Data driven management in Industry 4.0: a method to measure data productivity

Giovanni Miragliotta*, Andrea Sianesi*, Elisa Convertini*, Rossella Distante*

**Department of Management Engineering, Politecnico di Milano, Via Lambruschini 4/B; 20156 Milano (Tel: +39 02 2399 278;e-mail: giovanni.miragliotta@polimi.it) *Department of Management Engineering, Politecnico di Milano, Via Lambruschini 4/B; 20156 Milano (e-mail: andrea.sianesi@polimi.it) *Department of Management Engineering, Politecnico di Milano, Via Lambruschini 4/B; 20156 Milano (e-mail: elisa.convertini@polimi.it) *Department of Management Engineering, Politecnico di Milano, Via Lambruschini 4/B; 20156 Milano(e-mail: rossella.distante@mail.polimi.it)*

Abstract: In the early 1900s, together with the birth of mass production, modern managerial approaches were conceived, under the motto "you can't manage what you don't measure". Since then, operations managers throughout the world had been getting used to measure the productivity of materials, machines and workers to control and improve their own businesses. Nowadays, in the Industry 4.0 era, the emphasis is shifting toward data, under the new motto "data is the new oil". Despite many managers pledging allegiance to the principles of data driven decision making, still no comprehensive approach exists to measure how good a company is at exploiting the potential of its own information assets; in other words, no "data productivity" measure exists. In this paper, we present a first method to define and measure data productivity. Relying on a comprehensive literature review, and inspired by the traditional OEE framework, this new method brings some innovative perspectives. First, data productivity is broken into data availability, quality and performance of the decision-making process using those data. Second, it includes both technical and organizational factors, helping companies to evaluate their current level of productivity, and actions to improve it. The model has been tested through three cases studies and it results as effectively implementable. The results obtained from its application reflect the expectations of companies'managers accelerating the cultural shift needed to fully express the potential of Industry 4.0.

Keywords: Data productivity, Performance measurement, data driven decision making, Industry 4.0, Information Management.

1. INTRODUCTION

The industrialization phenomena that characterized different historical centuries, has been strongly influenced by technological progresses called Industrial Revolutions. These revolutions have always drive business to drastic increase in productivity. The first was triggered by the introduction of the steam power, in the middle of the eighteenth century; the second started conventionally in 1870 with the establishment of electricity, chemical products and crude oil, together with the changing concept of mass production; the third, one century later, referred to the effects provided by the increasing usage of electronics and IT in manufacturing industry; finally, in 2011, the term Industrie 4.0 was introduced at the Hannover Messe in Germany.
Industry 4.0 is "a vision of the future of Industry and Manufacturing in which Information Technologies are going to boost competitiveness and efficiency by interconnecting every resource (data, people and machinery) in the Value Chain" (Politecnico di Milano 2017).
The interconnection between the resources is the cardinal element of this revolution and the exchange of data is becoming the new flow to manage with the same relevance of the management of the materials and products one. Indeed, digitalization, intelligence and connection are the three pillars of the Industry 4.0 that poses the data as the principal actor of this new paradigm.
According to the research reported by the Industry 4.0 Observatory of Politecnico di Milano in 2017, companies currently have acquired knowledge and mastery of the Industry 4.0 paradigm. However, this knowledge is still linked to the theoretical concepts on the importance of *data* and there are currently no tools that measure their *productivity*).

In this scenario, companies' attention is more and more moving towards *data* and towards the *information asset* that they own. Indeed, companies are aware of the importance of the data and their potentiality, but currently they have not metrics to measure the efficiency and the *value* that their information asset is giving or can give to them. In addition, before the current data-revolution, the value of their information asset has been not proper valuated, because of the difficulties of set a proper measure and the limited attention of top managers to this asset. This research has the aim to provide a structure measure of *data productivity*, answering to the questions: What's mean that a *data* is

productive? How much a data can produce *value* for a company?
The structure of the paper is the following: in the first part, all the topics on which the research is based are presented; afterwards the definition of data productivity is introduced and its related assumptions; then the procedure to calculate the data productivity measure is presented; finally, there are the main insights of the model application.

