



HAL
open science

Classifying Data Mapping Techniques to Facilitate the Digital Thread and Smart Manufacturing

Laetitia V. Monnier, William Z. Bernstein, Sebti Foufou

► **To cite this version:**

Laetitia V. Monnier, William Z. Bernstein, Sebti Foufou. Classifying Data Mapping Techniques to Facilitate the Digital Thread and Smart Manufacturing. 18th IFIP International Conference on Product Lifecycle Management (PLM), Jul 2021, Curitiba, Brazil. pp.272-283, 10.1007/978-3-030-94399-8_20 . hal-04195222

HAL Id: hal-04195222

<https://inria.hal.science/hal-04195222v1>

Submitted on 4 Sep 2023

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



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Classifying data mapping techniques to facilitate the digital thread and smart manufacturing^{*}

Laetitia V. Monnier^{1,2}, William Z. Bernstein², and Sebti Fofou¹

¹ Laboratoire LIB, Université de Bourgogne Franche Comté, Dijon, France
sfoufou@u-bourgogne.fr

² Engineering Laboratory, National Institute of Standards and Technology,
Gaithersburg, MD 20899, USA {laetitia.monnier,william.bernstein}@nist.gov

Abstract. Representing manufacturing information and maintaining data interoperability are crucial to achieving Smart Manufacturing (SM), also referred to as Industry 4.0. Throughout the Product Lifecycle (PL), data standards have been developed in "silos" based on the specific needs of each PL stage. Dealing with multiple data standards, formats, and representations makes it challenging to transfer information throughout PL stages. Mapping techniques and data translation methods can aid in solving such data heterogeneity challenges. Recent works have demonstrated that general techniques for data mapping and analytics can facilitate more holistic viewpoints. Such viewpoints can help stakeholders make better decisions at the design phase, wherein cost and PL impacts such as those associated with environmental sustainability, are mostly fixed. This paper focuses on analyzing the potential of different types of mapping techniques to enable data interoperability between the multiple PL data formats and standards. We investigate the diversity of data formats in each of the PL data standards. Then, we focus on analyzing the challenges of successfully achieving PL interoperability. Finally, we define the data mapping and translation types with the goal of comparing different approaches to help realize interoperability across global PL management.

Keywords: Data Mapping · Interoperability · Smart Manufacturing · Standards · Systems Integration · Product Lifecycle · Industry 4.0

1 Introduction

The realization of *smart* manufacturing systems remains a core research topic for the Product Lifecycle Management (PLM) community. Smart manufacturing examines the intersection of emerging digital technologies such as Internet of Things (IoT) and artificial intelligence (AI), with advanced engineering capabilities including, for example, multi-material manufacturing and factory automation [12]. At the heart of its implementation, smart manufacturing is rooted in standard data representations (or data models) [6]. Such data models help organize data across a product's lifecycle in a standard and consistent manner to

^{*} Supported by the National Institute of Standards and Technology.

facilitate decision making [8]. One of the key challenges to augment decision making in smart manufacturing systems is the lack of interoperability across these data models. This is partly due to the way in which each of these standards has been developed [11]. Often, standards are constructed with limited scope. For example, to support exchange of product design information between the design and the process planning stage different data structures and languages are used, e.g., EXPRESS for the design stage and eXtensible Modeling Language (XML) format for the planning stage. This has led to a number of challenging issues to coordinate, understand, and visualize links between disparate data streams and structures [3].

Our previous work [18] focused on one such link between data structures from machining process planning to manufacturing execution – specifically between numerical control (NC) code and MTConnect data. The study showcased the difficulty in generalizing or fully automating a direct mapping approach and relied on a priori knowledge of the production environment. To provide a broader perspective on these issues, this paper investigates two different strategies for approaching the mapping problems: (i) using “direct” low-level mapping techniques, e.g., aligning coordinate systems, and (ii) introducing an additional layer in order to have a common data structure between different PL stages, e.g., through an ontology.

To summarize, the research contributions of this work include the following.

