ORIGINAL PAPER

FAIR for digital twins

Diana Peters[1] · Sirko Schindler[1]

Abstract

The continuing drive towards digitization in manufacturing leads to an increasing number of digital twins for monitoring and controlling all kinds of processes. While these capture crucial data of all individual steps and allow for analysis and optimization, more often than not the underlying models are confined to individual systems or organizations. This hinders data exchange, especially across institutional borders and thus represents an important barrier for economic success. Similar challenges in the scientific community led to the emergence of the FAIR principles (Findable, Accessible, Interoperable, and Reusable) as guidelines towards a sustainable data landscape. Despite the growing presence within academia, their transfer to industry has not yet received similar attention. We argue that the existing efforts and experiences in science can be exploited to address current data management challenges in industry as well. An improved data exchange within organizations and beyond can not just lower costs, but also opens up new opportunities ranging from discovering new suppliers or partners to improving existing value chains.

Keywords Digital twin · FAIR · MBSE · Semantics

1 Introduction

Digital twins are an essential building block in the digitization of all kind of manufacturing processes. They are "a complete virtual description of a physical product that is accurate to both micro and macro level" [1]. This connection between the physical and virtual worlds is bidirectional: Changes to the physical object get represented in its digital counterpart and the digital twin can act as a surrogate to control the physical object. While the precise definition of digital twins varies between authors, their importance for manufacturing processes is well documented [1, 2].

In the following, we will consider applications of digital twins in the (aero)space domain, like [3–5]. In particular, we will focus on general data management aspects that are independent of concrete implementations or use cases. For

the most part, we will also omit the aspects of digital twins that allow to directly control the physical world in form of, e.g., machines or any other actuators.

The basis of all digital twins is data. But this data does not originate from a single, homogeneous source. Instead, a multitude of sources contribute bits and pieces to the final digital twin. On a macro level, all partners along the value chain may add to the description of a product. Here, we see information like specifications of different components of the final product or the chain of suppliers involved in creating it. On the other hand, there is the micro level that includes each and every machine that is involved in the manufacturing process. This may add data about the environment and parameters during production or the properties of the final product with respect to set tolerance limits. This includes test results of the final product and data about the test environment.

The exact scope and amount of data collected will vary between use cases. Nevertheless, there is one important aspect to all of them: The growing number of data sources leads to quite some challenges due to heterogeneity on all levels. These become even more pressing when integrating digital twins that were designed for different purposes and were not intended to be combined.

Diana Peters and Sirko Schindler contributed equally to this work.

✉ Diana Peters
diana.peters@dlr.de

✉ Sirko Schindler
sirko.schindler@dlr.de

1 Institute of Data Science, German Aerospace Center (DLR), Mälzerstraße 3-5, Jena 07745, Germany

Science suffers from similar challenges: Every moment, a plethora of projects generates a sheer endless amount of data. An ideal, albeit currently rather unrealistic, approach would be to evaluate all existing datasets before setting up new, costly projects for data collection. However, in practice, the data volume alone makes it impossible to conduct an exhaustive search. Another major obstacle to data reuse is the heterogeneity both on data and on metadata level. Countless coding schemes are in use and most domains are quite far from reaching consensus on how to structure or describe their data. This impedes not only human use, but extends to any automated approach to exploit this wealth of data.

As a response, the FAIR principles (Findable, Accessible, Interoperable, and Reusable) have emerged in science as guidelines towards a sustainable data landscape [6]. At the core of their current implementation efforts is the use of Semantic Web technologies on all levels of data management. Open, controlled vocabularies provide the concepts needed to describe both data and metadata[1]. Those concepts can be augmented with additional descriptions and other information and they can be connected to one another, forming the Linked Data Graph [13]. As a result, (meta)data is not isolated anymore, but is now contextualized. Both human and machine users will have an easier time to understand (meta)data and to relate given data annotations with the task at hand.