2. LITTERATURE REVIEW

2.1 Industry 4.0 and Data

Data is one of the main pillars of Industry 4.0, the 4th generation of manufacturing that uses concepts such as cyber-physical systems, virtual copies of real equipment and processes, and decentralized decision making to create a smart factory or "Factory 4.0".
Within the Industry 4.0 world, *big data* impact can be summarized in six Cs: connection (sensors and networks); cloud (computing and on-demand); cyber- (model and memory); content/Context (meaning and correlation); community (sharing and collaboration); customization (personalization and value). In this new paradigm, a key role is played by the *Industrial Internet of Things* (IoT). IoT allows companies to capture data about process and products more quickly, to have global visibility on the overall supply chain, to work with more efficient and intelligence operations that, thanks to autonomous data collection and analysis, allows on-the-fly decision making (IDC Digital Universe 2014). The effective extraction and use of information embedded in the data have become the next frontier to drive innovation, competitiveness, and growth in manufacturing, as highlighted by McKinsey in a series of studies see McKinsey, 2011, 2015.

2.2 Data versus Information

"Data can be defined as a symbol, sign, or raw fact", see Mingers, 2006. Usually in literature the concept of data is associated or put in contrast with the information one. So far, two different definition of *information* are present:
1. information is "data that has been processed in some way to make it useful [..] information can be objectively define relatively to a particular task or decision", see Mingers, 1996. This definition of information implies that the concept of data is objective.
2. "Information equals data plus meaning", see Checkland et al., 2000. This definition implies that information is subjective, from a set of data different information can be created. Information is subjective because depends on the receiver and on the context in which the message is conveyed see Jumarie, 1990. Information cannot be viewed as an independent entity because has attributes and reflects the intention of the sender and the receiver, see Oppenheim et al., 2003. Different observers may generate *different information* from the *same data* given their differing values, beliefs, and expectations, see Lewis, 1993.

2.3 Data and Information Asset

During the late 1990s theories of *knowledge asset* and their contribution to organisational wealth become popular together with the definition of *information as an asset*. The role of information as an asset was introduced in 1994 by the Hawley committee, which define information asset as: "data that is or should be documented and that has value or potential value". This concept, in addition, treated information like traditional assets such as plant and machinery, see S.H Black et al., 1982. *Information asset* is defined also as "the ability to provide data and information to users with the appropriate levels of *accuracy*, *timeliness*, *reliability*, security, confidentiality, *connectivity*, access and the ability to tailor these in response to changing business needs and directions", see Mithas et al., 2011. "V*alue* should be assigned *not only* to *data* but also to the *system* allowing for its *exploitation*", see Ahituv, 1989. Because Information Asset is an intangible asset it is difficult to measure, see Evans et al., 2015.
Data can be in the form of structured, semi-structured, or unstructured data. There are several data types such as Master data, that provides the most business relevant information about a product, a supplier, a customer, etc. or Transactional data, that describes the event that happens in a moment referring to one or more master data element. These two types of data together with Tacit Knowledge can be consider the Information Asset of a company. Tacit Knowledge is hidden in human brains and refers to communication among people who share their knowledge and observations, in both formal and informal ways.

2.4 Measuring productivity in a manufacturing company

Productivity is commonly defined as a ratio of a volume measure of *output* to a volume measure of *input* use. While there is no disagreement on this general notion, productivity literature and its various applications proposes different measure of productivity, see OECD Manual, 2001.
Looking at the Efficiency aspect of productivity, a production process is operating in full efficiency, in an engineering sense, if it achieved the maximum amount of output that is physically achievable with current technology, and given a fixed amount of inputs, see Diewert E. W. et al., 1999.
Manufacturing companies are often interested in measure productivity in terms of *efficiency*. The *Overall Equipment Effectiveness* (OEE) measure gives a comprehensive response to this need, underling different *losses* of the manufacturing process under analysis. Measuring OEE is, indeed, a manufacturing best practice that provides insights on where it is necessary to act to improve the productivity. The most significant way to calculate the OEE index is based on *three factors*: *Availability*, *Performance*, and *Quality*.

