

Towards Addressing Requirements to Identification Posed by the Digital Transformation

Rustam Mehmandarov^{1,2}, Dag Hovland⁴, Torleif Saltvedt³ and Arild Waaler¹

¹*SIRIUS Centre, Department of Informatics, University of Oslo, Norway*

²*Computas AS, Norway*

³*Equinor ASA, Norway*

⁴*Bouvet ASA, Norway*

Abstract

Creating and maintaining a machine-readable mapping of the relationships between the various ways of identifying industrial assets across IT applications, domains, and actors is challenging in large-scale industrial systems. This challenge is usually addressed by using the manual labor of subject matter experts and by creating manual mappings. The automated solution for this challenge has been under-investigated previously. To this end, this paper proposes a classification of identifiers needed for identifying assets on an industrial scale and proposes an approach to digital transformation to address the problem by building upon a model-based approach that has been gaining popularity in recent years. We illustrate our approach with a real industrial example at Equinor, the largest, state-owned Norwegian energy company.

Keywords

Asset Management, Identifier, Asset Identification, Data Integration, Software Interoperability

1. Introduction

Background. In industrial asset management, one of the essential daily routines is the management of many industrial assets, e.g., pumps, pipes, and valves in oil and gas production. Those assets are represented and identified in various ways in the documentation and supporting IT application systems. For example, at Equinor – the largest, state-owned energy company in Norway – there are more than 100 installations, with over 9 million objects identified with TAG numbers. Each TAG number is managed in a wide range of applications, leading to tenfold millions of identifiers.

The Challenge. As the same objects are identified with various identification systems, today, we depend on human intervention from subject matter experts (SMEs) in one or several fields to know that we are talking about the same object or asset. With the rise of digitalization and the increased need for data consolidation and integration, we see a clear need to have those mappings generated and presented in a reproducible and machine-readable way. Furthermore,

SemIIM'22: International Workshop on Semantic Industrial Information Modelling, 30th May 2022, Hersonissos, Greece, co-located with 19th Extended Semantic Web Conference (ESWC 2022)

✉ rustamm@ifi.uio.no (R. Mehmandarov); dag.hovland@bouvet.no (D. Hovland); salt@equinor.com (T. Saltvedt); arild@ifi.uio.no (A. Waaler)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

today's working practice is often based on the sequential processing of data by SMEs and then by data engineers leading to bottlenecks and unnecessary delays in the process.

The Idea. To address this challenge, we aim to automatically generate the mappings between the identifiers and the objects by inferring the relationship mapping using model-based integration instead of the current document-based data exchange. We introduce our model-based approach that will serve SMEs and data engineers working with data pipelines and integrations, including the creation of automated machine-to-machine data integrations.

The Vision. The industry's move towards model-based documentation yields many improvements to the current document-based approach. The model-based approach combined with the suggested models will simplify the data exchange process and improve support for management of change (MoC).

Both of these tasks are challenging and resource-intensive in the current work practice – as the data exposed about an asset will vary significantly based on the context, the task at hand, and the role of the system actor. This differentiation is challenging as various user groups and applications would generally need to map different identification methods for the same asset, as well as to have the ability to represent the asset in different ways based on the context and the task at hand.

This approach will make it possible to process data on different levels in parallel instead of the more sequential processing we see today. The proposed solution will help optimize these processes, considering the various needs of user groups or automated machine-to-machine interfaces, consolidating the information about an asset across multiple domains and applications, and improving software interoperability and human collaboration.

2. The Current State of Affairs

Today's document-centered practice requires a lot of resources and manual work to consolidate and exchange information across the value chain and life cycles of engineering projects and assets. We are seeing a shift towards model-based approaches that can support subject matter experts, digital twins, and automated data exchange through APIs, data mesh [1, 2], data pipelines, or similar.

However, creating models across multiple domains, applications, and value chains, raises the need for a universal way to refer to an object represented by various distinct identifiers. It also questions how these object references should be managed and mapped.

Despite being a common problem in the industrial setting, data integration and mapping is still an area yet to be widely researched. It has commonly been solved using manual labor, and proprietary solutions [3, 4]. This work is typically done by SMEs working closely with IT and data integration experts, such as data engineers and data integrators.

Furthermore, this complex problem can be divided into several challenges. To understand those challenges and opportunities better, we must look closely at the actors and how each actor identifies the assets and the relationships between them. Change management for identifiers at hand is yet another challenge for the industry that we shed light on in this paper.