While these criteria are more and more embraced within the scientific community, also in the context of digital twins [14], their transfer to industry has not yet received similar attention. Exceptions start to emerge, e.g., in a blog post about FAIR principles for a digital twin that is "a virtual representation of the data of something in the real world" [15] or with the recent creation of a Semantic Industry (SemInd) W3C Community Group [16].

At this point, we also want to correct a common misconception: FAIR is not equivalent to open! While it is true that in science there is also a strong drive towards Open Science, the concepts involved are independent of each other. Implementing the FAIR principles in closed environments like companies will already yield considerable benefits for these organizations. Similarly, the restricted data exchange with external partners will be eased substantially without compromising intellectual property any more than by traditional means.

The remainder of this paper is organized as follows: First, in Sect. 2, we will expand on the FAIR principles and outline their potential benefits in an industry context. Later in Sect. 3, we discuss the way forward, which steps need to be taken, and which challenges remain. Finally, we give a brief conclusion (Sect. 4) and outlook (Sect. 5).

2 FAIR in an industrial context

In the following, we will explain each of the FAIR principles based on their definition in [6] and describe their potential impact in the (aero)space manufacturing domain. We will focus on digital twins in particular, but the core ideas are transferable to all industrial datasets in a similar fashion.

2.1 Findable

The first principle is findability—it is the basis for the other three. Data can neither be accessed, nor connected (interoperability), nor reused, if it is not found in the first place. When designing, for example, a spacecraft, engineers want to find components by manufacturers that fit specific needs, e.g., a battery with a particular capacity, maximum mass, and maximum size. Due to the rise of Model-based Systems Engineering (MBSE) [17], more and more machine-actionable models of spacecrafts themselves exist, usually including requirements for the different components being used.

So, there is a machine-actionable model of the "search request". However, the descriptions of manufacturable components are still only available in form of PDF data sheets. While these are accessible to human engineers, they are almost entirely opaque to automated systems. It requires substantial efforts to extract the contained information and create a machine-actionable description (see, e.g., [18]).

On the other hand, manufacturers might also want to find customers—and/or data about the needs of these potential customers. What are their requirements, especially those that are currently *not* met by the manufacturer or their products? Other aspects, besides finding products, manufacturers, and customers, are the quality and reliability of data. How can someone be sure that the product information they *did* find is the most recent? And actually from the manufacturer and not from someone else? Some product or product lines also change over time, so is this *thing* with slightly different parameters still the same product or a different one?

All of these questions concern different aspects of findability in this context.

2.2 Accessible

As mentioned above, it would be desirable to get machine-actionable data directly from manufacturers, not just PDF

---

[1] The Semantic Web and by extension knowledge graphs are an idea to make knowledge accessible to automated use by machines [7, 8]. Current implementations reuse Web technologies such as URLs for referencing [9] and HTTP for communication [10]. In addition, RDF and its serializations are used to store information [11] while SPARQL is used to query it [12]. This provides a set of references (URLs) to well-defined concepts that can be used to semantically annotate (meta)data, replacing free-text annotations in many places and scenarios.

data sheets that need to be processed before being usable. This concerns also the accessibility of data—it should be present in a usable format, preferably one that also the tools consuming the data can directly understand without further conversions needed. Closed, proprietary file and exchange formats are a main barrier here. Almost always, they result in a closed ecosystem and confine their users to a rather restricted set of tools to use, a so called "vendor lock-in". While this favors a few software (and sometimes hardware) vendors, it slows innovation and advancement in the industry as a whole.

Another aspect is authorization, since maybe not all data should be available for everyone. As mentioned above: FAIR does not imply Open Data or free access to data. Datasets can be described in a FAIR way while access might still be restricted. The restriction can either only affect the access to the data itself or already apply to accessing the metadata: If a person is not authorized they might not even learn about the existence of the dataset as already this information might provide some insight and may violate the intellectual property of its owner. In all such cases, FAIR requires the description of any restrictions to be available in a machine-actionable format. Further, the data access protocol has to be able to support authentication and authorization. Both demands combined allow to both document and enforce access restrictions for sensitive data.