2.5 Data value and Productivity

In literature there are *not researches* that provides to managers and academics a metrics or a model that allow them to measure the *Data productivity*. The main insights coming from the literature on the theme are:
1) some companies are aware of the *potentially* of the *data* they own, but are not able to exploit them see Ladley, 2010;

2) *data-driven* companies *better* perform in terms of *productivity* than other companies, see McAfee et al., 2012;
3) companies have, very often, other priorities than improving the *information asset*, simply because it is *not* an *accountant voice*. Taylor refers to the potential value of information asset, a concept that does not exist in traditional accounting, so that will always result in no value being attributed to information for financial reporting purposes, see Taylor, 1986.

Early studies on the value of information are considered together by Ahituv in 1989. He considered 4 different way to assess this value (quantitative, normative, realistic, perceived) concluding that each method presents different limits and none of them give a unique indicator of the data productivity of a company, see Aihtuv, 1989. Strassmann (Strassmann 2004) defined a quantitative measure of the Information productivity: the Information Productivity Index (IPI), see Strassmann, 2004. However, this measure is based on financial indicators and accounting voices that do not permit to measure the productivity of a single data element, because of their aggregated nature.

3. METHODOLOGY

The model was built up by joining personal researches together with an *empirical research* phases from three interviews with companies' managers, as case studies. Face-to-face interviews were useful not to force interviewees to provide punctual responses, expressing their ideas freely. This provides to the research original points of view that were very useful to design a comprehensive model. The *companies* selected belonging to manufacturing industry and have already implemented projects on Industry 4.0. After the first phase of model design, the issue of *measurability* was approached. In particular, an analysis of different possibility to measure each characteristics of the model has been fundamental for its correct application. The validation of the model, in terms of its measurability and its capability to correctly measure the Information Asset productivity, comes from the successful application of the model to the three companies object of the first empirical phase.

4. DATA PRODUCTIVITY

4.1 Data productivity and closed themes

To correctly address the research questions, an essential preliminary step is the definition of the borders of this research analysis (fig.1).

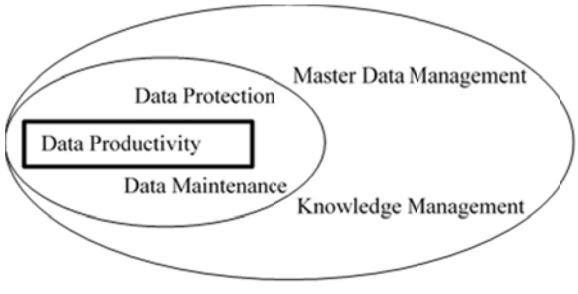

Fig. 1. Data productivity's borders of analysis representation

Master data management (MDM) incorporates business applications, Information Management methods, and data management tools to implement the policies, procedures, and infrastructures that support the capture, integration, and subsequent shared use accurate, timely, consistent, and complete master data", see Loshin, 2009.

Data, properly managed and monitored, in the master data system allow companies to rely on a unified and coherent data asset for all their applications, thanks to consistent and high-quality information, see Loshin, 2009.

The distinction among the concepts of knowledge, information, and data has often been addressed by researchers and academics. As previously presented, a commonly view considers: data as raw numbers and facts; information as processed data; knowledge as authenticated information, see Dretske, 1981, Machlup et al., 1983, Vance, 1997. Indeed, knowledge is defined as information possessed in the mind of individuals: it is personalized information related to facts, procedures, concepts, interpretations, ideas, observations, and judgments.

Knowledge management refers to identifying and leveraging the collective knowledge in an organization to help the organization compete, see Von Krogh, 1998. Knowledge management is purported to increase innovativeness and responsiveness, see Hackbarth, 1998, Alavi et al., 2001.

