Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 51 (2020) 1125–1131

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

30th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2021)
15-18 June 2021, Athens, Greece.

Availability of Manufacturing data resources in Digital Twin

Sara Moghadaszadeh Bazaz[a*], Mika Lohtander[a], Juha Varis[a]

[a] Lappeenranta-Lahti University of Technology, P.O. Box 20, FI-53851 Lappeenranta, Finland

* Corresponding author. Tel.: +358-41-7513311. *E-mail address:* sara.moghadaszadeh.bazaz@lut.fi

Abstract

Data and information are the basis of the integrated manufacturing environment. The latest, synonym of the integrated manufacturing environment is a digital twin. The concept of a digital twin has grown profoundly in recent years. Several models and frameworks are suggested to create a digital twin in manufacturing. Various information and data flows are involved in any manufacturing case. Diagnosis and gather required data and information is the first and basis of creating digital twin. Data ownership is a significant issue because of the large amount of isolated data islands and due to different data ownership relationships of the whole production chain. This study focuses on the ownership paradigm in manufacturing and digital twin. The reliable data resources in manufacturing the metal-based products are recognized. Available and unavailable data in the turning process are studied and classified in general form to cover every condition as a case study.

Keywords: Data ownership; digital twin; manufacturing data and information; turning process

1. Introduction

Fourth Industrial Revolution is the transition from an electro-mechanical environment to an integrated and information-centric environment. Big Data, Internet of things, artificial intelligence, augmented and virtual reality, and digital twin can be called as components of the Industry 4.0 [1]. Data and information are the common basis of these technologies. Understanding the meaning of data, information, and knowledge and how to distinguish them would help to perceive the position of data in the industry. There are many definitions of data in articles and books. Hicks defined data as a representation of facts in a form that is suitable for interpretation, transferring, or processing. Data are the fact from the recording of measurable events [2]. Professor Biagetti defined datum as every unit which is able to raise human knowledge and it is possible to record it. Information is the processed data that can be transmitted. Information is the meaningful output from a computer program [3]. Professor Barreto defined information as a set of significant sings that has the ability to create knowledge. Wersig and Neveling expressed the nature of the information phenomenon as a communication process between the sender and the recipient of the message. Therefore, the expression of the information concentrates on the origin and the final stage of the communication process [4]. “Knowledge is structured and organized information that has developed inside of a cognitive system or is part of the cognitive heritage of an individual” [5].

Data and information are the basis of the integrated manufacturing environment. The properties, characteristic, and behavior of the physical product is built with conversion data to information and knowledge in virtual space. Providing the initial required data is the first and important step in making a digitalized process. In the manufacturing industry, there are numerous data related to a product such as material data, design data, and production data [6]. Some of these data and information are available in general books, articles, or handbooks but some particular data belongs to the specific owners. creating a comprehensive digitalized environment that is able to mimic the real product requires a vast amount of data

2351-9789 © 2020 The Authors. Published by Elsevier Ltd.

10.1016/j.promfg.2020.10.158

and information related to the demand process and desired problem.

In manufacturing, data can be raw values, parameters, and characters, which are called unstructured facts. Information consists of data that is structured with features with communicative ability. Information, thus, comprises items such as degrees of freedom, interfaces between parts and assemblies, and it shows the relationship between components. Knowledge is information with accessible meaning. Some examples of knowledge are functionality, geometry and non-geometric information, and product variant [7].

The concept of a comprehensive digitalized environment which imitates real production is called a digital twin. Digital twin introduced by Dr. Michael Grieves in 2003. In terms of manufacturing, "the digital twin consists of a virtual representation of a production system that is able to run on different simulation disciplines that is characterized by the synchronization between the virtual and real system, thanks to sensed data and connected smart devices, mathematical models and real time data elaboration. The topical role within Industry 4.0 manufacturing systems is to exploit these features to forecast and optimize the behavior of the production system at each life cycle phase in real time." [8].

1.1. Manufacturing digital twin

Dr. Grieves introduced digital twin as a virtual representation of what has been produced which contains 3 main parts: real space, virtual space, and the connections of data and information that ties the virtual and real products together [9,10]. In the digital twin, the data and information play an important role in communication from physical space to virtual space to use in analyzing, predicting, or optimizing then the results are fed back to the physical space [11].