A last aspect of accessibility concerns times when data is not available anymore. This may, e.g., be due to the cost of storing it or due to a newer version superseding a dataset. This happens, for example, with newer versions of technical data sheets or if products are no longer produced and therefore the respective data sheets become invalid. In any case, the metadata for all datasets should be maintained—including a statement that this dataset is no longer available and possibly a reason for its disappearance. Users looking for a dataset are immediately notified of the situation and do not invest time in trying to find something that is no longer there.

2.3 Interoperable

One aspect of Findability was to identify a *fitting* product, based on requirements. For this, one does not only have to actually find data about products, but this also has to be compared to requirements—so, interoperability, both among product descriptions and with the initial requirements, is needed. In practise, this affects multiple levels. First, all the data formats must be understandable and compatible (preferably, they are all the same); then, the product parameters must also be comparable semantically. For semantic comparability, it must be clear if, for example, "weight" and "mass" refer to the same physical property; and if they always include or exclude the weight of an insulation layer,

or packaging, or anything else. It must also be clear, in which unit the value of a property is given. In the example of weight and mass: Is it kilogram, gram, pounds, ounces,... something else? To enable interoperability, the units must first be known and then can be converted or unified. A few examples of where mistakes in the understanding of units can lead to when building spacecraft, can be found in NASA's Space Math series [19].

Even if we look just at MBSE tools and their underlying data models—which we classified as already machine-actionable—these models are usually not interoperable, even if they are based on the same standards and developed by agencies like DLR [20] and ESA [21]. Therefore, information exchange is again often carried out via other means like documents that are intended for human use.

2.4 Reusable

In the space domain, engineers do not only want to reuse the data and the virtual twin by reusing a previous model, but it is also relevant to reuse or "use again" particular physical components—so, this also affects the physical twin. During a mission, data about the alteration, aging, errors, breakdown, etc. of components is recorded. This data may inform decisions in future projects on which components to use for particular requirements. However, this requires data in a shape that can actually be reused — including proper documentation of all relevant conditions at the time of recording. So, besides the other criteria discussed so far (findable, accessible, interoperable), the reusability of a component is also based on the reusability of the underlying data.

The reusability of data also leads to other benefits: Models could be reused to save time. If missions are overall similar, the basic components of a satellite stay the same; sometimes there is also the "2.0" version of a mission. Due to the long mission and therefore data life cycles, modeling techniques and/or tools being used have changed in the meantime. Without detailed descriptions of data characteristics, history, and other relevant attributes, previous models are rendered no longer usable. Even a migration to the current environment might not be possible anymore (or would require prohibitive resources), as the necessary knowledge of a model's inner workings has been lost over time. If proprietary tools were used, migration often requires the support of the initial tool vendor, which might no longer be available.

Another aspect of reusability of information is the whole area of "lessons learned". This knowledge is usually already recorded—but often just in a human understandable form, i.e. as reports. The information concerned here is also usually not very structured, but contains decisions, pros and cons regarding a particular solution, and the link between an actually operational system and its (early) design with its requirements. Such approaches hide valuable knowledge in

documents often disconnected from the corresponding data. This makes it rather cumbersome for humans and machines alike to benefit from past experience and apply the gained insight in upcoming projects. However, in a well structured and machine-actionable representation, these hard-won lessons can more easily be reused and can thus prevent any stakeholder from repeating unnecessary mistakes or provide them with well-tried best-practices.

Finally, the FAIR principles include as one of the reusability principles: "R1.1. (meta)data are released with a clear and accessible data usage license" [6]. This demand is related to the previous discussion about enforcing restrictions on data access. It requires companies to state in a clear and machine-actionable way who can use their data under which circumstances. As stated previously already: The fulfillment of FAIR principles does not imply Open Data or no restrictions to data access. The restrictions must just be made clear in a machine-actionable way.

3 A path forward

While we previously outlined the potential benefits of applying the FAIR principles, we will now shift towards approaches for their implementation. We will discuss current developments in the scientific community and how they might be translated into an industrial context. But not all challenges already have solutions. So, we will also highlight remaining challenges applying to both industrial and scientific environments.