Data protection theme is present in literature and in companies since the origin of the data itself. Indeed, since data present in companies are often sensitive, their content should be properly protected. Technical security of data regards whether and how the data is secured against possible incidents of different nature: components failure, software bugs, hacking, fraud or theft.

Data maintenance theme is related to physical storage. In particular, software solutions and database can become faster and more performant if they are subjected to the proper maintenance. Indeed, to make use of the technical innovations, systems should be up to date and be maintained properly, see Spruit et al., 2015.

4.2 Data productivity definition

The *lack* of a *measurement* tool by which companies can understand their current level in the management of *Information asset* and of *data*, heart of the Industry 4.0, is a gap present in the literature. To correctly address this gap, a structured research was essential to define the boundaries of the theme under analysis and to define the *three* main *pillars* of the "*Data Productivity model*" developed: *data, productivity, decision-making process.*

4.3 Data productivity and model's assumptions

The aim of the model is to measure the productivity of data and Information Asset related to a specific process selected for the analysis. The *data* pillar it is presented as the *primary element* of the model.

Looking at the object of the analysis, this research does a step aside from the research work conducted since now, looking back at the data per se, not transformed into information nor into knowledge.

Data are now present in factories in a huge quantity and it is possible to combine, interpret them in several different ways for different purposes. For this, the entity of analysis is not a peculiar information, a peculiar transformation of the data, but it is the data itself.

In the model proposed the focus it is on a *specific decision-making process*. It is possible, indeed to notice that focusing on a specific process allows the identification of the causes of poor performance of the Information Asset in that specific context. In other words, a metrics on the overall performance of the *Information Asset productivity* would drive managers not to understand easily the result obtained and which type of actions are needed to be taken and the related priorities to correctly improve.

Therefore, after the selection of a process to analyse, all the *data of interest* for the process are identified and at the same time also the tools that help the decision maker to take the decision are selected. For this, to assess the Information Asset Productivity of the company it is necessary to apply the model to more than one decision-making process, looking at different areas of applications.

The starting point for the development of the *Data Productivity model*, is the *parallelism* with the *OEE efficiency metrics* in manufacturing. This efficiency definition of productivity, express by the concept of output/input, can be applied to the productive field but also to the digital one, that considers data the primary feed of all processes. Therefore, the main question that guided the model building was: how the Information Asset of a company can be designed as a productive resource highlighting the "*losses*" that occur during the *decision-making process*.

The OEE metrics, is often computed in relation to a specific machinery/process/produced product. In this way, indeed, it is possible to analyse the causes of low efficiency in details, on the contrary this is almost impossible if the overall plant is considered. Looking to a specific element allows to identify also the smaller losses and their causes that are difficult to capture in an aggregate analysis.

The structure followed by this research is a pure comparison between OEE factors and the Information Asset characteristics coming from the literature, by adopting the path of *normative model*. Indeed, taking the OEE factors (by OEE industry standard) as starting point, the model was devised assuming that also the Information Asset characteristics (the one related to the decision-making process) follows a scheme similar to the OEE metrics one. That means, the model was designed by applying the same logic of the OEE metrics based on the three main factors: availability, quality and performance and a series of Information Asset characteristics coming from the literature.

Like the OEE, the *Data Productivity Index* is composed by *Data Availability*, *Data quality*, *DSS performance,* as shown in the following formula (1):

$$\boldsymbol{Data\ Productivity\ Index = Data\ Availability * Data\ Quality * DSS\ Performance} \quad \textbf{(1)}$$

The *Data Availability* factor has the aim to identify all the data at the disposal of the decision maker and it is composed by the "*interesting data*" and the "*generated data*". The interesting data are the ones that are of interest for the decision-making process, not considering if they are or not actually available. Of these data, a portion is selected based on the ones actual available to the decision maker from external or internal company's *sources* of data. The cause of this first loss of data is the absence of appropriated IT and storage infrastructure (e.g. sensors). This is often related to poor management, that fails in a proper ideation, design and implementation of these solutions.