3. Approach

We want to start describing our approach with introducing classifications essential to understanding and solving the challenge at hand.

3.1. Classification of Identifier Systems

In engineering systems, we often see various ways of identifying assets. We want to start by distinguishing two groups of identification systems based on their uniqueness outside their context and the amount of information they carry. Later we would also like to separate identifiers into categories based on their lifecycle management. These categorisations are based on the classification of the typical identifiers used in the engineering systems.

We would like to start by dividing identifiers into two distinct categories based on their usage:

- **Descriptors** – often context-dependent identifiers (i.e., unique only within a specific context), bearing encoded information about the asset and breakdown structures that SMEs use. Those often are presented in a human-readable format.
- **Identifiers** – identifiers used to identify data entries uniquely, often not intended to be human-readable or convey meaningful information.

While descriptors are easy for the experts and engineers to use within their domains, they carry too much information from that domain and are often dependent on the context to be unique. In other words, a descriptor has no value if it can not uniquely identify an asset outside its context. Therefore, we need identifiers to be able to supply that identification. On the other hand, even though identifiers are more likely to be unique outside their context, they will typically be less human-readable. As we can see, descriptors and identifiers have specific usages in respectively engineering systems and the underlying IT applications and can not be interchanged.

Furthermore, we need to look at the classification of identifiers from the lifecycle management perspective. This classification is fundamental when looking at change management for the data and assets. We usually see this in engineering systems where data has to be exchanged and updated across the value chain, domains, applications, or similar. The four main categories are:

- **Self-managed identifiers** can be managed by one entity without synchronizing the identifier generation or naming with other entities. This is typically internal identification that is being used within a specific context.
- **Co-managed identifiers** have to be synchronized across multiple entities and thus can not be easily changed without proper synchronization and clarification with other parties involved.
- **Unmanaged identifiers** that can be generated in a distributed manner and do not need any management from any party, except for agreeing on the algorithm, such as UUIDs and GUIDs.
- **Centrally managed identifiers** – their use and assignment have to be managed and coordinated by one specific body, e.g., TAG numbers in the current engineering practice.

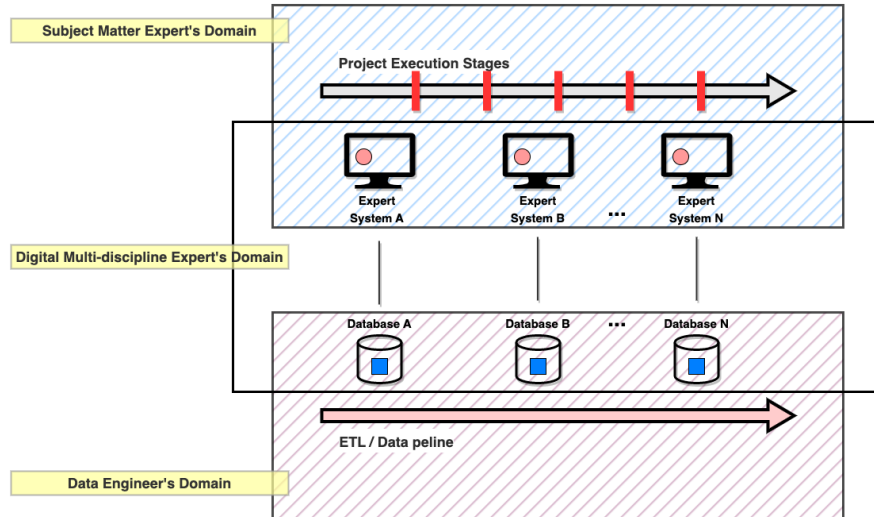


Figure 1: Domains of SMEs, Data Engineers, and Digital Multi-Discipline Experts.

The introduced classification is also in line with the Industrie 4.0 view on identifier management [5] and integrates well into Reference Designation System (RDS) codes defined in ISO/IEC81346-1 [6].

So far, we have introduced two ways of classifying the identifiers based on their usage and lifecycle. However, we still need to address the challenge of being able to map various kinds of identifiers across multiple applications and domains. We also need to make the data mapping and exchange process less sequential so that the various system actors can work on their parts without creating bottlenecks for each other.

3.2. Classification of System Actors

To help address the issues mentioned above, we would like to introduce three types of actors in expert systems involved in building and maintaining assets in engineering:

- ***Subject matter experts (SMEs)***
- ***Data engineers / IT experts***
- ***Digital multi-discipline experts.***

If we combine this classification with the identifier classification described above, we get a much clearer picture with the separate areas of responsibilities shown in Fig. 1.

