



**HAL**  
open science

# Challenges in Data Life Cycle Management for Sustainable Cyber-Physical Production Systems

Mélanie Despeisse, Ebru Turanoglu Bekar

► **To cite this version:**

Mélanie Despeisse, Ebru Turanoglu Bekar. Challenges in Data Life Cycle Management for Sustainable Cyber-Physical Production Systems. IFIP International Conference on Advances in Production Management Systems (APMS), Aug 2020, Novi Sad, Serbia. pp.57-65, 10.1007/978-3-030-57997-5\_7. hal-03635699

**HAL Id: hal-03635699**

**<https://inria.hal.science/hal-03635699v1>**

Submitted on 8 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Challenges in Data Life Cycle Management for Sustainable Cyber-Physical Production Systems

Mélanie Despeisse<sup>[0000-0002-1772-3884]</sup> and Ebru Turanoglu Bekar<sup>[0000-0003-4858-4386]</sup>

Chalmers University of Technology, Gothenburg, Sweden  
[melanie.despeisse, ebrut]@chalmers.se

**Abstract.** Rapid technological advances present new opportunities to use industrial Big Data to monitor and improve performance more systematically and more holistically. The on-going fourth industrial revolution, aka Industrie 4.0, holds the promise to support the implementation of sustainability principles in manufacturing. However, much of these opportunities are missed as social and environmental performance are still largely considered as an afterthought or add-on to business as usual. This paper reviews existing data life cycle models and discusses their usefulness for sustainable manufacturing performance management. Finally, we suggest possible directions for further research to promote more sustainable cyber-physical production systems.

**Keywords:** Data Life Cycle, Industrial Big Data, Sustainability Assessment, Sustainable Manufacturing, Sustainability Performance.

## 1 Introduction

Over the last four decades, sustainable development gained importance and is now a top priority in many organizations' and governments' agenda. While sustainability is widely recognized as an imperative for all human activities conceptually, it still lacks presence in organizations' operations as evidence by the lack of action to tackle global challenges and to respect planetary boundaries [1]. Recent technological progress present new opportunities to address these global challenges. Several studies have looked at the sustainability implications of digital technologies in manufacturing [2, 3]. Most notably, advanced data management systems and cloud computing hold promise to systematically collect and manage the data necessary for a sustainability assessment [2]. In turn, this would allow a more holistic approach to improve the performance of manufacturers and their value chain [3] while avoiding rebound effects.

This study aims to reconcile digital technologies and sustainability in manufacturing by investigating how data life cycle models—note that the expression “life cycle” is used in two manners in this paper: “data life cycle” and “life cycle perspective”—can be combined with and support the implementation of sustainability methodologies. We discuss common challenges encountered along the data life cycle, and provide recommendations for sustainability performance management. Finally, we suggest directions for further work to promote more sustainable cyber-physical production systems based on the challenges identified and recommendations listed in this paper.

## 2 Background

### 2.1 Sustainability Methodologies

Numerous methodologies have been developed and demonstrated as effective to support sustainability performance management at different levels [4]. These methodologies can be broadly categorized according to four dominant purposes: (1) *assessment tools* are used to quantify the system's performance or impacts; (2) *improvement tools* are used to identify potential solutions to increase performance or decrease negative impacts; (3) *decision-making tools* help prioritize the solutions to meet predefined targets; finally, (4) *enforcement tools* (e.g. policies and regulations) have a strong influence on decision-making. Many of these tools require high-quality data to generate meaningful and useful results. On-going efforts and collaborations have increased data availability and quality [5], most notably for life cycle-based methodologies [6] such as life cycle assessment (LCA), life cycle costing (LCC), social life cycle assessment (sLCA), and their combination in life cycle sustainability assessment (LCSA) [7]. But these methodologies are still rarely used by organizations.