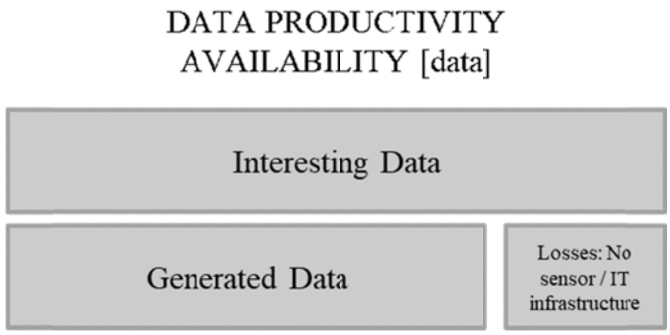

Fig. 2. Data productivity's Availability factor representation

The *Data Quality* factor selects the data that are "*timely*", "*completed and correct*". Form the *generated data*, only the ones that are characterized by a correct *frequency of collection* are considered. In other worlds, the ones that respond as rapidly as required by the user or necessitated by the process, allowing the decision maker to work with reliable data. The data that are not *timely* because of a too slow data collection are reported as losses. From the *timely data*, the *complete and correct* ones refer to those that do not include wrong elements and are composed by all the relevant attributes. Each data can be evaluated in terms of percentage of completeness and correctness.

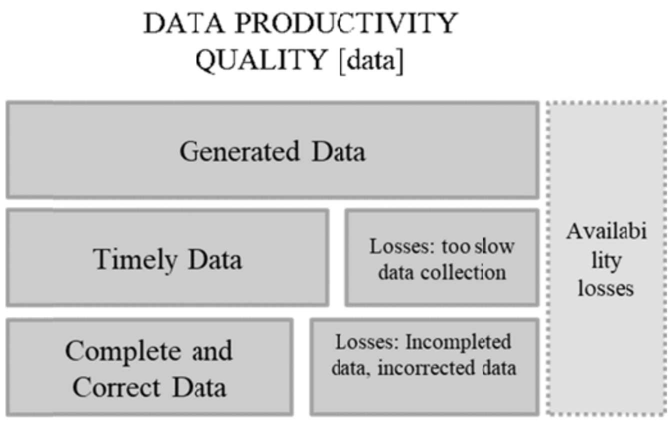

Fig. 3. Data productivity's Quality factor representation

The *DSS Performance* factor refers to the measurement of how the data are managed by the decision support system and the decision-maker in the process. In particular, "*valorized*" data refers to the set of the generated data that are managed by *algorithms*. In this case the losses are related to the weakness of the algorithms applied to the data. Therefore, after the identification of the data that are object of algorithms, an evaluation of their performance determines the percentage of loss because of a not proper exploitation of the data. The "*On time*" data characteristic is related to the data that arrive in a *due time* useful for the decision-making process because of performing and fast algorithms. The "*integrated*" data are those supported by an infrastructure that collects data coming from *different sources* in a *single platform* in order to ease their *consultancy for the* decision maker. The losses are identified by the data that are not integrated *with* each other. The data presented in the decision-making process through indicators are called "*aggregated*" data. The losses in this case are related to the difficulties in using data without looking at *summary indicators* especially because of the usually restricted time availability to make the decision or the presence of several data to look at. Then the "*used*" data are those *used*, at the end, *by the decision maker*. These further distinction is based on the *bounded rationality* concept that indicates the limited rationality of individuals in taking decisions, highlighting their cognitive limitations combined by the time availability. The data losses for this reason are more numerous in case of a no structured decision-making process or a not supporting software or tools for taking the decision.