Seen from the perspective of a *subject matter expert*, information about assets resides in a set of systems varying in the domain (e.g., electrical engineering, mechanical engineering), project execution stages (i.e., project lifecycle stages), and value chain (e.g., supplier, contractor). SMEs often relate to project-specific *descriptors* (e.g., TAG numbers, other engineering numbering systems) and use that information to identify objects. The same identifier can be traced throughout various systems, domains, and disciplines that may also have their own descriptors. Descriptors

are also made to be human-readable and often contain encoded breakdown structure information. Such descriptors are used across various systems and diagrams, such as Piping and Instrumentation Diagrams (P&ID), Master Equipment Lists (MEL), or process flow diagrams (PFD).

The data of the engineering applications used by SMEs is stored in databases, which are often tailored to these systems. Data import and export usually happen through application programming interfaces (APIs) or specialized data pipelines. The data from the expert systems lies in the databases and other ancillary IT systems. That data typically uses other types of identification, which we refer to as *identifiers*. The *data engineers* and *IT experts* working on creating and maintaining the IT applications, databases, data integrations, and data pipelines usually use identifiers as a primary object identification method. These identifiers are often technical and designed to be unique within their context and focus less on human readability. Typical examples include GUIDs, UUIDs, auto-incremented numbers, or serial numbers for the equipment.

The challenge is that the mapping between descriptors and identifiers across applications needs to be explicit. In addition, SMEs and data engineers often only stick to either of the distinct categories of identifiers – descriptors or identifiers. Both types are specialized for the workflows in which they are used and, therefore, are not optimal for use by other workflows. These challenges underline the importance of supporting both types of identification.

In addition to the two user groups already mentioned, we would also like to introduce the *digital multi-discipline experts* that work in the cross-section of one or several domains and IT, who map the information and help the two other user groups to move forward. This group would often need to have a good understanding of both domain knowledge and the IT applications, as well as types and mapping of the identifiers.

3.3. Requirements

To summarize the needs described above, we will need to create a solution that will be able to solve the following challenge:

Create tailored data "views" to show relevant data to a specific actor. Views will be based on which task and at what stage of the project lifecycle it needs to be performed, consolidating all the available data from various applications.

Furthermore, we have identified the following requirements to address the challenge:

1. Data about an asset should be presented based on the actor's context, need, and specific task the actor needs to perform. The tasks are defined by factors like domain, project and product lifecycles, and value chain.
2. There should be non-manual ways of mapping descriptors and identifiers across domains, applications, and value chains to facilitate seamless data integration.

3.4. Digital Transformation

The proposed approach will build further on the model-oriented way of working through better digitalization initiatives, like Industrie 4.0, Asset Administration Shell [7] and digital twins [8], to name a few. The idea is to facilitate the creation of various "views" for the