3.1 Shared semantics

Shared vocabularies are the basis for any FAIR description. They provide the unique identifiers to use for concepts, objects,... everything! This ensures that it is clear for everyone if two *things* are the same (but have, e.g., different labels) or are actually different. Replace string descriptions by concepts, i.e. use said unique identifiers to link to the concept of, e.g., the physical property of mass. This embeds your data in a system of other physical properties, including units of measurement and their conversion. If you used just the term "mass", it would be unclear what was meant (the physical property, a big amount of something, a fair, a religious mass,...) and it would be highly language dependent. This leads also to the area of disambiguation: Use different identifiers for different meanings. In the "mass" example, all the different concepts should have different unique identifiers. On the other hand, use identical identifiers for identical concepts, even if they have different names. "Weight" in many contexts can be defined as a synonym for "mass" and would therefore be an alternative label for the same concept with the same identifier. Furthermore, hierarchies can be built on top of these concepts, e.g., to indicate hypernyms or

hyponyms[2]. This way, systems may, e.g., know that a solar panel is a kind of power supply and thus inherits a certain set of attributes from the more general concept.

Similarly, all kinds of other connections can be made between any two concepts to encode basically every required information. In particular, this allows for so called *deep semantics*, an approach where concepts are further described as atomic as possible. The result contains multiple layers of different granularity. On the very top, there is, e.g., a single concept for "the reaction wheel #2 in the satellite xy, built by manufacturer z". In many situations this high-level concept suffices to describe the properties of that very reaction wheel, but sometimes the individual components are required, too. While as a human it is pretty easy to see which manufacturer and which satellite are involved, automated systems have a much harder time and need this information to be represented explicitly. So following the previous example, we need two more concepts: one for the the particular satellite and one for the respective manufacturer. All three concepts are then linked to encode all information in a machine-actionable way. The addition of these more fine-grained concepts then allows to query, e.g., for other reaction wheels of the same manufacturer or the other building blocks within the same satellite. Similarly, high-level concepts can be related to each another more easily, thus increasing the interoperability among systems. Matching concepts no longer requires complex natural language processing to determine the meaning of possibly long labels, but can instead rely on the structured, detailed information explicitly provided alongside the corresponding concepts. An example of deep semantics and the decomposition of complex concepts in the field of observable properties is provided in the output of RDA's I-ADOPT working group [22].

There are different knowledge graphs[3] that already exist and can be reused, extended, and built upon. The probably most commonly known is Wikidata[4] [23], though it is very generic and more domain-specific graphs are needed alongside: Buchgeher et al. compiled a review of knowledge graphs in production contexts [24]. NFDI4Ing[5] is the National Research Data Infrastructure for Engineering Sciences in Germany, focusing on research data, but also cataloging[6] knowledge graphs that might be relevant for (engineering)

---

[2] Hypernyms are more general concepts (higher up in the hierarchy), whereas hyponyms are more specialized (lower in the hierarchy).

[3] Knowledge graphs are large networks linking concepts to one another and provide additional information. They can be publicly available like Wikidata [23] and may be limited a specific organization like the ones underlying most modern search engines.

[4] https://www.wikidata.org.

[5] https://nfdi4ing.de/.

[6] https://terminology.nfdi4ing.de.

industry. The initiative MaterialDigital[7] is targeting material data, though projects to develop knowledge graphs there are mostly still at the beginning. The Industrial Ontologies Foundry (IOF)[8] creates ontologies with a focus on the manufacturing and engineering industry. OSMoSE[9] is an initiative by ESA to create an ontology for MBSE in the space domain, though here it seems still unclear how the results will be published. Several ontologies for units also already exist; Keil et al. [25] compared them regarding scope and provided concepts. These are mere examples for a broad range of ongoing initiatives. We advice to use existing knowledge graphs when available over building new ones to save valuable time and effort and to ensure interoperability with other systems that build upon the same knowledge graphs.

