.J. A case for Data Democratization. In Proceedings of the AMCIS, Salt Lake City, UT, USA, 10–14 August 2020.
10. Loshin, D. Data Intelligence: Empowering the Citizen Analyst with Democratized Data. *Erwin* **2019**, *1*, 2–3.
11. Kim, H.; Kim, H.R.; Kim, S.; Kim, E.; Kim, S.Y.; Park, H.-Y. Public Attitudes Toward Precision Medicine: A Nationwide Survey on Developing a National Cohort Program for Citizen Participation in the Republic of Korea. *Front. Genet.* **2020**, *11*, 283. [CrossRef]
12. Bellin, E.; Fletcher, D.D.; Geberer, N.; Islam, S.; Srivastava, N. Democratizing information creation from health care data for quality improvement, research, and education-the Montefiore Medical Center Experience. *Acad. Med.* **2010**, *85*, 1362–1368. [CrossRef]
13. Schuh, G.; Anderl, R.; Gausemeier, J.; ten Hompel, M.; Wahlster, W. *Industrie 4.0 Maturity Index: Die Digitale Transformation von Unternehmen Gestalten (Acatech Studie)*; Herbert Utz Verlag: Munich, Germany, 2017.
14. Statista Research Department. Volume of data/information created, captured, copied and consumed worldwide from 2010 to 2025. *Statista* **2022**, *1*. Available online: https://www.statista.com/statistics/871513/worldwide-data-created/ (accessed on 19 May 2022).

15. Chea, J.; Shen Lai, P. Accelerated Digital Transformation: Research, perspective and guidance from data scientists and thought leaders. *SAS* **2022**, *1*, 4–15.
16. Thieullent, A.-L.; Jiang, Z.; Perhirin, V.; Tolido, R.; Jones, S.; Svahn, T.; Clarke, N.; Buvat, J.; Manchanda, N.; Puttur, R. The data-powered enterprise: Why organizations must strengthen their data mastery. *Capgemini* **2020**, *1*, 3–29.
17. Legner, C.; Lefebvre, H. How Communities of Practice Enable Data Democratization Inside the Enterprise. *ResearchGate* **2022**, *2*, 1796.
18. Fahey, S. The democratization of big data. *J. Natl. Sec. L. Poly.* **2014**, *7*, 325–332.
19. Crosby, L.A.; Johnson, S.L. The democratization of data. *Mark. Manag.* **2001**, *10*, 8–15.
20. Yoder, R.T. Digitalization and Data Democratization in Offshore Drilling. In Proceedings of the Offshore Technology Conference, Houston, TX, USA, 6–9 May 2019.
21. McLaughlin, R.; Young, C. *Data Democratization and Spatial Heterogeneity in the Housing Market*; Harvard Joint Center for Housing Studies: Cambridge, MA, USA, 2018.
22. Alexander, J.C.; Joshi, G.P. Smartphone Application-based Medical Devices: Twenty-first Century Data Democratization or Anarchy? *Anesth. Analg.* **2016**, *123*, 1046–1050. [CrossRef]
23. Dong, W.; Kwok-Leung, T.; Qiang, M. Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators. In Proceedings of the 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 19–21 June 2017; IEEE: Piscataway, NJ, USA, 2017.
24. Duan, L.; Da Xu, L. Data Analytics in Industry 4.0: A Survey. *Inf. Syst. Front.* **2021**, *1*, 1–17. [CrossRef] [PubMed]
25. De, D.; Bhattacharyya, S.; Rodrigues, J.J.P.C. *Blockchain Based Internet of Things*; Springer Singapore: Singapore, 2022; ISBN 978-981-16-9259-8.
26. Da Xu, L.; Viriyasitavat, W. Application of Blockchain in Collaborative Internet-of-Things Services. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 1295–1305. [CrossRef]
27. Ramasamy, L.K.; Kadry, S. (Eds.) *Blockchain in the Industrial Internet of Things*; IOP Publishing: Bristol, UK, 2021; ISBN 978-0-7503-3663-5.
28. Rohweder, J.P.; Kasten, G.; Malzahn, D.; Piro, A.; Schmid, J. Informationsqualität—Definitionen, Dimensionen und Begriffe. In *Daten-und Informationsqualität*; Hildebrand, K., Gebauer, M., Hinrichs, H., Mielke, M., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2015; pp. 25–46. ISBN 978-3-658-09213-9.
29. Sprague, R.H. A Framework for the Development of Decision Support Systems. *MIS Q.* **1980**, *4*, 1–26. [CrossRef]
30. Keen, P.G.W.; Morton, M.S.S. *Decision Support Systems: An Organizational Perspective*; Addison-Wesley: Reading, MA, USA, 1978; ISBN 0201036673.
31. Druzdzel, F.; Flynn, R.R. *Decision Support Systems*; Encyclopedia of Library and Information Science: New York, NY, USA, 2002; Volume 2.
32. Spath, D.; Ganschar, O.; Gerlach, S.; Hämmerle, M.; Krause, T.; Schhlund, S. *Produktionsarbeit der Zukunft—Industrie 4.0*; Fraunhofer Verlag: Stuttgart, Germany, 2013.
33. Lanza, G.; Hofmann, C.; Stricker, N.; Biehl, E.; Braun, Y. Auf dem Weg zum digitalen Shopfloor Management: Eine Studie zum Stand der Echtzeitentscheidungsfähigkeit und des Industrie 4.0-Reifegrads. *Karlsr. Inst. Für Technol. (KIT) Karlsr.* **2018**, *1*, 34–36.
34. Yarlagadda, R.T.; Syed, H.H. Data Models In Information Technology. *ResearchGate* **2016**, *1*, 1–8.
35. Schwade, F.; Schubert, P. Social Collaboration Analytics for Enterprise Collaboration Systems: Providing Business Intelligence on Collaboration Activities. In Proceedings of the Conference on System Sciences, Hawaii International Conference on System Sciences, Hilton Waikoloa Village, HI, USA, 4 January 2017.
36. Kucharska, W.; Kowalczyk, R. Trust, Collaborative Culture and Tacit Knowledge Sharing in Project Management—A Relationship Model. In Proceedings of the 13th International Conference on Intellectual Capital, Knowledge Management & Organisational Learning, Ithaca, NY, USA, 2016.
37. Gartner. Gartner Forecasts Worldwide Social Software and Collaboration Market to Grow 17% in 2021. Available online: https://www.gartner.com/en/newsroom/press-releases/2021-03-23-gartner-forecasts-worldwide-social-software-and-collaboration-market-to-grow-17-percent-in-2021 (accessed on 19 May 2022).
38. Melović, B.; Mitrović, S.; Zhuravlev, A.; Braila, N. The role of the concept of LEAN management in modern business. *MATEC Web Conf.* **2016**, *86*, 5029. [CrossRef]
39. Imai, M. *Kaizen: (Ky'zen): The Key to Japan's Competitive Success*, 1st ed.; McGraw-Hill: New York, NY, USA, 1986; ISBN 007554332X.
40. Malik, S.; YeZhuang, T. Execution of Continuous Improvement Practices in Spanish and Pakistani Industry: A Comparative Analysis. In Proceedings of the 2006 IEEE International Conference on Management of Innovation & Technology, Singapore, 21 June 2006; Wilde, O., Ed.; John Wiley: Hoboken, NJ, USA, 2007; pp. 761–765, ISBN 1-4244-0147-X.
41. Otsuka, K.; Jin, K.; Sonobe, T. *Applying the Kaizen in Africa*; Springer International Publishing: Cham, Switzerland, 2018; ISBN 978-3-319-91399-5.
42. Kucharska, W.; Wildowicz-Giegiel, A. Company Culture, Knowledge Sharing and Organizational Performance: The Employee's Perspective. In Proceedings of the 18th European Conference on Knowledge Management, Barcelona, Spain, 7 September 2017.
43. Michael, M. Organisational Culture: Definitions and Trends. *Res. Gate* **2018**, *1*, 1–7.
44. Gamble, J.R. Tacit vs explicit knowledge as antecedents for organizational change. *JOCM* **2020**, *33*, 1123–1141. [CrossRef]