1. A brief review of the data formats associated with well-known data standards for smart manufacturing,
2. An articulation of available mapping method types across these standards, and
3. Commentary on potential tradeoffs to realize the digital thread based on our understanding of the data mapping and integration.

2 Product lifecycle data standards format challenges

Standards are vital for the communication and exchange of Product Lifecycle (PL) data. In this section, we review several internationally recognized manufacturing standards that aim to improve interoperability across the product lifecycle. This background will help position the data strategies discussed in Section 3. Note that we particularly focus on manufacturing standards that directly relate to the production stage, i.e., data representations that deliver design information to manufacturers, process data generated by the manufacturer itself, and inspection data related to each product produced.

2.1 Design stage

ISO 10303, commonly referred to as the STandard for Exchange of Product data (STEP), is a series of standards for the computer-readable representation and exchange of product manufacturing information. The primary purpose of the

ISO 10303 series is to properly curate and exchange product model data. The series provides specific application protocols (APs) to address domain-specific concerns, e.g., computer numerical control (CNC) programming with AP238 and automotive design process with AP214.

EXPRESS is the modeling language used for STEP models, as specified in ISO 10301-11 [13]. This language has been developed in order to textually describe and to graphically represent, e.g., through EXPRESS-G, product data using models, schemas and constraints. EXPRESS is an object-oriented language that aims to detail the relationships between the data objects within the data models.

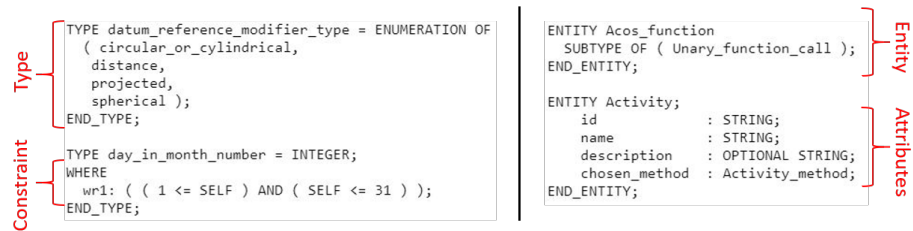


Fig. 1. Excerpts of EXPRESS source code and constructs (based on STEP AP242 Application Reference Model [14])

As seen in Figure 1, the EXPRESS syntax consists of *Entities*, which represents the object in the product description, *Types*, which characterize the different objects, *Attributes*, that relate the properties of the objects, and *Constraints*, that express the product constraints.

STEP AP242 is the ISO 10303 Application Protocol (AP) that supports “Managed model-based 3D engineering” [14] providing means to fully represent geometric dimensioning and tolerancing (GD&T). The standard is based on several modular information models that enable a precise description of the geometric requirements of the design. The models use the EXPRESS language in order to describe the semantics of the schema. The data schema itself is defined as an XML schema (see Figure 1).

2.2 Manufacturing stage

Manufacturing activities generate a wide variety of data including NC code from process planning to equipment-related data from process execution. To represent such different types of data, multiple standards have emerged aiming to capture the characteristics (input and output) of each manufacturing operation.

NC Code describes numerical control standards such as EIA RS274 and ISO 6983 (also known as M and G codes) languages used to program machine tool instructions. The machine's controller follows the given code, which can consist of tool paths, feed rates, and spindle speeds. It is possible to derive an implicit timeline from the step-by-step instructions in a GCode program. Each line of a machine control program contains instructions for the machine [24], usually in the form of a code number, e.g. GCode or MCode, and some data related to the instruction (see Figure 2). The purpose of NC Code is to state actions that the machine will need to execute based on an enforced order of operations creating an implicit timeline. These sets of instructions can result in complex actions and, based on their inherent goal to focus on machine interaction, are not easily understandable by humans.

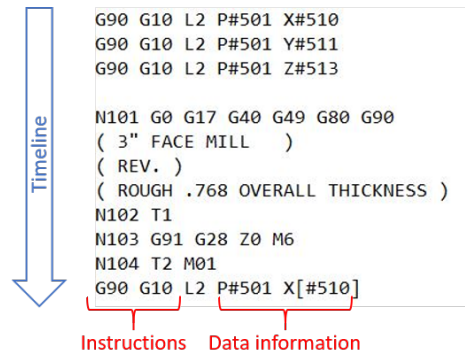


Fig. 2. Example of GCode operation instruction data format.