During recent years, the concept of digital twin expanded rapidly. Thus, many definitions are introduced by researchers. In reference 12, the authors express the digital twin in manufacturing by integrating real-time information communication and artificial intelligence. The digital twin model integrates all information, multi-physics, and possible environments that influence the final product to imitate the behavior of the physical space in order to predict, make decisions and optimize a manufacturing process or production line.

The creation of a digital twin for manufacturing differs from digital twin in other fields due to the wide variety of data and information involved, which necessitates the making of a generalizable model. In view of a large number of different data owners, it is infeasible to access all required data to create an accurate digital twin, and the user is therefore confronted with the issue of different digital twin levels with respect to available data and information.

In this survey, different data sources of a metal-based product and their corresponding owners are explained. Most of the possible data owners for each data category is studied according to the manufacturing expert's information. Moreover, relevant available and unavailable data are introduced.

The paper is structured as follows: Section 2 deals with the research problem and question; Section 3 presents the using data in digital twin and data ownership paradigm; Section 4 identifies the various data resources and their place in the digital twin framework; finally, Section 5 and 6 propose some discussing and concluding remarks.

2. Research problem and question

This paper investigates the various resources of data and information in manufacturing metal-based products. The research question is identifying the possible reliable data resources to create a comprehensive digital twin of a manufacturing process or a production line. As a sub-question, the researchers detected the available and unavailable e required data to create a comprehensive digital twin of a manufacturing process or a production line. Depending on customer demands and accessibility to data and information, the level of smartness of the digital twin may vary. Accuracy, reliability, and validity are the factors that determine the level. As primary data are used for training in artificial intelligence methods, the absence of one or more parameters may impact other relevant parameters.

This paper is following the previous paper of the authors about proposing the comprehensive digital twin framework in manufacturing [12]. Figure 1 shows the layout of the proposed digital twin framework which will be explained in the method section. This paper is discussing the cyber-physical data store layer. This layer contains data from different resources. In this paper, different resources of required data are discovering.

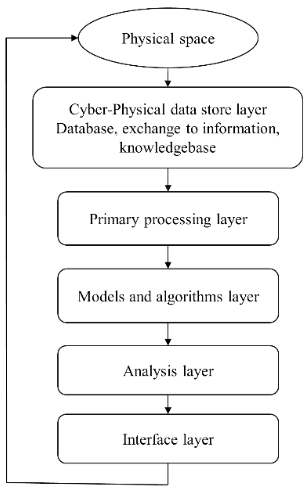

Fig. 1. Digital twin general framework for manufacturing based on [12].

3. Method

The study has been conducted based on industry experience in understanding the production environment. The ownership types are studied to attribute the data resources in manufacturing.

3.1. Role of data in the digital twin framework

In the methodology of this survey, the authors focus on classifying the type of data used in their earlier work about a framework for comprehensive digital twin in manufacturing [12]. The paper expressed the position of using the data in the digital twin framework, type of data, and the different types of data owners. The advantage of this framework is that if the user can use sufficient data and information as an input, the framework can be used to solve various problems. Therefore, access to sufficient data is the barrier to making an extensive digital twin. Figure 1 shows the 5 layers framework of the digital twin in manufacturing. This framework consists of 5 layers: cyber-physical data store layer, primary processing layer, models and algorithm layer, analysis layer, and interface layer.

The paper pays attention to the cyber-physical data store layer. The provision of an accurate description of the physical object and its data is a basic and crucial phase in the creation of a virtual model. A wide variety of data resources are involved in modern manufacturing, which complicates data treatment. Data resources are divided into five groups: equipment data, material and product data, environmental data, management data, and internet data. Equipment data consist of real-time data, information, and knowledge of equipment properties, performance, and operating conditions. Material and product data include performance, inventory, and context of use. Environmental data comprise properties such as temperature, humidity, and air quality. Workers are categorized as environmental data. Management data consist of data collected from manufacturing information systems and computer-aided systems. Internet data include user data collected from e-commerce and social networking platforms and public data from open websites [6, 12].