### 2.2 Data Life Cycle Models

Data life cycle refers to the entire process from generation to use and possibly destruction of the data, sometimes divided into two, three or more stages. In 1993, Levitin and Redman presented two cycles for data acquisition and data usage with the addition of data storage connecting these two cycles [8]. They went on to detail those cycles into a data life cycle model with the enhanced acquisition and usage cycles. To address comprehensively such data quality issues, an extension of Levitin and Redman's data life cycle model [8] was developed by Yoon et al. [9] to include additional phases focusing on metadata before dealing with the data itself. More recently, data life cycle models used by Big Data researcher included the same three main of activities for data management (data generation/collection, storage, processing/utilization), e.g. [10]. Some models detail further some intermediate and end-of-life activities with added emphasis on Big Data and data security [11]. Other models focus on Big Data analytics and machine learning [14, 15]. Major differences between models: inclusion of a planning phase before the collection phase, a pre-processing/calibration/refinement phase before the use phase, and an assessment/improvement phase either alongside or integrated in the use phase. These phases are indeed not always included in industrial practices, resulting in poor data quality, suboptimal exploitation, or data misuse/misinterpretation.

### 2.3 Cyber-Physical Production Systems

The technologies comprised under the trend of "digitalization" can be categorized using the smart factory architecture or cyber-physical production systems (CPPS) comprising four layers: (1) *physical resources*, (2) *network*, (3) *cloud services*, and (4) *terminal* (edge computing) [16]. In addition, a better understanding of Big Data characteristics

and research challenges [17] also helped develop more effective data management systems and data analytics to support process control [18] and decision-making [19]. Finally, the 5-level cyber-physical systems (CPS) architecture [20] and Big Data architecture [19], as well as the maturity model for CPPS [21] provide good foundations to address some of the technical challenges in the digital transformation of manufacturing systems. Much of the work done has potential to align with and support sustainability methodologies (e.g. resource efficiency and collaboration for circular economy). But digitalization can also unlock new levels of productivity and accelerate our linear economy, thereby further transgressing planetary boundaries already exceeded.

## 2.4 Industrial Big Data

Research challenges, opportunities and enabling technologies for Big Data analytics have been discussed from the different perspectives in recent literature studies [22–24]. Big Data challenges can be grouped into three categories [22]: (1) *data challenges* related to data characteristics, e.g. “the seven Vs” [11]; (2) *process challenges* related to data handling, e.g. capture, clean, integrate, transform, store, visualize, etc.; (3) *management challenges*, e.g. security, privacy, information sharing, operational costs, data ownership, governance and ethical concerns. Specific challenges for sustainable smart manufacturing from a product life cycle perspective have been reviewed and categorized into data storage, analysis and processing, and value, also raising concerns about security and ownership [23]. In addition, inadequate technical infrastructures remains one of the biggest challenges of industrial Big Data [24]. Therefore, effective data management using appropriate expertise and secure infrastructure are necessary for Big Data analytics to support smart and sustainable manufacturing [25, 26].

## 3 Challenges in Data Management for Sustainability

In this paper, we have adopted advanced data life cycle model shown in Fig.1 based on Levitin and Redman [8] and using the terminology from Yoon et al. [9] as we found it adapted to focus on challenges associated with each stage of the data life cycle. We also put a strong emphasis on intermediate activities related to data quality assessment and continuous improvement of the data model from the more experimental and innovative approaches to data management. This is to highlight typically overlooked activities, e.g. assess data quality, repair data if possible, discard faulty data which cannot be repaired, obtain new values, and update metadata when necessary.

This section presents typical challenges and resulting recommendations for industrial data management to support sustainability performance monitoring and improvement. We especially emphasize the importance of planning, quality checkpoints, data manipulation, and feedback loops which are often overlooked. To structure our discussion, we grouped activities as follows:

- Data model architecture: metadata creation, structure and refinement;
- Data acquisition and storage: data creation, manipulation, storage and refinement;
- Data exploitation: data assessment, representation/visualization and utilization.