Even though a lot of approaches and best practices for shared semantics already exist, some challenges remain. One major question concerns the maintenance of vocabularies: Who should be able to add new concepts or new links between concepts? If this is done by domain experts, then how can the technical quality of the vocabulary be ensured? If this is done by data architects, then how can the correctness from domain perspective be ensured? Are mixed groups needed?

3.2 Metadata

Use semantic (and machine-actionable) metadata. Metadata describes data(sets) or objects and provides summary information. While commonly this is geared primarily at human users, machine-actionable metadata also allows machines to "understand" what they are dealing with. This allows for further automating workflows like more efficient search engines: Instead of transferring entire, potentially huge, datasets, looking into it, and then deciding if it is useful for a particular request, only the metadata may be acquired and the decision is initially based on this. For this purpose, metadata must be extensive and meaningful enough to provide all the information about datasets or models that may be needed at some point: In particular, necessary attributes will already differ between different stakeholders in the same context. It might be even more difficult when the consumer of the provided metadata is not part of the same organisation or ecosystem, like search engines, and thus exact information about which attributes are needed remain somewhat unclear. Furthermore, the scope may include future use cases—new applications might arise that have their own requirements regarding attributes. Here, the challenge is to find the right

balance between the conciseness of a summary and the demands of current and future applications.

Another major aspect is access—who should be able to see and access which (kinds of) data? While generic information like the definition of physical quantities can be shared freely, other aspects like detailed description of supply chains may have to be kept strictly confidential. Metadata alone can already reveal that the company providing a particular dataset (even if the access to the dataset itself is restricted) is working on a particular topic. So, it might be desirable to keep the metadata-set small or provide only a subset outside of your organization—which might affect its findability in a negative way. One approach trying to address these issues is Gaia-X[10]. Here, so called data spaces are supposed to control access to semantically enriched information. Time will tell if Gaia-X is actually able to solve this challenge, at least on a European level.

3.3 Provenance

Document provenance. Provenance data helps contextualizing data and answer questions about its origins. It helps, e.g., to create an overview of the history of a specific component, including the environment conditions it endured. This can help to assure, e.g., the compliance with physical storage requirements (temperature range, humidity range, etc.) and is sometimes needed for certification [26].

It can also help with accountability—who provided which data, who changed it, when, and why? To achieve this, provenance data must be added over the whole life cycle of a component, which leads to the main challenge of this area: The provenance data must be provided by different parties or rather extracted from their different systems. So, this data must also be FAIR and especially interoperable, to be combined into one description of the history of a component.

Looking beyond the scope of a single object, one may also consider entire collections of entities like all products of a specific type. Keeping detailed provenance records allows to conduct comparative analysis [27] like trying to identify the common source of reoccurring issues or quantifying the impact of certain factors. Relying on provenance instead of individual (maybe largely manual) analysis can increase the validity and hence also confidence in the discovered results.

3.4 Open formats

To increase interoperability, use open and machine-actionable formats. Here, "open" refers to the format of data storage, not the data itself. Using proprietary formats of any source will lead to a vendor lock-in at some point. Open and machine-actionable formats in particular ensure that other

---

[7] https://materialdigital.de/.

[8] https://www.industrialontologies.org/.

[9] https://mb4se.esa.int/OSMOSE_Main.html.

[10] https://gaia-x.eu/.

vendors can also develop software and tools to exploit existing data. While this may not yield an immediate benefit right now, it substantially eases the inclusion of or transition to new systems once this becomes necessary.

A repeated concern is that the transition to semantically annotated data requires replacing entire software stacks. A counter example is JSON-LD[11]. Based on the structure of common JSON[12], it adds the means to annotate all entries in a JSON file. However, existing implementations generally require no change at first: JSON-LD is fully backwards compatible. They just ignore the additional entries[13]—no existing entry is changed. On the other hand, newer systems may make use of the added information and are able to translate existing entries into the corresponding semantic concepts. This approach allows to keep the existing infrastructure and only migrate to new systems where needed and appropriate. Furthermore, JSON-LD in addition to being machine-actionable retains the advantages of JSON itself: It is compact and easy to parse since there are only few syntax elements in the standard itself. Further, it is human-readable, which eases troubleshooting during development and deployment.