3.2. Data ownership

The collection of data and construction of the database is one of the foundation layers of the digital twin model for manufacturing, and it is necessary to collect data from reliable resources both in the public domain and from various service providers. Consequently, addressing the question of data ownership becomes an issue.

Data ownership implies the ability to create, access, modify, package, and gain profit from selling data [13]. In the digital twin for manufacturing model, it is the responsibility of the owner of the data to collect the correct data for required tasks, so any kind of change or improvement in the data is under "the duties of the data owner." The data owner has a responsibility to provide, package, and format the data for the client. Additionally, there are business rules for data processing operations and the owner is responsible for managing and adhering to the rules, for example, by following data standards as regards data format. When a data owner has a set of data from different suppliers, the owner should be responsible for supplier management. This management includes negotiating arrangements with each provider and making data delivery agreements and data quality criteria with each data supplier.

Ownership paradigms can be helpful to define importance in the ownership problem. These paradigms are creator, consumer, compiler, enterprise, funding organization, decoder, packager, reader, the subject, purchaser, and everyone as owner [14].

Creator of data is the common ownership. A designer creates the design data, or a manufacturer creates the manufacturing specific data. A consumer of data for example using data in simulation should have ownership right. The consumer paradigm represents consumers need to have an ownership to use data in processing because the consumer gains the value from the data. Collecting data and arrange information defines compiler ownership. In the case of the digital twin, for example, gathering database, converting to the information, and extracting knowledge are owned by compiler. In the big organization, there is a policy that the company owns subtend input and output data. In funding organization, there are two owners: first, the company that creates the data, second, the company who paid to create the data. For special data which are locked, who is permitted to decode is the owner. Packager paradigm creates the format information for specific use. Reader adds information to an information resource. The subject as owner implies on privacy and copyright right. It is issue when another party claims the ownership of the same data. Purchaser paradigm expresses a buyer claims the ownership of data. Everyone as owner shows the global data ownership that data should be available for knowledge to improve global knowledge for everyone such as researchers, engineers, or experts [14,15].

Data ownership is a complicated issue because of the considerable resources expended on data creation, data management and data usage. The problem is a data control issue about the flow and value of information. In the context of this study, the issue of value is particularly pertinent when the data has more than one owner. In such cases, therefore, the degree of ownership needs to be determined to ascertain the value gained from usage of the data for each creator. Information privacy, business rules and Information Island determine how it is possible to use data or information, how information moves from one source to another, and responsibility for control of the information and information flow. In a factory, Information privacy contains manufacturing properties and design parameters of a key product, business rules can be the environmental properties and management privacy.

4. Results

Studying the required data to form a digital twin and possible resources to obtain the data incorporate with discussion about available and unavailable data and the problem of multi ownerships is classified in the result section.

Table 1a. Ownership information of data in general machine job or metal based manufacturing process - Ownership levels structure.

Ownership levels structure of production manufacturing data				
Data	**Ownership levels**	**Description**	**Data ownership**	**Place in Digital twin layer**
Technical Design	Ownership of product rights		Creator Purchaser	Database-Cyber physical data store
	Ownership of characteristics of parts			Information conversion-primary processing layer
	Manufacturing knowledge of parts			Models and algorithm
				Analyzing layer
Raw material for production	Product owner knowledge how material behaves in its purpose	Capabilities and costumer demand	Creator Purchaser	Database & knowledgebase-Cyber physical data store
	How material behaves under certain manufacturing	How material behave under certain manufacturing circumstance	Creator Purchaser	
		Save material in production		
Manufacturing Machines and controller	Machine technical data and its equipment	Third party production parameters	Creator Purchaser Funding organization	Database & knowledgebase-Cyber physical data store
		Companies own production parameters		
		Third party manufacturing company		LISA- primary processing layer
	Rights to use firmware	Right to use internal data warehouses	Creator Purchaser Reader	
		Add your own data to data warehouse		Models and algorithm
		Right to create own data		
Tools	Tool provider	Mechanical tool properties owner	Creator Purchaser	Database & knowledgebase-Cyber physical data store
				Information conversion-primary processing layer
				Models and algorithm
Operation	Machine manufacturer		Creator Purchaser	Database - Cyber physical data store
Operation	Operator of company using the machine		Packager	Information conversion-primary processing layer Analyzing layer
Software	Firmware programmer (mostly inside the machine producer company sometimes outside)		Creator Purchaser Reader	Database & knowledgebase-Cyber physical data store
	Third party programming (outside, post processor, auxiliary			LISA- primary processing layer
	Post processor			Analyzing layer
Maintaining	Producer company	Maintenance handbook	Funding organization	Database- Cyber physical data layer
	Maintenance company	Overall maintenance		
	Service provider	Maintenance, system process, collect data for analyzing		
Method of processing	Machine producer provider		Funding organization	Database & knowledgebase-Cyber physical data store
	Raw material provider			Information conversion-primary processing layer
	Service provider			
Costumer demands	Costumer	Order an ideal product (knows 100% demand)	Proposal, offer, agreement	Database & knowledgebase-Cyber physical data store
	Other involved companies	Depending on the type of contract it can be many layers of data owners	Contract: information (research basic knowledge)	User Interface
Blackmailing information production	All data include Experience information, measurement data, accidently discovered data,	In very high cost	All	