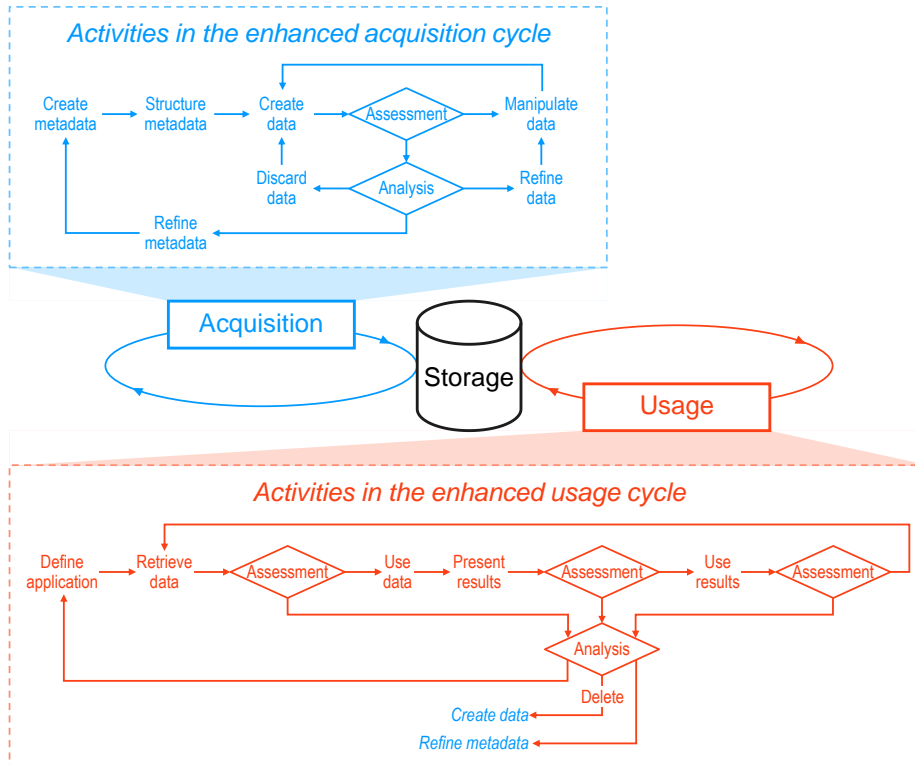


Fig. 1. Data life cycle model (adapted from [8, 9]).

### 3.1 Data Model Architecture: Create, Structure and Refine Metadata

We assume a manufacturing organisation is developing a new data management system for their operations, thus starting the data life cycle with goal and scope definition, which in turn informs the metadata creation. This first planning stage is overlooked. With the cost of information and communication technology (ICT) assets rapidly falling, data is easily collected, much of which is of limited usefulness as little consideration was given to the (long-term) applicability and purpose of the data model. Thus, it is critical to reflect on what is to be achieved, and when possible, experiment with small-scale projects to develop organisational capabilities. Lessons learnt from such pilot applications are invaluable in developing a more robust data model before scaling up. This also allows clarification and alignment with organisational sustainability goals.

The data model architecture is then defined and documented accordingly. A typical limitation encountered at this stage is a reductionist view of manufacturing (gate-to-gate), thereby ignoring externalities and indirect impacts outside the manufacturing site. Lack of standards for system boundaries lead to inconsistencies in sustainability indicators and incompatibilities when combining different assessment methods. This issue is exacerbated with the intrinsic complexity and interdependencies between the three dimensions of sustainability.

Another typical weakness in developing data management systems is the lack of feedback loop and maintenance activities resulting in missed opportunities to correct structural flows and other recurring defects in the data model. This problem is particularly present in organization relying on outdated ICT infrastructure and legacy systems. Although such systems still perform satisfactorily, the organization may miss on new value opportunities and lag behind competitors who are making the transition to state-of-the-art practice and technology. However, upgrading ICT assets as well as skills and capabilities are costly and take time. This has been a central question surrounding the changing nature of jobs in manufacturing to leverage new digital technologies.