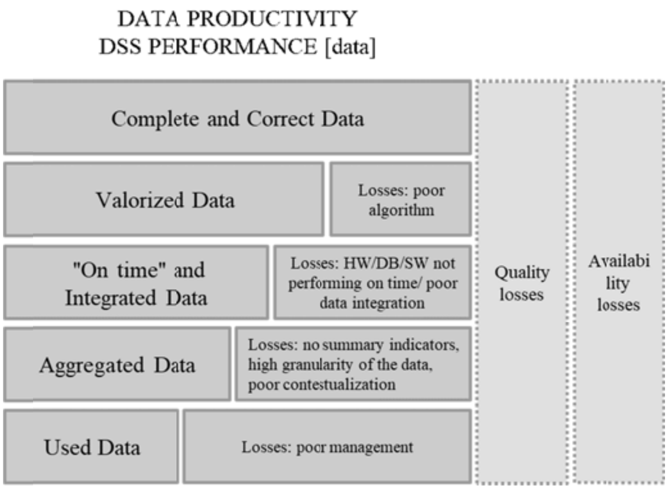

Fig. 4. Data productivity's Performance factor representation

In addition, the model is contextualized in the *Supply Chain* of the company, because the entity of the *turbulence* in the *decision-making process* is concordance with the relevance of data and Information asset value. This implies that the result obtaining by the model application as to be contextualized by the company looking at the process analysed. There are different models in literature that provide indications on how to assess the risk of a company. The *Supply Chain risk assessment* considered for this dissertation is based on the assignment of a grade based on the level of impact and probability of occurrence of each risk driver impacting in the Supply Chain. The *risk drivers* proposed are adapted from Chopra and Sodhi and are divided in 8 macro areas: disruption, delay, IT systems, forecasting & communication, intellectual propriety (IP), legal and economics, procurement, inventory, see Chopra et al., 2004.

.

A matrix that combine the risk drivers with the main decision-making processes of a company allows to obtain a measure of the risk level for each process. This consequently implies the possibility to evaluate the relevance of obtaining a good result by the Data Productivity index.

The results obtained by the model applications need to be contextualized looking at the analysed company, the selected decision-making process, the overall *Data Productivity Index* result and its specific factors (Data Availability, Data Quality and DSS Performance). To clearly present these insights the model developed has been considered together with the *structure* of the *decision-making process* from the *model* "*Thinking first*" presented by Simon. This model considers the decision-making process composed by *five phases*: *intelligence*, *design*, *choice*, implementation, control and review. Only the first three are the ones considered for the model application, because the scope of the model is to measure the Data and Information Asset productivity to assess their level of goodness for taking a decision. The three factors of the *Data Productivity Index* have an impact on all of these three phases, but with different impact. The *intelligence* phase is characterized by a more predominant importance of the *Data Availability* and the *Data Quality*; the *design* one considers as the most relevant the "*valorized*" data and the "*On time*" and "*integrated*" data characteristics; the *choice* one, instead, needs "*aggregated*" and the "*used*" data characteristics.

5. CONCLUSIONS

Because of the *novelty* of the *model* proposed, it is difficult to establish the goodness of the quantitative result obtained by the model application to the *three case studies*, that range from 40% to 60%, but they are in line with the expectations of managers. In addition, the result obtained by an OEE best in class company is about 85%, therefore the numbers obtained reflect: the novelty of the measure and the needed path that companies must follow to improve these results; the *organizational*, *technological* and *cultural changes* that companies must embrace to better manage data, Information Asset, and decision-making process. The *aim* of the model is to assess the *AS-IS situation* of the *Information Asset* of the company related to a specific process, for this a comparison between the same metrics after some projects that has the aim to improve the information asset management is also value added. Because of the novelty of the model and the fact that the presented applications are the first ones, it is possible to improve the model working on the modality of applications, on the involvement of the actors, on the extension of the model application also to other processes. In addition, it is possible to refine the characteristics investigated by the model with other experts' opinions or by insights coming from more applications. Overall, it is possible to conclude that the model posits a first approach to measure Information Asset productivity, from its applications relevant insights emerged that helps companies to assess their AS-IS performance on these elements and to identify pattern of *improvements*.