Table 1b. Ownership information of data in general machine job or metal based manufacturing process - data statues.

	Data statues	
Data	**Accessible or available data**	**Inaccessible or unavailable data**
Technical Design	Information from patents Method, materials, geometry,	Newest Patent pending, information inside intranet
Raw material for production	Ingredients, how to produce, Basic properties of materials	Combination materials properties
Manufacturing Machines and controller	Industrial rights of the machine (design, main structure, patent's features, industrial properties, adjustments, equipment) Collected data from machine	Company specific information.(hidden information) Experience based information (important in all sections)
Tools	Parameters, functionality	Material, characteristics of tool structure
Operation	instructions on how to perform one step at a time, knowledge of operating	All the steps at a time
Software	Guideline to use, abilities	All adjustment and capabilities
Maintaining	All actions, parameters, adjustment, failure modes	Failure data
Method of processing	Mechanical and chemical actions, material related production parameters	Some of raw material properties, chemical properties
Costumer demands	Final information in the contract are available about the aim of product	Some data are hidden in intermediaries
Blackmailing information production	Needs more research	Needs more research

Since collecting appropriate data is a vital part of the proposed digital twin framework for manufacturing process, access to the comprehensive database is an issue. As figure 2 demonstrates there are several sources for required data, information and knowledge for a product.

For instance, there are two owners for technical design. Because data in this field can be from costumer, who order the product for specific use and data from design company that contains detailed design data based on the possible production technology.

Table 1a and 1b express the data, information and knowledge involved in making a product that need to be collected in the database and knowledge base in order to create the digital twin of the product. It should be noted that some data have more than one data owner. Generally, some data and information are available at different levels of the digital twin framework and different amounts of data are available at different levels. For example, it can be seen that technical design information, details of methods for a specific process, material properties, tool information, manufacturing machine data, operating data, software data, data control and customer needs are involved data sources with different owners.

In the technical design, for instance, some information is available about the method, materials and geometry, as is data about the composition of the material, how it was produced, and its basic characteristics. However, information about the effects of combinations of the properties of materials is unavailable. In addition, parameters and functionality of tools are available at the technical design stage but information about tool materials and the characteristics of the tool structure are unavailable. Knowledge of the tool material and tool characteristics is essential for choosing appropriate methods and tools. In the proposed digital twin framework, artificial intelligence is adopted, which can overcome some of the deficiencies, but sufficient data must be available for training. Furthermore, as better training data sets can lead to better outcomes, it is important to consider ways to obtain unavailable data and the possible extra costs involved.