MTCconnect is a standard exchange protocol that facilitates semantically consistent retrieval of machine log information [19]. The data transferred can include realized feed rate, tool position, and tool path, depending on the machine and its controller. The underlying data format used is XML which contains read-only data that, in this case, contains the logging data of the operating machine. An XML schema is provided in order to relate the XML entities featured in the XML logging file. Each XML entry is a link to an explicit timeline that corresponds to the moment when the machine controller has logged the operation.

2.3 Inspection stage

QIF is a Quality Information Framework (QIF) that describes the data needed to inspect the quality of a product [4]. Multiple types of information, e.g., tolerances, inspection plans, metrology data, and reporting data, are used in order to efficiently allow the analysis of the end product [17]. Figure 3 represents the

QIF architecture consisting of six XML Schema Definition (XSD) models. Each schema has been implemented to represent different types of information used for capturing inspection processes. The schema are named *QIFRules*, *QIFResults*, *QIFPlans*, *QIFProduct*, *QIFStatistics*, *QIFMeasurementResources*.

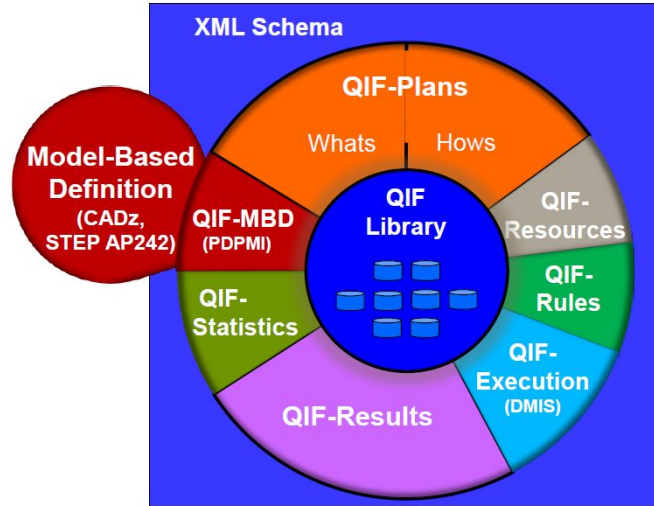


Fig. 3. QIF architecture representation adopted from Michaloski et al. [17].

3 Product lifecycle interoperability

3.1 Data interoperability

Wegner et al. [25] define interoperability as “the ability of two or more software components to cooperate despite differences in language, interface, and execution platform.” The authors highlight two principal methods for achieving this interoperability: data standardization and data bridging. However, interoperability can apply to different notions such as semantic, syntactic, structural, and system [9, 21].

Different standards have been developed in manufacturing for each PL stage. This standardization approach has allowed to normalize the data flow between the PL stages and to facilitate information sharing between different systems. However, due to the wide diversity of the standard formats, this sharing is far from being seamless. Thus, semantic interoperability has become a challenge in this domain where several data standards are used across the PL stages.

3.2 Manufacturing interoperability

Ray et al. [22] discuss the needs of interoperability in manufacturing and showcase an approach that would allow semantic interoperability in order to help overcome the challenges associated with global supply chains. There are three different approaches that could achieve this kind of flow of information: (i) a “point-to-point” approach, which is a customized solution that companies develop depending on their own needs; (ii) a conformity approach, meaning that a consensus must be made between the different proprietary solutions available; and (iii) a “neutral, open, published standards” based approach, focusing on developing global solutions applicable to all manufacturing organizations.

Bringing this method back to the PL presents multiple challenges. Standards have already been in place in the majority of the PL stages, bridging the data standards’ semantics between these stages is a means to facilitate the upstream and downstream flow of information. Data mappings across the standard representations, or realizing the *digital thread*, can be enhanced by (1) developing low-level one-to-one mappings between salient representations, or (2) adding generic layers to bridge the generic schemas across the PL stages, such as through ontology.