45. YouGov. Data-Driven Companies Are Resilient Companies. Available online: https://public.tableau.com/app/profile/tableau.research/viz/YouGovSurvey/Data-DrivenCompaniesareResilient (accessed on 14 May 2022).

46. Lerch, C.; Jäger, A.; Heimberger, H. Lean 4.0: Smart und schlank produzieren—Potenziale und Effekte einer neuen Management-disziplin: Potenziale und Effekte einer neuen Managementdisziplin. *Fraunhofer ISI* **2021**, *80*, 2–8.

47. Bertagnolli, F. *Lean Management: Einführung und Vertiefung in die japanische Management-Philosophie*; Springer Gabler: Wiesbaden, Germany, 2020; Volume 1, pp. 256–264. ISBN 978-3-658-31239-8.

48. Dillinger, F.; Formann, F.; Reinhart, G. Lean Production und Industrie 4.0 in der Produktion. *ZWF* **2020**, *115*, 738–741. [CrossRef]

49. Harvard Business Review. How CEOs Can Lead a Data-Driven Culture. Available online: https://hbr.org/2020/03/how-ceos-can-lead-a-data-driven-culture (accessed on 17 May 2022).

50. Soulejman Janus, S. *Becoming A Knowledge Sharing Organization: A Handbook for Scaling Up Solutions through Knowledge Capturing and Sharing*; World Bank Publications: Washington, DC, USA, 2016.

51. Brüning, A.; Gluchowski, P.; Kaiser, A. Data Governance—Einordnung, Konzepte und aktuelle Herausforderungen. *Chemnitz Econ. Pap.* **2017**, *15*, 1–8.

52. Ehrlinger, L.; Wöß, W. Automated Data Quality Monitoring. In Proceedings of the ICIQ 2017, Little Rock, AR, USA, 6–7 October 2017.

53. Rti. Data-Centric Middleware: Modular Software Infrastructure to Monitor, Control and Collect Real-Time Equipment Analytics. Rti Whitepaper. Available online: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiFh87Ph_b4AhWdX_EDHcRcDPwQFnoECAMQAQ&url=https%3A%2F%2Fwww.rti.com%2Fhubfs%2Fdocs%2FRTI_Data_Centric_Middleware.pdf&usg=AOvVaw16O9ISxjihOowvXW_snImY (accessed on 17 May 2022).