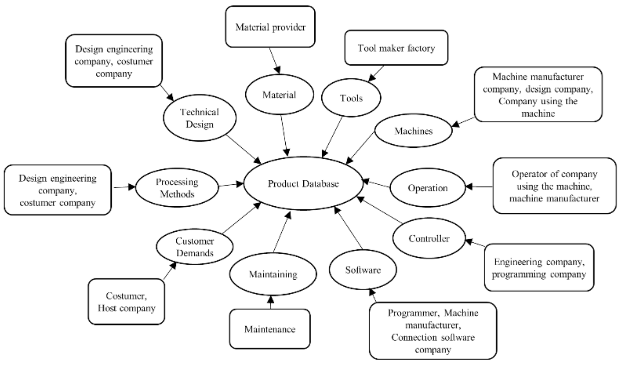

Fig. 2. Different data owners of a product details

4.1. Ownership point of view

Table 1a and 1b describe ownership information for data in a metal manufacturing product. The product technical design contains different types of data and, consequently, there are several owners and types of ownership. Product technical design data are the most important data which can be used in four layers of digital twin: cyber physical data store, information conversion- primary processing layer, models and algorithm, analyzing layer. In many cases, a number of different companies are involved in designing a product. The lead company owns the idea and production, e.g. a new or updated machine (product), and the other companies engage in design of some parts. The data that the companies own contain existing knowledge and general technical features, which are public data, information and knowledge. However, the technical solution or structure developed by the company belongs to the company as a part of Intellectual Property Rights (IPR). Overall, the data ownership of technical design can be classified at three levels. At the first level, ownership of product rights that belong to all involved companies depends on the roles of the respective companies in the design. The second level is ownership of technical characteristics of the parts. Manufacturing knowledge of parts is the last level and contains material properties, mechanical behavior, software usage rights, and understanding of how to operate the machines used. The raw material needs to be used in cyber-physical data store layer to convert to information or knowledge. The raw material for the products contains general knowledge of the properties of the material, which is in public ownership. The private ownership divides into two levels: A production company owns knowledge of how the material behaves, its special purposes, and how the material behaves under certain manufacturing circumstances, which can lead to material and cost savings.

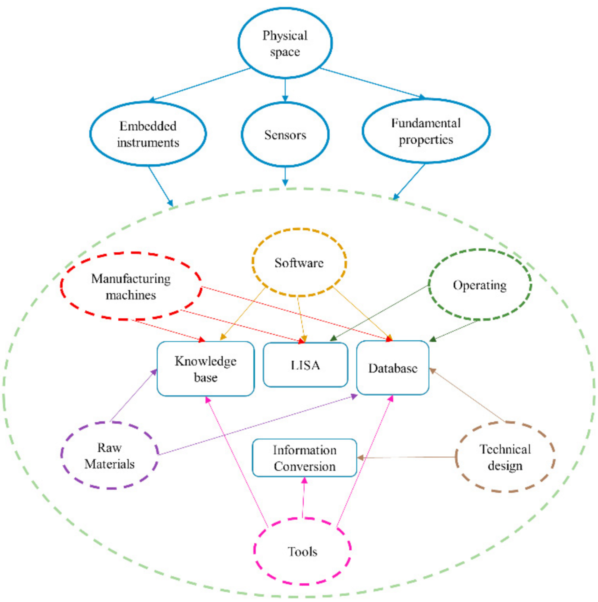

Fig. 3. Manufacturing data collected for the first and second layers of the digital twin. Blue shows the physical space and green shows the cyber space

A manufacturing machine contains technical information, equipment and firmware which placed in the Cyber physical data store, the information conversion- primary processing layer, and the models and algorithm. The host company offers these data, information, and rights to use the software. Operation of a machine contains some level of ownership, which means, for example, that the manufacturer provides operating guidance to their clients, and then based on their analysis, the host company can provide a second level of operating guidance. The host company owns the new information as well. Factory maintenance owns timetables and scheduling that are based on experiences of shareholders. However, the actual action to repair belongs to the company carrying out the maintenance, which can include failure mode analysis and other factors. The service provider collects data from the process after analyzing the data determines the time how to fix a failure and when repair operations should be carried out. This maintenance information is gained from general knowledge, experience and analysis of the product and relevant process. In the digital twin, for example, these data used to predict the failure in parts or improve the life of the parts.