6. REFERENCES

Ahituv N. (1989). Assessing the value of information: Problems and Approaches. *Proceedings of the International Conference on Information Systems (ICIS) 1989*, 45.

Alavi M., Leidner D.E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly,* 25 (1), 107-136.

Black S.H, Marchand D. (1982). Assessing the value of information in organisations: a challenge for the 1980s. *The Information Society*, 1, 191-225.

Checkland P, Scholes J. (1990). *Soft Systems Methodology in Action*. John Wiley and Sons, New York.

Chopra, S., Sodhi M. (2004). Managing risk to avoid supply-chain breakdown. *MIT Sloan Management Review,* 46(1).

Diewert Erwin W., Lawrence D. (1999). Measuring New Zealand's Productivity, *Treasury Working Paper*, 99(5).

Dretske F. I. (1981). *Knowledge and the Flow of Information.* MIT Press, Cambridge Mass.

Evans N., Price J. (2015). Information Asset Management capability: The Role of the CIO. *Twenty-first Americas Conference on Information Systems, 2016.*

Hackbarth G. (1998). The Impact of Organizational Memory on IT Systems. *Proceedings of the Fourth Americas conference on Information Systems*, 588-590.

Infor publication (2015). *Big Data in manufacturing: A compass for growth.* Infor, New York.

Jumarie, Guy. (1990). *Relative Information: Theories and Applications.* Springer-Verlag Berlin Heidelberg, Berlin.

Ladley J. (2010). *Making Enterprise Information Management (EIM) Work for Business*. Elsevier, Burlington.

Lewis, P. J. (1993). Linking soft systems methodology with data-focused information systems development. Information Systems Journal, July, 169-186.

Loshin D. (2009). *Master Data Management*. Elsevier, Burlington.

Machlup F., Mansfield U. (1983). *The Study of Information: Interdisciplinary Messages.* Wiley-Interscience, New York.

McAfee A., Brynjolfsson E. (2012). Big data: The Management Revolution. *Harvard Business Review*, 90 (10), 1-9.

McKinsey Global Institute (2015). *The Internet of Things: Mapping the Value Beyond the Hype.* McKinsey & Company, New York.

McKinsey Global Institute (2011). *Big data: The next frontier for innovation, competition, and productivity.* McKinsey & Company, New York.

Mingers J. (2006). *Realizing Systems Thinking: Knowledge and Action in Management Science*. Springer, Berlin.

Mingers J. (1996). An Evaluation of Theories of Information with Regard to the Semantic and Pragmatic Aspects of Information Systems. *Systemic Practice,* 9(3), 187-209.

Mithas, Sunil, Ramasubbu, Narayan, Sambamurthy V. (2011). How Information Management Capability Influences Firm Performance. *MIS Quarterly,* 35(1), 237-256.

OECD Organisation for Economic Co-Operation and Development (2001). *Measuring Productivity.* OECD Publications, Paris.

Oppenheim C., Stenson J., Wilson R.M.S. (2003). Studies on information as an asset I: definitions. *Journal of Information Science,* 29 (3), 159–166.

Politecnico di Milano (2017). *Industria 4.0: la grande occasione per l'Italia.* Observatory Industry 4.0, Milano.

PwC (2016). *PwC's 2016 Global Industry 4.0 Survey: Building the Digital Enterprise.* PwC, London.

Spruit M., Pietzka K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior,* 51, 1068-1076.

Strassmann P.A. (2004). *Defining and Measuring Information Productivity*. The Information Economics Press, New Canaan.

Taylor R.S. (1986). *Value-added Processes in Information Systems*. Praeger, Westport.

Vance D. M. (1997). Information, knowledge and wisdom: The epistemic hierarchy and computer-based information systems. *Proceedings of the 1997 America's Conference on Information Systems.*

Von Krogh G. (1998). Care in Knowledge Creation. *California Management Review*, 40(3), 133-153.