Recommendations regarding desired data model characteristics:

- **Model relevance & usefulness:** the model addresses organizational sustainability goals and effectively supports sustainability performance management.
- **Model scope & completeness:** the scope is comprehensive enough to capture relevant physical and information flows, and to enable a holistic performance assessment which accounts for externalities and potential rebound effects.
- **Model essentialness & minimally redundant:** the model complexity is minimised by including only data elements fundamentally required to meet the organization's goals, and is implemented using minimal duplication to meet user needs.
- **Model correctness & consistency:** the model is developed and maintained with appropriate standards to maintain uniform data definitions and avoid data misinterpretation, especially given the multidisciplinary nature of sustainability.
- **Data management:** the model is developed and maintained by a central function which applies systematically data quality engineering methods, and data is managed with distributed control and appropriate data representation to meet end-user needs (local exploitation under the coordination from the central function).
- **Accessibility & searchability:** the model is structured so it is possible to access specific data based on end-users' requirements.
- **Data sharing:** the model addresses data sharing requirements within the organisation and among its collaborators based on accepted standards for metadata definitions to support data exchanges; especially critical given the importance of collaborative networks to achieve higher levels of industrial and societal sustainability.
- **Continuous improvement:** the model is regularly assessed and maintained in a flexible, adaptable manner to meet future requirements (including new applications) using user feedback to determine the value provided to end-users and implement corrective measures (updates) based on lessons learnt.

### 3.2 Data Acquisition and Storage: Create, Manipulate, Store and Refine Data

To create data, various means are used, such as measurement, observation, analysis, or copy. A first challenge related to this activity is a lack of purpose in data collection which can lead to an ineffective model as discussed in the previous section. Another challenge in this early stage of data creation is data improperly recorded, especially for raw data acquisition through sensing, measuring and/or manual entry. Due to a lack of quality checkpoints, flawed and potentially useless data are recorded, contributing to the

typical problems of data overload, especially considering the sheer volume of data generated by modern ICT systems. It is therefore crucial to ask the questions “why do we need the data?” and “is it the correct data for our purpose?” before collecting data.

Data scarcity and unreliability are common hurdles in system modelling and assessment. Implementing sustainability methodologies is often limited by external data availability, trustworthiness, transparency and other data quality issues leading to incomplete or unreliable assessment results. Estimations, calculations and assumptions are sometimes necessary to fill gaps in data, for instance due to inaccessibility to the process or lack of technical solutions for what needs to be measured, e.g. sensor technology does not exist. Incompleteness can also result from other technical issues, e.g. data collected by smart agents and stored in server, but missing connection thus cannot be transferred.

Other gaps are associated with data not collected due to lack of interest or awareness, e.g. waste, pollution and other non-value adding flows not monitored unless unregulated, or data collected with inadequate granularity and quality to be useful in assessing manufacturing performance/impact and identifying improvements. In manufacturing systems, data granularity and sparsity issues are often due to unavailability of machine-level data. However, progress in sensing technologies are quickly removing these barriers (while creating new ones, such as the aforementioned data overload problem).

Even when the right data is collected and stored, there are a number of common pitfalls organizations need to avoid, such as inconsistent terminology or coding practices, inaccurate or missing contextual information (metadata) leading to misinterpretation and misuse, amongst other issues. Once again, these issues need to be addressed in early planning stages and when creating the data model architecture (section 3.1).

Recommendations regarding desired data acquisition characteristics:

- **Completeness, conciseness & relevance:** all (and only) data values necessary for the agreed purpose are acquired.
- **Correctness, reliability & accuracy:** data values are assessed and (when possible) repaired/corrected to ensure they are free from faults or recording defects.
- **Precision & granularity:** data values match the level of detail (e.g. machine, plant, company levels), the time period, and the precision required by the user.
- **Currency & timeliness:** data values are updated as often as required by the user.