As Table 1b classifies, different available and unavailable data and information of each part of the manufacturing process or production line has a role in different components of the digital twin layers. Understanding the lack of data and information helps to find out accuracy of results from digital twin. For example, many new details regarding technical design are inaccessible because of the competitive market. The same issue happens about the new combined material data. In the case of manufacturing machines and controller, tools, operation, maintaining, and method of processing, experimental data from workers and experts in companies are not available or recorded. Figure 3 presents the first and second layers of the digital twin for manufacturing and the components involved. The figure shows in a manufacturing process or production line, appertain to which parts of the layers. Many instruments and resources collect data and information from the physical space. These data are transferred to the first and second layers of the digital twin. The processed data, information and knowledge are used in the next layers for prediction, optimization, and developing a production line or a process.

5. Discussion

To sum up, descriptions of virtual objects utilize a great deal of data from many different sources and this data is owned by multiple resources. The data typically concerns technical designs, materials, tools, machines, operations, controllers, software, maintenance, and processing methods. As data and information in each area comes from different sources, data ownership is not always clear. For instance, the manufacturer of the machine owns some data about the machine that is being used in the manufacturing process; however, other data concerning the same machine is owned by the product design company and the company using the machine. Therefore, usage of the digital twin encounters issues with data ownership, responsibility for data, access, rights and control. Further questions include whether all the required data resources can be found and whether the required data will be provided.

6. Conclusion

According to the crucial role of data and information in creating reliable digital twin, engineers are confronted with the question of access and permission to collect and use data. The complex ownership structure of data involving many different companies and individuals can become an obstacle. Moreover, if it is possible to acquire and use all the data needed to create a digital twin, the model and results from the digital twin may be claimed by different owners. The underlying issue is how data, information and knowledge gained from the digital twin can be fairly shared between all the owners of the data used to construct and operate the digital twin. Development of an ideal comprehensive digital twin for manufacturing faces the hurdle of unavailable and inaccessible data, and future work should endeavour to find a solution to the data ownership issue. One solution that should be investigated is the possibility of giving the ownership rights of digital twin outputs as compensation for granting access to required data.

References

[1] Cheng GJ, Liu LT, Qiang XJ, Liu Y. Industry 4.0 development and application of intelligent manufacturing. In 2016 international conference on information system and artificial intelligence (ISAI) 2016 Jun 24 (pp. 407-410). IEEE.
[2] Hicks BJ, Culley SJ, Allen RD, Mullineux G. A framework for the requirements of capturing, storing and reusing information and knowledge in engineering design. International journal of information management. 2002 Aug 1;22(4):263-80.
[3] Biagetti T, Sciubba E. Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems. Energy. 2004 Oct 1;29(12-15):2553-72.
[4] Wersig G, Neveling U. The phenomena of interest to information science. The information scientist. 1975 Dec;9(4):127-40.
[5] Peirce CS. Collected papers of charles sanders peirce. Harvard University Press; 1960.
[6] Lohtander M. On the development of object functions and restrictions for shapes made with a turret punch press. Lappeenranra University of Technology; 2010 Dec 17.
[7] Lanz M. Logical and semantic foundations of knowledge representation for assembly and manufacturing processes. 2010.
[8] Garetti M, Rosa P, Terzi S. Life cycle simulation for the design of product–service systems. Computers in Industry. 2012 May 1;63(4):361-9.
[9] M. Grieves, Digital twin: Manufacturing excellence through virtual factory replication, White paper. 2003.
[10] M. Grieves, Digital Twin: Manufacturing Excellence through Virtual Factory Replication, 2014.
[11] Qi Q, Tao F. Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. Ieee Access. 2018 Jan 15;6:3585-93.
[12] Bazaz SM, Lohtander M, Varis J. 5-Dimensional Definition for a Manufacturing Digital Twin. Procedia Manufacturing. 2019 Jan 1;38:1705-12.
[13] Loshin D. Enterprise knowledge management: The data quality approach. Morgan Kaufmann; 2001. (chapter 2. Who owns information)
[14] Lee J, Lapira E, Bagheri B, Kao HA. Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing letters. 2013 Oct 1;1(1):38-41.
[15] Fishbein EA. Ownership of research data. Academic medicine: journal of the Association of American Medical Colleges. 1991 Mar;66(3):129-33.