The digital thread is an idealized concept that harmonizes a set of technologies allowing a seamless transfer of information between disparate systems. West et al. [26] argued that a true digital thread (and digital twin) of holistic manufacturing systems can become too expensive to implement and maintain. Gielingh et al. [7] underline that “enabling international standards is now available. However, the current technology and the current business models for the offering of solutions appear to be insufficient.” Solutions developed to bridge this gap can be separated into five categories: manual re-input of data, standard systems, direct translation, neutral format translation, and shared databases.

4 Digital thread : direct mapping/translation

Here, we discuss the aforementioned five data mapping techniques with respect to their relevance in improving data interoperability across manufacturing systems. Each information model and data standard used throughout the Product Lifecycle stages presents specific challenges to allow a flow of information. Thus, the complexity of realizing a transfer of information between data formats can be differs drastically.

4.1 Manual re-input of data

This solution includes the translation of all data from one system to another. It is time-consuming and does not guarantee a long-term interoperability of all the systems used in the different Product Lifecycle stages. It also an additional layer of complexity by requiring a validation stage to check translated data and then correct errors. Such corrections often must be done manually since no methods guarantee the accuracy of the manual translation.

Fischer et al. [5] analyzed the Model-Based Enterprise (MBE) strategies adopted by small and medium enterprises (SMEs) in order to improve data interoperability. The study revealed that a large number of SMEs still use manual re-input of the design data for quality inspection. Part of the problem is that the design model generally contains shape and geometry but does not include information needed for performing the inspection process itself, including machine-interpretable instructions.

This method is mostly manual and needs trained technicians who are able to understand and analyze the data. The advantage of the data re-input approach is that no complex implementation is needed for setting it up. However, it is time-consuming as it relies on human knowledge and can be difficult to maintain and share with other stakeholders.

4.2 Standard systems

This solution implies the adoption of a global strategy that will be deployed between all stakeholders. Standards data and systems are selected and implemented with the goal of realizing a unified PL. System operators will only have to be trained once since the underlying framework adheres to a “global” standard. Theoretically, all incorporated technologies and their information infrastructure should be able to be shared and transferred across the entire lifecycle.

Helu et al. [10] showcased a Product Lifecycle Test Bed that aims to promote the uses of cyber-physical infrastructure to create a digital thread across the entire PL. Even though the benefit of developing a fully connected smart manufacturing system is significant, implementing such a structure has been proven to be challenging and expensive. It may be an effective solution at the level of a single company. However, once coordination between different companies is required, all stakeholders must adopt and adhere to the same foundational standards. Considering that industries work with several partners, e.g., suppliers and vendors, enforcing a global system can be arduous and unrealistic.

4.3 Direct translation

This solution uses the same idea of language translators. Direct translation begins by indexing all the words of one language and their meaning in the target language. Each data format is cataloged to create a dictionary of word-to-word translations. This method is straightforward. Many translators and parsers already exist in this form to transfer information between heterogeneous systems.

Several efforts aim to translate and use EXPRESS with different systems. For example, the EXPRESS to XML translator [2] represents the EXPRESS model schema in the XML format, a widely used data format across many domains. Further, the EXPRESS to C++ [23] and EXPRESS generic translator [15] better facilitates translation and the processing of the schema. However, this method involves a heavy maintenance meaning that every change in the standard must also be reflected in the translators.

4.4 Neutral format translation

Neutral format translation relies on the use of an intermediate neutral format that acts as a bridge between the different data standards. This format must be oriented towards the exchange of the data and be common to all target systems. Translators must be implemented for all standard representations to transfer information towards the intermediate data exchange format. This more generic method leverages the *direct translator*³ idea but adds a medium between the individual data models, which may allow for a single “multi-format” translator.