### 3.3 Data Exploitation: Assess, Present and Use Data

After data is created, manipulated and stored, it can be accessed and used for different applications. Here, a typical challenge result from misaligned objectives between the model developers and model users as they are often at different levels in the organization. Most of the challenges related to data model architecture and metadata during the planning stages (section on 3.1) will have an impact in the usage cycle, thus need to be addressed early. Other challenges regarding data accessibility, searchability and usability can prevent end-users from engaging with the data management system. A typical cause for this disincentive to use data is the lack of tools and user-friendly interfaces to help access and make sense of the data easily. Recent developments in cloud compu-

ting, connectivity as well as mobile devices are helping in this regard. Augmented, virtual and mixed reality tools are also increasing the usability and intuitiveness of the data by using (more) realistic representations and context-sensitive information.

Despite the rapid growth in industrial Big Data, there is still a lack of tools to make use of this data and especially real-time data to perform continuous sustainability assessments. While sustainability assessment may not be necessary for day-to-day management of operations, it should not be considered as a mere reporting exercise, but as a basis for decision-making. We strongly advocate that data should be used more systematically to support sustainability performance management, thus should be integrated in business-as-usual rather than treated as the occasional “sustainability project”.

Recommendations regarding desired data exploitation characteristics:

- **Accessibility:** data can be retrieve when needed by the user.
- **Timeliness:** data is presented only when needed.
- **Correctness & consistency:** the data is free from retrieval fault, data is presented unaltered from what was stored and according to established standards.
- **Detail & conciseness:** data is presented at an appropriate level of detail according to the user requirements without data loss.
- **Clarity & intuitiveness:** data is presented in a manner that ensure it is understood and used as intended.
- **Presentation & visualization:** data is presented in a manner that facilitates user comprehension.
- **Flexibility & customizability:** data presentation can be customized between systems, formats, media to match user specific needs.
- **Portability & compatibility:** data can move between applications without data loss.

## 4 Conclusion

With advances in ICT, collecting data is becoming ever easier, resulting in industrial Big Data with large amounts of data from multiple sources using different representation methods and structural characteristics. But many organizations still lack data management strategies and architectures to define key parameters and to ensure data quality. To solve these problems, a systematic and standardized approach is needed. This papers describes some typical challenges in data management for sustainability and outlines recommendations regarding desired characteristics for various stage of the data life cycle. Further research is required to identify critical data and prepare the data adequately to enable advanced analytics, and especially machine learning methods for industrial applications of sustainable production strategies.

## References

1. Rockström, J., Steffen, W., Noone, [...], Foley, J.: Planetary boundaries: Exploring the safe operating space for humanity. *Ecol. Soc.* 14, (2009).
2. Stock, T., Obenaus, M., Kunz, S., Kohl, H.: Industry 4.0 as enabler for a sustainable development: A qualitative assessment of its ecological and social potential. *Process Saf. Environ. Prot.* 118, 254–267 (2018).