3.5 Knowledge-aware systems

Semantic, interconnected descriptions enable for a broad range of applications that require access to advanced domain knowledge — previously a field exclusive to human experts. A prime example is search. Major search engines rely on large graphs containing semantically connected information to interpret and thus better respond to user queries[14]. Similarly, specialized systems like Google's Dataset Search[15] build on semantic annotations—in this case using the `schema.org` vocabulary[16] [28]—to provide access to a wealth of information. We also witness an increasing number of information providers who annotate their data accordingly to increase their visibility and allow for more targeted search services [29]. In a parallel to traditional Search Engine Optimization (SEO), we expect that this will become the norm and ignoring this trend will lose substantial business opportunities.

Beyond mere search, machine-actionable descriptions also allow to further automate tasks heavily relying on expert knowledge. An example is matching complex

requirements with an existing supply: In the project Factory of the Future[17], we described the capabilities of individual robots to make allocations for a sequence of tasks to be performed [30]. In another, still ongoing project, we want to suggest suitable ways of manufacturing for requested parts. Here, the focus is on the three-dimensional shape of those parts and allowed tolerances in the final product [31].

Another increasingly important field is question-answering. Unlike conventional search engines, queries do not yield a list of possible sources to get an answer. Instead, based on a knowledge base, the answer is returned immediately, relieving users of the need to work their way through the given *possible* sources [32].

4 Conclusion

In this paper we pointed out how the application of the FAIR principles, originating in research data management, can also benefit industrial contexts and in particular digital twins. We put our focus on examples in the space domain, though many can be generalized towards other production domains. In the last section we pointed out how current approaches to implement the FAIR principles in research might be translated in an industrial context and where open challenges remain—both for industry in particular and for achieving FAIR data ecosystems in general.

Precise and contextualized attributes as advocated by the FAIR principles enable the automation of tasks that formerly required substantial manual effort and were exclusive to human experts. Adding similar capabilities to digital twins allows to escape the information silos that are oftentimes created. Digital twins are no longer exclusive to a specific organization or production process, but can more easily be combined, compared, or partly automated.

5 Outlook

It will of course be interesting to see actual applications of the presented "path forward" in companies, especially to better understand remaining challenges. Some insights can already be gained by the usage of knowledge graphs for digital twins used internally within companies, like IBM [33] and Bosch [34] presented. Both papers mention a challenge we highlighted earlier in this paper: It is very difficult for domain experts to do the semantic modelling, so there is a big need for tool support in this area.

Further, the information exchange between different organisations remains challenging. While the existing publications concern only a single organisation, we envision

---

[11] https://json-ld.org/.

[12] https://www.json.org/.

[13] Of course, existing systems *can* check for unknown entries and handle this as an error, but the JSON(-LD) standard itself offers backward compatibility.

[14] https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels/.

[15] https://datasetsearch.research.google.com/.

[16] https://schema.org/.

[17] https://factory-of-the-future.dlr.de/.

the exchange of information between different stakeholders gaining importance in an increasingly digitized world. Here, aspects like authentication, authorization, and accessibility come more into focus.

Declarations

Conflict of interest No funds, grants, or other support was received and the authors have no relevant financial or non-financial interests to disclose.

References

1. Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B.: Characterising the digital twin: a systematic literature review. CIRP J. Manuf. Sci. Technol. **29**, 36–52 (2020). https://doi.org/10.1016/j.cirpj.2020.02.002

2. Wu, J., Yang, Y., Cheng, X., Zuo, H., Cheng, Z.: The development of digital twin technology review. In: 2020 Chinese Automation Congress (CAC), pp. 4901–4906. IEEE (2020). https://doi.org/10.1109/CAC51589.2020.9327756

3. Yang, W., Zheng, Y., Li, S.: Application status and prospect of digital twin for on-orbit spacecraft. IEEE Access **9**, 106489–106500 (2021). https://doi.org/10.1109/ACCESS.2021.3100683