Previous works have considered developing the idea of implementing a “black-box” algorithm in order to link information between GCode and MTConnect files [18]. Here the data from each file is retrieved and then an algorithm matches the corresponding data from one standard to another. The goal is to extend this type of mapping throughout the data Standards used in the Product Lifecycle.

Thus, this solution is not restricted to the case when only two systems are involved but could also be used when multiple actors with different systems must collaborate. However, using a universal format as an intermediate language between the data standards still raises issues concerning the translators’ maintenance and forces all stakeholders to agree on a generic data exchange format.

4.5 Shared databases

This solution elevates the *neutral format translation*⁴ by swapping the generic data exchange format used in the previous method with databases that curate information needed to be transferred between the different data standards. The information of each standard is stored in a common database that can then be linked together and queried. Data pre-processing and translators must be developed depending on the data format in order to transfer the information to the database. This solution adds the possibility to construct a “data lake” [20], e.g., a database containing heterogeneous data types and formats. Data lakes can act as a recipient of information to be transferred between the different data standards of the PL stage.

Ontologies have also been leveraged to store the information of the Product Lifecycle stages. For example, to enhance overall data access and manipulation, Barbau et al. [1] developed a method to translate STEP AP242 files to an ontology [1]. Soonjo et al. [16] took advantage of the previous OntoSTEP work by extending this idea to QIF in transferring the inspection data to a corresponding knowledge graph. Here, the knowledge graph containing the entities of the ontology can be seen as a database shareable between different stages and in which additional information can be added. However, maintenance issues may occur as one needs to ensure that the pre-processing of the data remains consistent with the regular updates of the standards. Furthermore, the method again forces stakeholders to orient around one particular database architecture.

³ See Section 4.3

⁴ See Section 4.4

5 End to end product lifecycle interoperability

5.1 Generic translation method comparison

In Section 4, we described different classes of methods that could improve interoperability and how they have been used in the manufacturing domain. On the one hand, the manual re-input of data (Subsection 4.1) and the standard systems adoption (Subsection 4.2) are customized solutions that are developed by industries to answer their specific needs. On the other hand, the direct translation (Subsection 4.3), neutral format translation (Subsection 4.4), and shared database (Subsection 4.5) methods are more generic solutions that can allow interoperability within a company but can also be applied to multiple stakeholders.

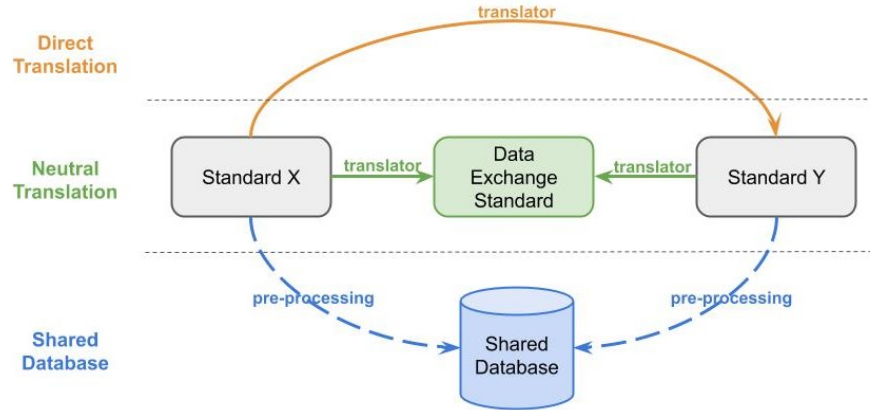


Fig. 4. Possible translation methods to transfer information between two data standards (X and Y).

Figure 4 delineates the implementation between the last three methods, i.e., direct translation, neutral translation, and shared database. The direct translation method used a translator to directly translate a standard representation to another data format. The neutral translation method uses a semi-link translator to translate the information to an intermediate data exchange format. The shared database method uses pre-processing algorithms to transfer information of each standard to a database. Each method uses a medium that can be either a translator, a different standard representation, or a common database to allow information transfer. Based on the use-case and the standard to map the method to apply can vary depending on the implementation feasibility, the maintenance complexity, the data retrieval technique, and the application/domain of the mapping.