3. Li, W., Alvandi, S., Kara, S., Thiede, S., Herrmann, C.: Sustainability Cockpit: An integrated tool for continuous assessment and improvement of sustainability in manufacturing. *CIRP Ann. - Manuf. Technol.* 65, 5–8 (2016).
4. Finnveden, G., Moberg, A.: Environmental systems analysis tools - An overview. *J. Clean. Prod.* 13, 1165–1173 (2005).
5. Fazio, S., Kusche, O., Zampori, L.: *Life Cycle Data Network — Handbook for data developers and providers.* (2016)
6. Bjørn, A., Margni, M., Roy, P.-O., Bulle, C., Hauschild, M.Z.: A proposal to measure absolute environmental sustainability in life cycle assessment. *Ecol. Indic.* 63, 1–13 (2016).
7. Sala, S., Farioli, F., Zamagni, A.: Progress in sustainability science: Part I. *Int. J. Life Cycle Assess.* 18, 1653–1672 (2013).
8. Levitin, A. V., Redman, T.C.: A model of the data (life) cycles with application to quality. *Inf. Softw. Technol.* 35, 217–223 (1993).
9. Yoon, V.Y., Aiken, P., Guimaraes, T.: Managing Organizational Data Resources: Quality Dimensions. *Inf. Resour. Manag. J.* 13, 5–13 (2000).
10. Jain, P., Gyanchandani, M., Khare, N.: Big data privacy: a technological perspective and review. *J. Big Data.* 3, (2016).
11. Khan, N., Yaqoob, I., Hashem, I.A.T., [...], M., Gani, A.: Big data: Survey, technologies, opportunities, and challenges. *Sci. World J.* 2014, (2014).
12. Borgman, C.L., Wallis, J.C., Mayernik, M.S., Pepe, A.: Drowning in data: Digital library architecture to support scientific use of embedded sensor networks. In: *Proceedings of the ACM International Conference on Digital Libraries.* pp. 269–277 (2007)
13. Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y.: Big Data for Remote Sensing: Challenges and Opportunities. *Proc. IEEE.* 104, 2207–2219 (2016).
14. Monostori, L., Markus, A., Van Brussel, H., Westkämpfer, E.: Machine learning approaches to manufacturing. *CIRP Ann. - Manuf. Technol.* 45, 675-x7 (1996)
15. Fisher, D., DeLine, R., Czerwinski, M., Drucker, S.: Interactions with big data analytics. *Interactions.* 19, 50–59 (2012).
16. Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., Yin, B.: Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges. *IEEE Access.* 6, 6505–6519 (2017).
17. Costa, C., Santos, M.Y.: Big Data: State-of-the-art concepts, techniques, technologies, modeling approaches and research challenges. *Int. J. Comput. Sci.* 44, 285–301 (2017)
18. Schmidt, M., Moreno, M. V., Schülke, A., Macek, K., Mařík, K., Pastor, A.G.: Optimizing legacy building operation: The evolution into data-driven predictive cyber-physical systems. *Energy Build.* 148, 257–279 (2017).
19. Santos, M.Y., Sá, J., Costa, C., [...], Costa, E.: A big data analytics architecture for industry 4.0. *Adv. Intell. Syst. Comput.* 570, 175–184 (2017).
20. Lee, J., Bagheri, B., Kao, H.-A.: A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manuf. Lett.* 3, 18–23 (2015).
21. Monostori, L., Kádár, B., Bauernhansl, T., [...], Ueda, K.: Cyber-physical systems in manufacturing. *CIRP Ann. - Manuf. Technol.* 65, 621–641 (2016).
22. Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V.: Critical analysis of Big Data challenges and analytical methods. *J. Bus. Res.* 70, 263–286 (2017).
23. Ren, S., Zhang, Y., Liu, Y., Sakao, T., Huisingh, D., Almeida, C.M.V.B.: A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing. *J. Clean. Prod.* 210, 1343–1365 (2019).
24. Al-Abassi, A., Karimipour, H., HaddadPajouh, H., Dehghantanha, A., Parizi, R.M.: Industrial Big Data Analytics: Challenges and Opportunities. In: Choo, K.-K.R. and Dehghantanha, A. (eds.) *Handbook of Big Data Privacy.* pp. 37–61. Springer, Cham (2020)
25. Dubey, R., Gunasekaran, A., Childe, S.J., Wamba, S.F., Papadopoulos, T.: The impact of big data on world-class sustainable manufacturing. *Int. J. Adv. Manuf. Technol.* 84, 631–645 (2016).
26. Raut, R.D., Mangla, S.K., Narwane, V.S., Gardas, B.B., Priyadarshinee, P., Narkhede, B.E.: Linking big data analytics and operational sustainability practices for sustainable business management. *J. Clean. Prod.* 224, 10–24 (2019).