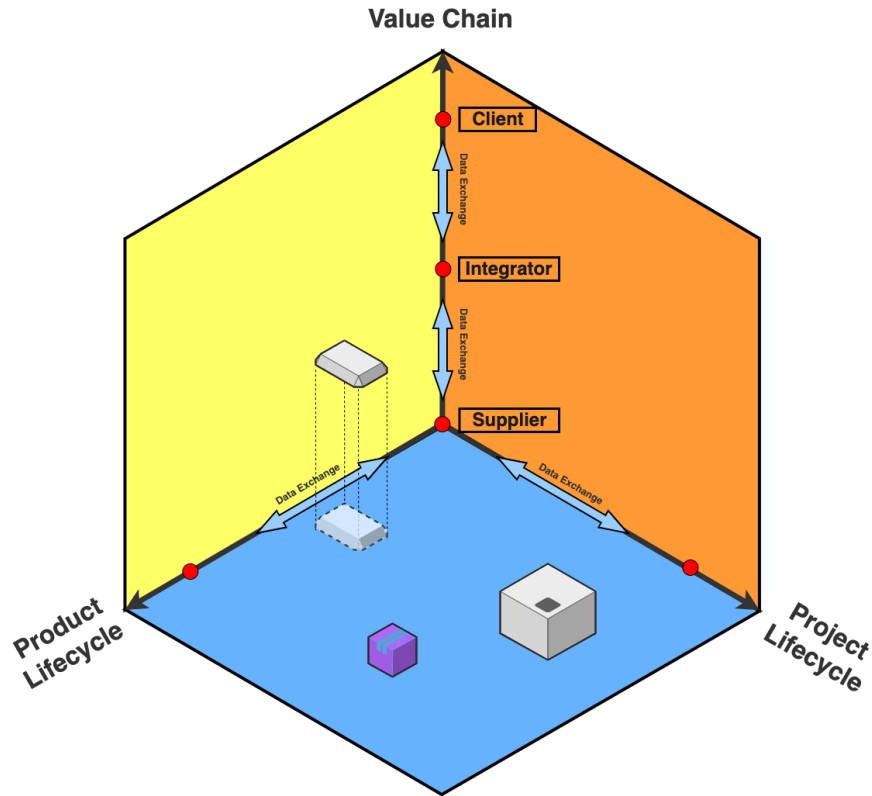


Figure 2: Various dimensions for the identification systems that are represented by a simplified visualization of different identification systems that can be applied to an asset. Each axis represents an example data flow between three dimensions: value chain, product lifecycle, and project lifecycle.

system's actors based on all the available information and aspects mentioned above. The variation and complexity of identification methods that will need to be mapped are illustrated as a multidimensional plane in Fig. 2. It should also be noted that this figure simplifies the real-world situation, where more than three dimensions are needed to map between different identification models. Identifiers for the same object will also not necessarily map one-to-one between identification models. However, they would often rather have one-to-many mappings that need to be addressed in the new approach of identifier mapping. Such identification mapping and support for multiple dimensions is an important step to support the Industrie 4.0 approach, specifically the Reference Architecture Model for Industrie 4.0 (RAMI 4.0) [9].

3.5. The Proposed Systematic Model

To satisfy the abovementioned needs, we propose classifying the user groups into three distinct categories and introducing systematic knowledge or models.

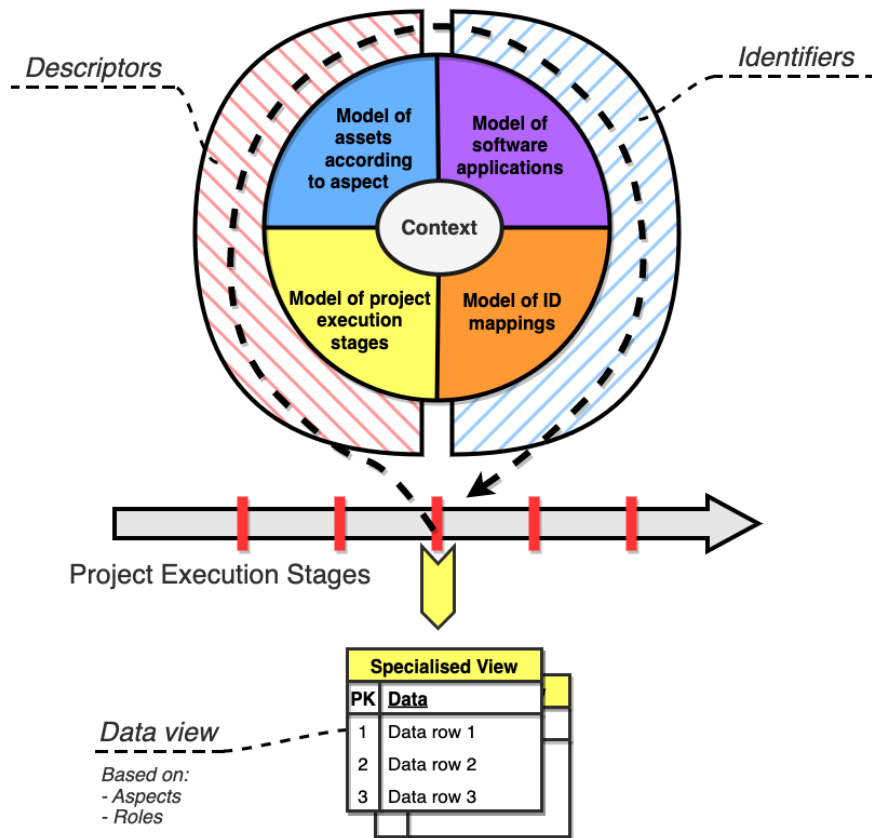


Figure 3: A conceptual view of the proposed model integrated data views.