4. Wang, L., et al.: Application and development prospect of digital twin technology in aerospace. IFAC-PapersOnLine **53**(5), 732–737 (2020). https://doi.org/10.1016/j.ifacol.2021.04.165

5. Peng, Y., Zhang, X., Song, Y., Liu, D.: A low cost flexible digital twin platform for spacecraft lithium-ion battery pack degradation assessment. In: 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–6. IEEE (2019). https://doi.org/10.1109/I2MTC.2019.8827160

6. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B.: The FAIR guiding principles for scientific data management and stewardship. Sci. Data (2016). https://doi.org/10.1038/sdata.2016.18

7. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Sci. Am. **284**(5), 34–43 (2001). https://doi.org/10.1038/scientificamerican0501-34

8. Hogan, A., Blomqvist, E., Cochez, M., D-amato, C., Melo, G.D., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., Ngomo, A.-C.N., Polleres, A., Rashid, S.M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., Zimmermann, A.: Knowledge graphs. ACM Comput. Surv. **54**(4) (2021). https://doi.org/10.1145/3447772

9. WHATWG: URL-living standard (2023). https://url.spec.whatwg.org/

10. Fielding, R., Reschke, J.: Hypertext transfer protocol (HTTP/1.1): message syntax and routing (2014). https://doi.org/10.17487/RFC7230

11. Cyganiak, R., Wood, D., Lanthaler, M.: RDF 1.1 concepts and abstract syntax (2014). https://www.w3.org/TR/rdf11-concepts/

12. Aranda, C.B., Corby, O., Das, S., Feigenbaum, L., Gearon, P., Glimm, B., Harris, S., Hawke, S., Herman, I., Humfrey, N., Michaelis, N., Ogbuji, C., Perry, M., Passant, A., Polleres, A., Prud'hommeaux, E., Seaborne, A., Williams, G.T.: SPARQL 1.1 overview (2013). https://www.w3.org/TR/sparql11-overview/

13. Bizer, C., Heath, T., Berners-Lee, T.: Linked data—the story so far. Int. J. Semantic Web Inf. Syst. **5**(3), 1–22 (2009). https://doi.org/10.4018/jswis.2009081901

14. Schultes, E., Roos, M., da Silva Santos, L.O.B., Guizzardi, G., Bouwman, J., Hankemeier, T., Baak, A., Mons, B.: Fair digital twins for data-intensive research. Front. Big Data **5** (2022). https://doi.org/10.3389/fdata.2022.883341

15. Wharton, M.: Digital twins—FAIRest of them all? (2021). https://www.linkedin.com/pulse/digital-twins-fairest-them-all-mark-wharton/. Accessed 01 Sept 2022

16. Chamanara, J.: Kick-off meeting report (July 2022)|Semantic Industries Community Group (2022). https://www.w3.org/community/semind/2022/07/08/kick-off-meeting-report-july-2022/. Accessed 01 Sept 2022

17. Ramos, A.L., Ferreira, J.V., Barceló, J.: Model-based systems engineering: an emerging approach for modern systems. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **42**(1), 101–111 (2012). https://doi.org/10.1109/TSMCC.2011.2106495

18. Opasjumruskit, K., Böning, S., Schindler, S., Peters, D.: Ontohuman: Ontology-based information extraction tools with human-in-the-loop interaction. In: Luo, Y. (ed.) Cooperative Design, Visualization, and Engineering, pp. 68–74. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16538-2_7

19. Odenwald, S.: Space Math VI. National Aeronautics and Space Administration (NASA), Greenbelt, MD, USA (2010). Chap. 14: Some famous unit conversion errors. https://www.nasa.gov/sites/default/files/files/Space_Math_VI_2015.pdf

20. Fischer, P.M., Lüdtke, D., Lange, C., Roshani, F.-C., Dannemann, F., Gerndt, A.: Implementing model-based system engineering for the whole lifecycle of a spacecraft. CEAS Space J. **9**(3), 351–365 (2017)