5.2 interoperability recommendation

	Direct translation	Neutral format translation	Shared databases
Standards translation	One to one	Multi-standards	Multi-standards data information
Needs	Specific	Global	Data center
Complexity	One translator for each couple of standards	Adapt to different standards	Omniscient of standards
Maintenance	Update for each standard	Partial update depending on the standards	Intermediate data pre-processing update
interoperability	No specific standards	Common and generic data exchange format	Common database

Table 1. Table comparing the characteristics of the different data mapping types.

As seen in Table 1, each mapping method has its own characteristics resulting in advantages and drawbacks based on the use-case on which the mapping must be applied.

Direct translation methods are usually easier to develop and apply to one-to-one translation. However, the maintenance cost and complexity of these methods are significant. Thus, direct translation mapping methods can be appropriate for use-cases that aim to translate data from a specific standard to another.

Neutral format translation is more relevant when multiple standards, e.g., more than two, are concerned. Moreover, this type of translation is more flexible and can easily accommodate changes. For example, if new standards need to be added to the mapping, we only need to define its intermediate translator. Moreover, when a standard is updated, the maintenance of the mapping, e.g., the updating of the mapping, is straightforward having only to apply the changes to the intermediate translator between the related standard and the data exchange standard. Yet, the implementation of such mapping requires significant efforts in creating the intermediate translator for each standard and set up the data exchange standard. Thus, this method is fit for use-cases that will continuously need to map multiple standards.

Shared databases mapping methods are also appropriate when multiple standards are concerned. However, these methods implement data pre-processing algorithms in order to organize and label inputs from different data formats into a common database. Thus, the intermediate pre-processing algorithm needs to be updated when an update is developed for a standard. The implementation of this mapping type requires some effort in the set-up of the common database and the development of the intermediate pre-processing algorithms. Thus, these methods are more appropriate for use-cases that need to continuously deal with multiple standards but focus on retrieving and analyzing the data contained in those standards.

6 Conclusion

The realization of a digital thread to achieve the smart manufacturing systems goal is proven to be arduous as this task can lead to the entire remodeling of the product lifecycle management systems. Indeed, due to the quantity of information and the diversity of standards used to create and manipulate these information, the implementation of true digital threads in smart manufacturing is too expensive to achieve and maintain. Thus, the method chosen to realize this digital thread is crucial for a long-term solution.

We tried to answer this need by analyzing the different techniques that could enable a global product lifecycle data interoperability based on the existing standards used in manufacturing. The main guideline for implementing a data mapping across several standards is to identify the requirements and the complexity of the chosen method. Therefore, we have intended to summarize the existing mapping techniques that could help improve smart manufacturing systems.

References

1. Barbau, R., Krima, S., Rachuri, S., Narayanan, A., Fiorentini, X., Fofou, S., Sriram, R.D.: Ontostep: Enriching product model data using ontologies. *Computer-Aided Design* **44**(6), 575–590 (2012)
2. Barkmeyer, E., Lubell, J.: Xml representation of express models and data. In: *ICSE Workshop on XML Technologies and Software Engineering* (2001)
3. Bernstein, W., Krima, S., Monnier, L., Shahid, M.: Securing, authenticating, and visualizing data-links for manufacturing enterprises. *Model-Based Enterprise Summit (MBE 2019)* p. 51 (2019)
4. Dimensional Metrology Standards Consortium: Part 1: Overview and Fundamental Principles in Quality Information Framework (QIF) - An Integrated Model for Manufacturing Quality Information (2014), <http://qifstandards.org/>, accessed on 03.31.2017
5. Fischer, K., Rosche, P., Trainer, A., Feeney, A.B., Hedberg, T.D.: Investigating the Impact of Standards-Based Interoperability for Design to Manufacturing and Quality in the Supply Chain. Tech. Rep. NIST.GCR.15-1009, National Institute of Standards and Technology, Gaithersburg, MD (2015)
6. Fofou, S., Fenves, S., Bock, C., Rachuri, S., Sriram, R.: A core product model for plm with an illustrative XML implementation. In: *Proceedings International Conference on Product Lifecycle Management*. vol(1). pp. 21–32 (2005)
7. Gielingh, W.: An assessment of the current state of product data technologies. *Computer-Aided Design* **40**(7), 750–759 (2008)
8. Hedberg Jr, T.D., Hartman, N.W., Rosche, P., Fischer, K.: Identified research directions for using manufacturing knowledge earlier in the product life cycle. *International Journal of Production Research* **55**(3), 819–827 (2017). <https://doi.org/10.1080/00207543.2016.1213453>
9. Heiler, S.: Semantic interoperability. *ACM Computing Surveys (CSUR)* **27**(2), 271–273 (1995)
10. Helu, M., Hedberg Jr, T.: Enabling smart manufacturing research and development using a product lifecycle test bed. *Procedia Manufacturing* **1**, 86–97 (2015)