1. **Project execution steps (facility lifecycle)** – a piece of information about steps in the lifecycle of a facility;
2. **Asset representation according to the aspect¹**, including the domain and asset lifecycle;
3. **Software applications including the data they represent and store** – information about what data is stored where and how it can be retrieved and updated;
4. **Relationships and mappings of identifier and descriptor types** – information about how the descriptors and identifiers are composed for each type to be able to create mappings

As we can see from Fig. 3, the context plays a central role in all those models. Context can vary from model to model but can typically be something that helps uniquely identify the

¹The *aspect* is defined by ISO/IEC 81346-1:2022[6, 10] as views to sort out and monitor the technical information of objects. The 81346-1 part defines four fundamental aspects that can be used in the ISO/IEC 81346 standard series and can be extended with other aspects.

information elements queried to the models. For instance, for a software applications model, that can be information about which application is requesting the information about a specific identifier or the current facility lifecycle stage.

Context is meant to add necessary information to a query or identifier to uniquely identify an object with a model's scope. We need this information since identifiers are often only unique with a specific scope. If we refer to those identifiers outside their scope, we need to add more information to ensure their uniqueness within the larger scope.

From the Fig. 3, we can also see that the different models will have different kinds of predominant identifiers used to refer to the object – some models will be using various types of descriptors, and some will be using multiple types of identifiers. To be able to extract the necessary data for the particular actor on a specific stage in the project lifecycle, we would have to be able to query all those models. This further underlines the need for a model mapping the identifiers in use.

More about the parts of the technical implementation for such a system has been drafted in work by Mehmandarov R. et al. previously [11].

4. Conclusion

This paper introduced a classification of identifiers, their types, and actors in the industrial context. Furthermore, we have defined a need and the requirements for identifying industrial assets posed by digital transformation.

We propose to address this problem by introducing several models in a machine-readable way. This approach will help us to automate the data integration between various endpoints and domains, as opposed to the time- and resource-consuming manual processes that exist in today's practice. This approach aligns well with the industry's current move to a model-driven approach and data exchange and distribution trends, such as data mesh architecture.

We are implementing our approach using an example of Equinor and other prominent actors in the Norwegian energy sector. In the future, we plan to validate our approach with more data and industrial users and implement it in an industrial evaluation environment. We also aim to develop a semantic theory of the approach in formal semantics and reasoning.

Acknowledgements

The work was partially supported by the SIRIUS Centre, Norwegian Research Council project number 237898.

References

- [1] Z. Dehghani, Data mesh principles and logical architecture, 2020. URL: <https://martinfowler.com/articles/data-mesh-principles.html>.
- [2] Z. Dehghani, Data Mesh, O'Reilly, 2022.
- [3] A. Soyly, E. Kharlamov, D. Zheleznyakov, E. Jimenez-Ruiz, M. Giese, M. G. Skjæveland, D. Hovland, R. Schlatte, S. Brandt, H. Lie, et al., OptiqueVQS: A visual query system over ontologies for industry, *Semantic Web 9* (2018) 627–660.
- [4] S. Fillinger, E. Esche, G. Tolksdorf, W. Welscher, G. Wozny, J.-U. Repke, Data exchange for process engineering – challenges and opportunities, *Chemie Ingenieur Technik 91* (2019) 256–267. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cite.201800122>. doi:<https://doi.org/10.1002/cite.201800122>.
- [5] Plattform-Industrie-4.0, AAS Reference Modelling, 2021. URL: https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/AAS_Reference_Modelling.pdf.
- [6] ISO, ISO/IEC81346-1 Industrial systems, installations and equipment and industrial products - Structuring principles and reference designations - Part 1: Basic rules, 2022. URL: <https://www.iso.org/standard/82229.html>.
- [7] E. Tantik, R. Anderl, Integrated data model and structure for the asset administration shell in industrie 4.0, *Procedia Cirp 60* (2017) 86–91.
- [8] E. Kharlamov, F. Martin-Recuerda, B. Perry, D. Cameron, R. Fjellheim, A. Waaler, Towards semantically enhanced digital twins, in: 2018 IEEE International Conference on Big Data (Big Data), IEEE, 2018, pp. 4189–4193.
- [9] K. Schweichhart, Reference Architectural Model Industrie 4.0 (RAMI 4.0), 2017. URL: https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference_architectural_model_industrie_4.0_rami_4.0.pdf.
- [10] ISO, The RDS 81346 Standard Series, 2022. URL: <https://www.81346.com/81346-1>.
- [11] R. Mehmandarov, A. Waaler, D. Cameron, R. Fjellheim, T. B. Pettersen, A semantic approach to identifier management in engineering systems, in: 2021 IEEE International Conference on Big Data (Big Data), IEEE, 2021, pp. 4613–4616. doi:<https://doi.org/10.1109/BigData52589.2021.9671515>.