21. de Koning, H.P., Gerené, S., Ferreira, I., Pickering, A., Beyer, F., Vennekens, J.: Open concurrent design tool-esa community open source ready to go. In: 6th International Conference on Systems and Concurrent Engineering for Space Applications, Stuttgart (2014)

22. Magagna, B., Moncoiffe, G., Devaraju, A., Stoica, M., Schindler, S., Pamment, A.: InteroperAble Descriptions of Observable Property Terminologies (I-ADOPT) WG-outputs and recommendations. RDA endorsed Recommendations (2022). https://doi.org/10.15497/RDA00071

23. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Commun. ACM **57**(10), 78–85 (2014). https://doi.org/10.1145/2629489

24. Buchgeher, G., Gabauer, D., Martinez-Gil, J., Ehrlinger, L.: Knowledge graphs in manufacturing and production: a systematic literature review. IEEE Access **9**, 55537–55554 (2021). https://doi.org/10.1109/ACCESS.2021.3070395

25. Keil, J.M., Schindler, S.: Comparison and evaluation of ontologies for units of measurement. Semantic Web J. **10**, 33–51 (2018). https://doi.org/10.3233/SW-180310

26. Dressel, F., Doko, A.: Common source & provenance at virtual product house. In: Deutscher Luft- und Raumfahrtkongress 2021. Deutsche Gesellschaft für Luft- und Raumfahrt - Lilienthal-Oberth e.V., Bonn, Germany (2021). https://elib.dlr.de/144515/

27. Johnson, M.A.C., Paradies, M., Dembska, M., Lackeos, K., Klöckner, H.-R., Champion, D.J., Schindler, S.: Astronomical pipeline provenance: a use case evaluation. In: Theory and Practice of Provenance (TaPP 2021) (2021). https://elib.dlr.de/142954/

28. Guha, R.V., Brickley, D., Macbeth, S.: Schema.org. Commun. ACM **59**(2), 44–51 (2016). https://doi.org/10.1145/2844544

29. Benjelloun, O., Chen, S., Noy, N.: Google dataset search by the numbers. In: International Semantic Web Conference, pp. 667–682. Springer (2020). https://doi.org/10.1007/978-3-030-62466-8_41

30. Schäfer, P.M., Steinmetz, F., Schneyer, S., Bachmann, T., Eiband, T., Lay, F.S., Padalkar, A., Sürig, C., Stulp, F., Nottensteiner, K.: Flexible robotic assembly based on ontological representation of tasks, skills, and resources. In: Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning, vol. 18, pp. 702–706 (2021). https://doi.org/10.24963/kr.2021/73

31. Köhler, T., Kleinhenz, L., Schäfer, P.M., Bergmann, J.P., Peters, D.: Development of a methodology for the digital representation of manufacturing technology capabilities. ESSN: 2701-6277 (2021). https://doi.org/10.15488/11259

32. Dimitrakis, E., Sgontzos, K., Tzitzikas, Y.: A survey on question answering systems over linked data and documents. J. Intell. Inf. Syst. **55**(2), 233–259 (2019). https://doi.org/10.1007/s10844-019-00584-7

33. Ploennigs, J., Semertzidis, K., Lorenzi, F., Mihindukulasooriya, N.: Scaling knowledge graphs for automating AI of digital twins. In: Sattler, U., Hogan, A., Keet, M., Presutti, V., Almeida, J.P.A., Takeda, H., Monnin, P., Pirrò, G., d'Amato, C. (eds.) The Semantic Web—ISWC 2022, pp. 810–826. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19433-7_46

34. Kalaycı, E.G., Grangel González, I., Lösch, F., Xiao, G., ul-Mehdi, A., Kharlamov, E., Calvanese, D.: Semantic integration of Bosch manufacturing data using virtual knowledge graphs. In: Pan, J.Z., Tamma, V., d'Amato, C., Janowicz, K., Fu, B., Polleres, A., Seneviratne, O., Kagal, L. (eds.) The Semantic Web—ISWC 2020, pp. 464–481. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62466-8_29