11. Helu, M., Joseph, A., Hedberg Jr, T.: A standards-based approach for linking as-planned to as-fabricated product data. *CIRP Annals* (2018)
12. Helu, M., Libes, D., Lubell, J., Lyons, K., Morris, K.C.: Enabling smart manufacturing technologies for decision-making support. In: *ASME 2016 international design engineering technical conferences and computers and information in engineering conference*. American Society of Mechanical Engineers Digital Collection (2016)
13. ISO, T.: 184/SC 4, ISO 10303-11: 1994 Industrial automation systems and integration-Product data representation and exchange-Part 11: Description methods: The EXPRESS language reference manual. International Organization for Standardization (1994), 00027
14. ISO 10303-42: Industrial automation systems and integration – Product data presentation and exchange – Part 242: Application protocol: Managed model-based 3D engineering. International Organization for Standardization, Geneva, Switzerland (2014)
15. Kahn, H., Filer, N., Williams, A., Whitaker, N.: A generic framework for transforming express information models. *Computer-Aided Design* **33**(7), 501–510 (2001)
16. Kwon, S., Monnier, L.V., Barbau, R., Bernstein, W.Z.: Enriching standards-based digital thread by fusing as-designed and as-inspected data using knowledge graphs. *Advanced Engineering Informatics* **46**, 101102 (2020)
17. Michaloski, J., Hedberg, T., Huang, H., Kramer, T., Michaloski, J.: End-to-end quality information framework (QIF) technology survey. US Department of Commerce, National Institute of Standards and Technology (2016)
18. Monnier, L., Bemstein, W.Z., Fougou, S.: A proposed mapping method for aligning machine execution data to numerical control code. In: *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. pp. 66–72. IEEE (2019)
19. MTConnect Institute: Mtconnect standard (2014), <http://www.mtconnect.org/>, accessed on 03.31.2017
20. Nogueira, I.D., Romdhane, M., Darmont, J.: Modeling data lake metadata with a data vault. In: *Proceedings of the 22nd International Database Engineering & Applications Symposium*. pp. 253–261 (2018)
21. Rachuri, S., Subrahmanian, E., Bouras, A., Fenves, S., Fougou, S., Sriram, R.: Information sharing and exchange in the context of product lifecycle management: Role of standards. *Computer-Aided Design* **40**(7), 89–800 (2008)
22. Ray, S.R., Jones, A.: Manufacturing interoperability. *Journal of Intelligent Manufacturing* **17**(6), 681–688 (2006)
23. Sauder, D., Morris, K.C.: Design of a c++ software library for implementing express: The nist step class library. In: *Proc. of EUG ‘95—The Fifth EXPRESS Users Group Conference*. pp. 21–22 (1995)
24. Smid, P.: *CNC programming handbook: a comprehensive guide to practical CNC programming*. Industrial Press Inc. (2003)
25. Wegner, P.: Interoperability. *ACM Computing Surveys (CSUR)* **28**(1), 285–287 (1996)
26. West, T.D., Blackburn, M.: Is digital thread/digital twin affordable? a systemic assessment of the cost of dod’s latest manhattan project. *Procedia Computer Science* **114**, 47–56 (2017)